THE UNIVERSITY *of* EDINBURGH

Networks, Complexity and Internet Regulation

Scale-Free Law

Andres Guadamuz

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy by Publication(PhD)

The University of Edinburgh

February 2013

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Contents

Figures and Tables

FIGURES

TABLES

Abbreviations

ARPANET	Advanced Research Projects Agency Network
AS	Autonomous System
CAS	Complex Adaptive System
CC	Creative Commons
CERN	European Organization for Nuclear Research
DDoS	Distributed Denial of Service
DMCA	Digital Millennium Copyright Act
DNS	Domain Name System
DoS	Denial of Service
DS	Dynamic Systems
DVD	Digital Video Disc
ECD	Electronic Commerce Directive
EFF	Electronic Frontiers Foundation
EU	European Union
EUR	Euro
FOSS	Free and Open Source Software
GBP	British Pound
GPL	General Public License
GSCC	Giant Strongly Connected Component
HADOPI	Haute Autorité pour la diffusion des oeuvres et la protection Internet
HTTP	Hyper Text Transfer Protocol
IAB	Internet Architecture Board

IANA	Internet Assigned Numbers Authority
ICANN	Internet Corporation for Assigned Names and Numbers
IESG	Internet Engineering Steering Group
IETF	Internet Engineering Task Force
IFPI	International Federation of the Phonographic Industry
IGF	Internet Governance Forum
IMPs	Message Processors
IP	Internet Protocol
IRC	Internet Relay Chat
ISOC	Internet Society
ISP	Internet Service Provider
IWF	Internet Watch Foundation
LAN	Large Area Network
P2P	Peer-to-peer
PING	Packet InterNet Groper
PRS	Performing Right Society
RIAA	Recording Industry Association of America
RO	Read-Only
RW	Read/Write culture
SCC	Strong Country Centrality
SLOC	Single Lines of Code
SMTP	Send Mail Transfer Protocol
SNA	Social Network Analysis
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UK	United Kingdom
UN	United Nations
US	United States
USD	United States Dollar

USPTO	United States Patent and Trademarks Office
W3C	Wide Web Consortium
WCT	WIPO Copyright Treaty
WIPO	World Intellectual Property Organization
WSIS	World Summit of the Information Society
WWW	World Wide Web

Cases

A&M Records v. Napster, 2002 U.S. App. LEXIS 4752.

AdobeSystems Inc. v. Tripod Inc., No. 1:96CV157 (N.D. W.Va.).

Arista Records LLC v. Lime Group LLC, 2010 U.S. Dist. LEXIS 46638.

Authors Guild v Google Inc, (United States District Court for the Southern District of New York, Docket No 2005 CV 8136, filed September 20, 2005).

Capitol Records, Inc. v. Thomas, 579 F. Supp. 2d 1210.

Frank Music v. CompuServe Inc., No. 93 Civ. 8153 (S.D.N.Y. 1993).

Gibbons v. Ogden 22 U.S. 1.

In re Aimster Copyright Litig., 2004 U.S. App. LEXIS 1449.

McCulloch v. Maryland 17 U.S. 316

McGraw-Hill Cos, Inc v Google Inc, (United States District Court for the Southern District of New York, Docket No 2005 CV 08881, filed October 19, 2005).

Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 125 S. Ct. 2764.

Playboy Enterprises, Inc. v. Frena, 839 F. Supp. 1552 (M.D.Fla. 1993).

Roadshow Films Pty Ltd v iiNet Limited [2010] FCA 24.

Sabam V. S.A. Tiscali (Scarlet), District Court Of Brussels, No. 04/8975/A (29 June 2007).

Sega Enterprises v. Sabella, No. C93-04260 (N.D. Cal. 1996).

Sony BMG Music Entertainment Sweden AB et al. v Niej et al, Stockholm Tingsrätt, Case B 13301-06 (2009).

Sony BMG Music Entertainment v. Tenenbaum, 672 F. Supp. 2d 217.

Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417.

Universal Music Australia v Sharman License Holdings [2005] FCA 1242.

Viacom International Inc., et al. v. YouTube Inc., et al., Nos. 07-Civ-2103 (LLS), 07-Civ-3582 (LLS) (S.D.N.Y. June 24, 2010).
Warner Bros. Entertainment Inc. et al v. RDR Books et al (575 F.Supp.2d 513).

Acknowledgments

Whenever I present at a conference, I always start with an apology, and this book is no exception to the rule. I am sure that I will miss mentioning people who have been very important in the completion of this work, but I tend to keep the formalities to a minimum, so apologies to those who I have missed.

This book is the result of generous funding from the Arts and Humanities Research Council, and therefore it is one of the deliverables for the Network Architecture research stream within the second round funding of the SCRIPT Centre for Studies in Intellectual Property and Technology Law at the University of Edinburgh.

The work has benefited from discussions, conference presentations, feedback and input from a vast network of colleagues and friends. Burkhard Shafer, Abbe Brown and Lilian Edwards have always been around to listen to me go on about networks, and their input and friendship is always appreciated. Special thanks go to Shawn Harmon and Judith Rauhofer for their friendship and support. Johanna Gibson provided encouragement when I first became interested in this subject, and her continuing friendship is always welcome. Fiona Macmillan gave me the opportunity to first present an earlier version of this work at one of her excellent conferences, and had to endure me using her as an example of a central hub in the conference network. She also kindly published a book chapter containing some initial sketches of the ideas presented here. Hector MacQueen and Sheldon Halpern were instrumental in publishing an article version of some of my ideas on networks and copyright with the Albany Law Review. Finally, Carol George provided some needed proof-reading in the final stages of completion.

The cover is a map of the Internet drawn using network visualisation tools from the OPTE Project (http://www.opte.org/), and is reproduced with permission. Colour

versions of all images in the book will be available at the website http://www.technollama.co.uk/book.

As usual, all faults and errors are my own.

License

THE EXTENT THIS LICENSE MAY BE CONSIDERED TO BE A CONTRACT, THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

1. Definitions

a. **"Adaptation"** means a work based upon the Work, or upon the Work and other pre-existing works, such as a translation, adaptation, derivative work, arrangement of music or other alterations of a literary or artistic work, or phonogram or performance and includes cinematographic adaptations or any other form in which the Work may be recast, transformed, or adapted including in any form recognizably derived from the original, except that a work that constitutes a Collection will not be considered an Adaptation for the purpose of this License. For the avoidance of doubt, where the Work is a musical work, performance or phonogram, the synchronization of the Work in timed-relation with a moving image ("synching") will be considered an Adaptation for the purpose of this License.

b. **"Collection"** means a collection of literary or artistic works, such as encyclopedias and anthologies, or performances, phonograms or broadcasts, or other works or subject matter other than works listed in Section 1(f) below, which, by reason of the selection and arrangement of their contents, constitute intellectual creations, in which the Work is included in its entirety in unmodified form along with one or more other contributions, each constituting separate and independent works in themselves, which together are assembled into a collective whole. A work that constitutes a Collection will not be considered an Adaptation (as defined above) for the purposes of this License.

c. **"Distribute"** means to make available to the public the original and copies of the Work through sale or other transfer of ownership.

d. **"Licensor"** means the individual, individuals, entity or entities that offer(s) the Work under the terms of this License.

e. **"Original Author"** means, in the case of a literary or artistic work, the individual, individuals, entity or entities who created the Work or if no individual or entity can be identified, the publisher; and in addition (i) in the case of a performance the actors, singers, musicians, dancers, and other persons who act, sing, deliver, declaim, play in, interpret or otherwise perform literary or artistic works or expressions of folklore; (ii) in the case of a phonogram the producer being the person or legal entity who first fixes the sounds of a performance or other sounds; and, (iii) in the case of broadcasts, the organization that transmits the broadcast.

f. **"Work"** means the literary and/or artistic work offered under the terms of this License including without limitation any production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression including digital form, such as a book, pamphlet and other writing; a lecture, address, sermon or other work of the same nature; a dramatic or dramatico-musical work; a choreographic work or entertainment in dumb show; a musical composition with or without words; a cinematographic work to which are assimilated works expressed by a process analogous to cinematography; a work of drawing, painting, architecture, sculpture, engraving or lithography; a photographic work to which are assimilated works expressed by a process analogous to photography; a work of applied art; an illustration, map, plan, sketch or three-dimensional work relative to geography, topography, architecture or science; a performance; a broadcast; a phonogram; a compilation of data to the extent it is protected as a copyrightable work; or a work performed by a variety or circus performer to the extent it is not otherwise considered a literary or artistic work.

g. **"You"** means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation.

h. **"Publicly Perform"** means to perform public recitations of the Work and to communicate to the public those public recitations, by any means or process, including by wire or wireless means or public digital performances; to make available to the public Works in such a way that members of the public may access these Works from a place and at a place individually chosen by them; to perform the Work to the public by any means or process and the communication to the public of the performances of the Work, including by public digital performance; to broadcast and rebroadcast the Work by any means including signs, sounds or images.

i. **"Reproduce"** means to make copies of the Work by any means including without limitation by sound or visual recordings and the right of fixation and reproducing fixations of the Work, including storage of a protected performance or phonogram in digital form or other electronic medium.

2. Fair Dealing Rights.

Nothing in this License is intended to reduce, limit, or restrict any uses free from copyright or rights arising from limitations or exceptions that are provided for in connection with the copyright protection under copyright law or other applicable laws.

3. License Grant.

Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) license to exercise the rights in the Work as stated below:

1. to Reproduce the Work, to incorporate the Work into one or more Collections, and to Reproduce the Work as incorporated in the Collections; and,
2. to Distribute and Publicly Perform the Work including as incorporated in Collections.

The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are

technically necessary to exercise the rights in other media and formats, but otherwise you have no rights to make Adaptations. Subject to 8(f), all rights not expressly granted by Licensor are hereby reserved, including but not limited to the rights set forth in Section 4(d).

4. Restrictions.

The license granted in Section 3 above is expressly made subject to and limited by the following restrictions:

1. You may Distribute or Publicly Perform the Work only under the terms of this License. You must include a copy of, or the Uniform Resource Identifier (URI) for, this License with every copy of the Work You Distribute or Publicly Perform. You may not offer or impose any terms on the Work that restrict the terms of this License or the ability of the recipient of the Work to exercise the rights granted to that recipient under the terms of the License. You may not sublicense the Work. You must keep intact all notices that refer to this License and to the disclaimer of warranties with every copy of the Work You Distribute or Publicly Perform. When You Distribute or Publicly Perform the Work, You may not impose any effective technological measures on the Work that restrict the ability of a recipient of the Work from You to exercise the rights granted to that recipient under the terms of the License. This Section 4(a) applies to the Work as incorporated in a Collection, but this does not require the Collection apart from the Work itself to be made subject to the terms of this License. If You create a Collection, upon notice from any Licensor You must, to the extent practicable, remove from the Collection any credit as required by Section 4(c), as requested.
2. You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital file-sharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary

compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

3. If You Distribute, or Publicly Perform the Work or Collections, You must, unless a request has been made pursuant to Section 4(a), keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing: (i) the name of the Original Author (or pseudonym, if applicable) if supplied, and/or if the Original Author and/or Licensor designate another party or parties (e.g., a sponsor institute, publishing entity, journal) for attribution ("Attribution Parties") in Licensor's copyright notice, terms of service or by other reasonable means, the name of such party or parties; (ii) the title of the Work if supplied; (iii) to the extent reasonably practicable, the URI, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work. The credit required by this Section 4(c) may be implemented in any reasonable manner; provided, however, that in the case of a Collection, at a minimum such credit will appear, if a credit for all contributing authors of Collection appears, then as part of these credits and in a manner at least as prominent as the credits for the other contributing authors. For the avoidance of doubt, You may only use the credit required by this Section for the purpose of attribution in the manner set out above and, by exercising Your rights under this License, You may not implicitly or explicitly assert or imply any connection with, sponsorship or endorsement by the Original Author, Licensor and/or Attribution Parties, as appropriate, of You or Your use of the Work, without the separate, express prior written permission of the Original Author, Licensor and/or Attribution Parties.
4. For the avoidance of doubt:
    a. **Non-waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme cannot be waived, the Licensor reserves the exclusive

right to collect such royalties for any exercise by You of the rights granted under this License;

b. **Waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme can be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License if Your exercise of such rights is for a purpose or use which is otherwise than noncommercial as permitted under Section 4(b) and otherwise waives the right to collect royalties through any statutory or compulsory licensing scheme; and,

c. **Voluntary License Schemes**. The Licensor reserves the right to collect royalties, whether individually or, in the event that the Licensor is a member of a collecting society that administers voluntary licensing schemes, via that society, from any exercise by You of the rights granted under this License that is for a purpose or use which is otherwise than noncommercial as permitted under Section 4(b).

5. Except as otherwise agreed in writing by the Licensor or as may be otherwise permitted by applicable law, if You Reproduce, Distribute or Publicly Perform the Work either by itself or as part of any Collections, You must not distort, mutilate, modify or take other derogatory action in relation to the Work which would be prejudicial to the Original Author's honor or reputation.

5. Representations, Warranties and Disclaimer.

UNLESS OTHERWISE MUTUALLY AGREED BY THE PARTIES IN WRITING, LICENSOR OFFERS THE WORK AS-IS AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR THE ABSENCE OF LATENT

OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO SUCH EXCLUSION MAY NOT APPLY TO YOU.

6. Limitation on Liability.

EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. Termination

a. This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Individuals or entities who have received Collections from You under this License, however, will not have their licenses terminated provided such individuals or entities remain in full compliance with those licenses. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this License.
b. Subject to the above terms and conditions, the license granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

8. Miscellaneous

a. Each time You Distribute or Publicly Perform the Work or a Collection, the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License.
b. If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.
c. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.
d. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You.
e. The rights granted under, and the subject matter referenced, in this License were drafted utilizing the terminology of the Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), the Rome Convention of 1961, the WIPO Copyright Treaty of 1996, the WIPO Performances and Phonograms Treaty of 1996 and the Universal Copyright Convention (as revised on July 24, 1971). These rights and subject matter take effect in the relevant jurisdiction in which the License terms are sought to be enforced according to the corresponding provisions of the implementation of those treaty provisions in the applicable national law. If the standard suite of rights granted under applicable copyright law includes additional rights not granted under this License, such additional rights are deemed to be included in the

License; this License is not intended to restrict the license of any rights under applicable law.

CREATIVE COMMONS NOTICE

Creative Commons is not a party to this License, and makes no warranty whatsoever in connection with the Work. Creative Commons will not be liable to You or any party on any legal theory for any damages whatsoever, including without limitation any general, special, incidental or consequential damages arising in connection to this license. Notwithstanding the foregoing two (2) sentences, if Creative Commons has expressly identified itself as the Licensor hereunder, it shall have all rights and obligations of Licensor.

Except for the limited purpose of indicating to the public that the Work is licensed under the CCPL, Creative Commons does not authorize the use by either party of the trademark "Creative Commons" or any related trademark or logo of Creative Commons without the prior written consent of Creative Commons. Any permitted use will be in compliance with Creative Commons' then-current trademark usage guidelines, as may be published on its website or otherwise made available upon request from time to time. For the avoidance of doubt, this trademark restriction does not form part of this License.

Creative Commons may be contacted at http://creativecommons.org/.

1. Introduction

> PSYCHOHISTORY - … Gaal Dornick, using non-mathematical concepts, has defined psychohistory to be that branch of mathematics which deals with the reactions of human conglomerates to fixed and social stimuli…
> Isaac Asimov, *Foundation*[1]

Isaac Asimov is usually placed at or near the top of any listing of the most important science fiction writers. His works have introduced a number of concepts that have captivated the imagination of the public and, perhaps more importantly, inspired scientists, to the extent that some of his ideas have been the subject of academic consideration. The Three Laws of Robotics, for example, introduced in his short story, *Runaround,* have informed research and debate in the fields of artificial intelligence, robotics and information technology.[2] Psychohistory, another important concept explored in Asimov's novels, could be described as the scientific prediction of the behaviour of large human conglomerates acting in large numbers. In his Foundation series, Asimov recounts the way in which this mathematical modelling of a large Galactic society is performed, and the problems that ultimately arise from trusting such mechanisms. Psychohistory is based on three postulates:

1. The population under study must be unaware that the predictions are taking place.
2. The predictions must be conducted over periods of three consecutive generations.
3. To ensure the accuracy of statistical probability, the population in question must number in the billions.

---

1. Asimov I, *Foundation*, London: Octopus Books (1983), p.17.
2. See for example: Clarke R, "Asimov's Laws Of Robotics: Implications for Information Technology" 26:12-27:1 *IEEE Computer* (1993-1994).

One of the main plots in the Foundation novels is that a hidden cabal is dedicated to making sure that history continues along the path predicted by its inventor, mathematician Hari Seldon. Asimov seems to imply that the predictive science of psychohistory is doomed to eventual failure because it can only foresee large events, and it does not (and cannot) take into account the actions of remarkable individuals. He suggests that human and robotic intervention is necessary for the accuracy of psychohistory, which can be taken as a satisfying compromise between determinism and free will, mechanism and individuality.

Despite this apparent indictment, the predictive capacity of psychohistory remains a powerful ideal for some. Nobel Prize winning economist Paul Krugman cites psychohistory as one of the reasons he studied economics:

> Those who read [science fiction] may be aware of the classic Foundation trilogy by Isaac Asimov. It is one of the few science fiction series that deals with social scientists – the 'psychohistorians', who use their understanding of the mathematics of society to save civilization as the Galactic Empire collapses. I loved Foundation, and in my early teens my secret fantasy was to become a psychohistorian. Unfortunately, there's no such thing (yet). […] As for social sciences other than economics, I am interested in their subjects but cannot get excited about their methods – the power of economic models to show how plausible assumptions yield surprising conclusions, to distil clear insights from seemingly murky issues, has no counterpart yet in political science or sociology. Someday there will exist a unified social science of the kind that Asimov imagined, but for the time being economics is as close to psychohistory as you can get.[3]

The present work does not assume to be a study in psychohistory. I am using the concept to illustrate a vital concept that will be proposed throughout the following pages. This book is concerned with a narrow and specific area of legal study, that of Internet regulation. Psychohistory cannot be written in this way, but the idea behind it remains. The underlying assumption in this work is that there are analytical and

3. Krugman P, *Incidents from my Career*, http://web.mit.edu/krugman/www/incidents.html.

descriptive tools that are more comfortable in the realm of mathematics than in the social sciences. Before describing the objective and reach of this work, I will try to explain the background to the idea that one can bring both together.

A SHORT HISTORY OF PSYCHOHISTORY

One of the presuppositions of the study of human interaction is that human behaviour is too complex and chaotic to allow anything even remotely like psychohistory to take shape. This seemingly insurmountable stumbling block is at the heart of the stark methodological division that exists between the natural and social sciences, a split that has become an almost unshakeable feature of modern academia, and that is played out on a daily basis in university campuses around the world. It is perhaps important to point out that although we have grown accustomed to the separation of the hard sciences and social disciplines, this division is a relatively recent development. While social sciences may be seen as the poor relative of scientific endeavour, they have, over extended periods of time, aspired to adopt methodological approaches used in the study of natural phenomena.[4] It was the work of authors such as Habermas, Bernstein and Marcuse that defined and expanded the gap and promoted the idea that the social sciences are an entirely separate set of disciplines, with their own methodology and approach to empirical research.[5] Since then, social science has become involved in critical theory, and increasingly split from the ideals of what Habermas calls materialistic science, becoming something else entirely.

---

4. Bernstein RJ, *The Restructuring of Social and Political Theory*, Philadelphia: University of Pennsylvania Press (1978), p.xvi.
5. Particularly relevant to this debate is: Habermas J, *Knowledge and Human Interests*, 2nd [English] ed, London: Heinemann Educational (1978), Chapter 3.

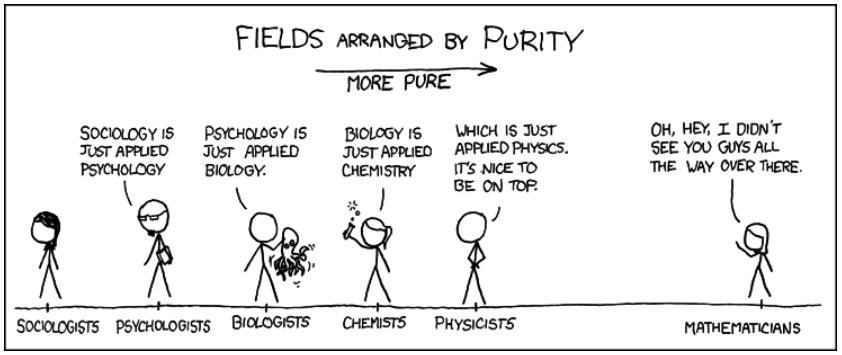


Figure 1.1 Fields of science according to purity[6]

The critical theory that has characterised many social sciences since the latter part of the 20th century can be seen as a reaction to the hierarchical and structured view of the world that had dominated Western thought since the Enlightenment. As a reaction to this materialist world, the social sciences adopted a non-hierarchical and unstructured way of looking at reality.[7] In certain extreme versions of critical theory a form of relativism rules, in which it is possible to deconstruct almost anything –including natural science– into its cultural origins. This trend further reinforced the schism between the natural and the social sciences, resulting in an acrimonious divorce and eventually to the Sokal hoax. Physicist Alan Sokal published an article in the prestigious social science journal *Social Text*, claiming to establish a critical theory of quantum gravity,[8] a spoof that served to polarise opinions in both areas of study. On the one hand, some natural scientists could not disguise their contempt and glee at the comeuppance of disciplines that some consider little more than gibberish.[9] The response of cultural theorists, on the other hand,

6. Xkcd, *Purity*, http://xkcd.com/435/ (released under a Creative Commons licence).
7. Hart K, *Postmodernism: A Beginner's Guide*, Oxford: Oneworld (2004).
8. Sokal AD, "Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity", 14:1-2 *Social Text* 217 (1996).
9. With varying degree of animosity. For some reactions, see: Koertge N, *A House Built on Sand: Exposing Postmodernist Myths About Science*, Oxford: Oxford University Press (1998).

ranged from the meek recognition that something might be wrong, to a barrage of invective directed at Sokal.[10]

An interesting introspection arose, however, out of the Sokal affair, and there seems to be genuine willingness to try to get past the science wars.[11] There is a legitimate argument to be made about the uselessness of furthering the current state of affairs. Should science remain split between the seemingly objective physical sciences and the presumably subjective social sciences? Is there room for philosophers to have a say about natural phenomena, and for mathematicians to comment on social issues? As has been hinted at already, the scientific split is relatively recent, and there is growing interest in reverting to a more interdisciplinary approach to the relationship between the natural and social sciences. Philip Ball calls it the physical modelling of human social systems,[12] which can be described as the use of methodological and empirical tools prevalent in the physical sciences to describe social interaction. In other words, the science of psychohistory is born.

The creation of a branch of study that employs tools used in the study of mould, gases and sub-atomic particles, and applies them to complex human behaviour, is the logical result of a line of thought that has been growing in credence since the Enlightenment: that social sciences have the capacity for more predictive precision, much like the so-called hard sciences of chemistry, physics and biology. Such a powerful idea may seem counter-intuitive to those in the academic world who have come to rely and thrive on the clear separation of disciplines described above. The idea that societies might respond along deterministic paths, and that their behaviour could be charted by physics and mathematics contradicts the concepts of agency and free will that have dominated much of philosophical thought in the last centuries. This modern idea that human affairs are

---

10. See for example: Newman F, "One dogma of dialectical materialism", 1 *Annual Review of Critical Psychology* 83 (1999).
11. Two thoughtful pieces in the journal *Physics Today* can be highlighted as offering a balance view of the affair: Gottfried K, "Opinion – Was Sokal's Hoax Justified?", 50:1 *Physics Today* 5 (1997); and Beller M, "The Sokal Hoax: At Whom Are We Laughing?", 51:9 *Physics Today* 7 (1998).
12. Ball P, "The Physical Modelling of Human Social Systems", 1 *Complexus* 190 (2003).

akin to the exact sciences can, however, be traced back to the 17th century, when several philosophers sought to address both natural and human philosophy. Gottfried Wilhelm Leibniz is perhaps one of the best examples of a man who was comfortable talking about the nature of matter[13] and comparative history,[14] and in whose works one may find in the same paragraph mathematical equations and musings about human freedom.[15]

One could argue that such overlap of magisteria was the logical result of the nature of human progress at that time, as philosophers dealt interchangeably with the natural world, theological discourse and social phenomena. One could also say that the eventual schism between social and physical sciences became necessary once the number of subjects of study became too vast for any one person to handle, preventing furtherance of knowledge in their field of study. The Renaissance Man has become a figure of times past, and specialisation is the norm. I answer these hypothetical objections with two questions. Were our predecessors wrong to try to look at human endeavours with the same analytical tools that informed their scientific thinking? Has this apparent divorce between mathematics and society been for the best?

At a time when the secrets of the universe were being unlocked, and during which nature displayed astonishing exactitude, it must have been tempting to assume that the mysteries of the inner workings of society would also eventually be uncovered to show similar clockwork precision.[16] Seventeenth century philosopher Thomas Hobbes is often referred to as the father of the mechanistic view of society.[17] Although he is better known for his political philosophy, Hobbes was clearly inspired by his mentor Francis Bacon, the father of natural philosophy. In his works we encounter a strong adherence to

13. Leibniz GWF, *Monadology and Other Philosophical Essays*, Indianapolis, IN: Bobbs-Merrill Company (1965).
14. Perkins F, *Leibniz and China: A Commerce of Light*, Cambridge: Cambridge University Press (2004).
15. Leibniz GWF, "Freedom and Possibility", in *Philosophical Essays,* Indianapolis, IN: Hackett Publishing (1989), pp.19–22.
16. Vinnicombe T, "Thomas Hobbes and the Displacement of Political Philosophy", 32:8 *International Journal of Social Economics* 667 (2005), p.668.
17. For example, Ball P, *Critical Mass: How One Thing Leads to Another*, London: Arrow Books (2004), pp.7–37.

rationality and the stricture of social systems that is the precursor of political thought in the following centuries. In the *Leviathan*, he wrote:

> To conclude, the light of humane minds is perspicuous words, but by exact definitions first snuffed, and purged from ambiguity; reason is the pace; increase of science, the way; and the benefit of mankind, the end. And, on the contrary, metaphors, and senseless and ambiguous words are like ignes fatui; and reasoning upon them is wandering amongst innumerable absurdities; and their end, contention and sedition, or contempt.[18]

Although Hobbes predates the work of Isaac Newton, his words herald a world in which it is the precision of science that presents us with the first glimpse of the attainability of objective truth. It is the clockwork universe unveiled by Newton that seems to have unleashed a new generation of philosophers intent on marrying the Hobbesian ideals of society and the exactitude of mathematics. Philip Ball comments that:

> A political scientist taking a chronological approach would track the trajectory of Hobbes's thought via Locke to later thinkers that believed there could be a 'calculus of society'. Along this path we would uncover Jeremy Bentham's utilitarianism in the late eighteenth century, an attempt to harmonize the individual's personal happiness with the interests of society. […] Bentham and the Philosophical Radicals, who included John Stuart Mill, paved the way for the socialism of Karl Marx.[19]

It is also in the 17th century that another vital relationship between mathematics and social sciences starts to appear: that of finance and economics. The foundations of a theory of supply and demand were, for example, famously laid by John Locke in a letter to the Members of Parliament in 1691.[20] Similarly, an often overlooked fact is that a few years later, in 1696, Sir Isaac Newton took on the role of Warden of the Mint, and in 1699 became Master of the Mint. He is credited (or discredited depending on your point

---

18. Hobbes T, *Leviathan*, New York: Barnes & Noble Publishing (2004), p.30.
19. Ball, supra note 17, p.34.
20. Locke J, "Some Considerations of the Consequences of the Lowering of Interest and the Raising the Value of Money", in Medema SG and Samuels WJ (eds), *The History of Economic Thought: A Reader*, London: Routledge (2003), pp.57–77.

of view) with having moved Britain from the silver to the gold standard[21] during this crucial period, thereby shaping an English monetary policy that was to endure into the 20th century. It is no coincidence that these two figures, more famous for their political and scientific works, were united in their interest in monetary policy. After all, the mystery of the markets may have seemed like just another area of potential discovery for the soundest minds of the time.

Given this background, it should come as little surprise that Adam Smith, the father of free market economics, was also a philosopher. Smith's earlier academic life was spent teaching logic and moral philosophy at Glasgow University, and it was only later that he turned his attention to law and economics.[22] He is perhaps the best representative of the line of thinkers that believed in hidden forces behind social phenomena. In both *The Theory of Moral Sentiments*,[23] and *The Wealth of Nations*,[24] Smith introduces the idea that market actors, while pursuing self-interest, are guided by an invisible hand that acts to the benefit of society. While much ink has been spent on discussing the precise meaning of Smith's invisible hand,[25] it is clear that Smith believed that human endeavours were controlled by hidden currents, expressing what was perhaps a precursor to the ideas of complexity and emergence that will be subject of this book.

At the other side of the political spectrum, philosopher and social scientist Friedrich Engels also dedicated considerable time to discussion of the natural sciences. In *Herr Eugen Dühring's Revolution in Science* and in his unfinished work, *Dialectics of Nature*, Engels proposes ways in which socialist dialectics could be applied to the latest developments in science and mathematics. Simply put, dialectics is a way of looking at history and society as opposition, negation and transformation in smooth and constant

21. For more about this, see: Findlay-Shirras G and Craig JH, "Sir Isaac Newton and the Currency", 55:218 *The Economic Journal*, (1945), pp.217–241.
22. Buchan J, *The Authentic Adam Smith: His Life and Ideas*, New York: W.W. Norton (2006).
23. Smith A, *The Theory of Moral Sentiments*, New York: A.M. Kelley (1966), IV.I.10.
24. Smith A et al, *An Inquiry into the Nature and Causes of the Wealth of Nations*, Indianapolis, IN: Liberty Press (1981), IV.2.9.
25. For example, see: Minowitz P, "Adam Smith's Invisible Hands" 1:3 *Economy Journal Watch* 381 (2004).

fluctuation. Engels believed that natural scientists could learn from the methodology contained in dialectics by forgetting their own preconceptions. He wrote:

> Nature is the proof of dialectics, and it must be said for modern science that it has furnished this proof with very rich materials increasing daily, and thus has shown that, in the last resort, nature works dialectically and not metaphysically. But the naturalists who have learned to think dialectically are few and far between, and this conflict of the results of discovery with preconceived modes of thinking explains the endless confusion now reigning in theoretical natural science, the despair of teachers as well as learners, of authors and readers alike.[26]

Karl Marx was heavily inspired by Engels, yet he goes further in his ideas about history. In truly psychohistorian fashion, Marx believed that not only was history shaped by Engels' dialectics, but that history could be read scientifically and that economic laws drove all markets, be they labour or commodities.[27] Those who could understand these laws could therefore foresee the result of future social conflicts.

These are just some illustrations of the strong philosophical tendency to borrow the language and methods of so-called hard sciences for use in the charting of social phenomena. There is an abundance of other scholars, scientists and thinkers who may be cited for their adoption of physical modelling,[28] but it is not the objective of this work to provide a comprehensive examination of them.

Despite the eventual divorce of the natural and the social described above, some of these ideas survived (and thrived) in the 20th century. The torch-bearer of interdisciplinary studies since the writings of Adam Smith has been economics, and in that discipline, one of the foremost examples of the attempt to understand human behaviour through the language of mathematics can be found in the discipline of game theory. Put simply, game theory is a systematised way of ascribing mathematical

---

26. Engels F and Dühring EK, *Anti-Dühring: Herr Eugen Dühring's Revolution in Science,* London: Progress Publishers, (1954), p.4.
27. Particularly in: Marx K, *Wage Labour and Capital*, Whitefish, MT: Kessinger Publishing, (2004), pp.32–36.
28. For a more detailed history of the physical modelling of social sciences, see Ball, supra note 17, chapters 1–4.

reasoning to decisions involving other players, and therefore trying to analyse strategic situations in order to attribute potential outcomes to each decision.[29]

While economics and game theory are indicative of the possibility of social mathematical modelling, it is perhaps the very existence of these disciplines that is to blame for the prevalence of the science wars. There is something distasteful about reducing human decisions to basic binary choices between favourable and unfavourable outcomes, as though human beings were machines with little rational choice in these decisions. Implicit in the physical modelling of social interactions described since the time of Hobbes lies the presumption that humans make predictable choices, that society is to an extent deterministic, and that history is nothing more than a collection of dialectic points and counterpoints. Looking at the science in this way, there is little wonder that some in the social sciences have rebelled against such a reductionist view of human beings. In the words of documentary maker Adam Curtis, game theory and other similar mathematical explanations of social phenomena offers us a "simplistic view of human beings as self-seeking, almost robotic creatures".[30]

OBJECTIVES

Despite objections to the idea that human affairs can be the subject of statistical predictive analysis, this book follows a line of thought similar to that which inspired some of the philosophers and scientists mentioned above. It is one of the starting premises of the present work that several social phenomena follow certain predictable patterns that can be quantified and accurately described using mathematical tools. If such assumption is warranted, as I believe that it is, then such predictive and descriptive tools could be very useful to the law in its efforts to regulate human affairs in a much more efficient manner. This book, then, starts with a general statement: that regulators

29. Davis MD, *Game Theory: A Nontechnical Introduction,* Rev. ed, London: Dover Publications, Constable (1997), pp.3–9.
30. Curtis A, *The Trap – What Happened to our Dream of Freedom*, BBC (2007), Episode 3, 0:17.

should try, wherever possible, to use the physical methodological tools presently available in order to draft better legislation. While such an assertion may be applied to the law in general, this work will concentrate on the much narrower area of Internet regulation and the science of complex networks.

The Internet is the subject of this book not only because it is my main area of research, but also because –without over-emphasising the importance of the Internet to everyday life[31]– one cannot deny that the growth and popularisation of the global communications network has had a tremendous impact on the way in which we interact with one another. The Internet is, however, just one of many interactive networks. One way of looking at the complex and chaotic nature of society is to see it as a collection of different nodes of interaction. Humans are constantly surrounded by networks: the social network, the financial network, the transport network, the telecommunications network and even the network of our own bodies. Understanding how these systems operate and interact with one another has been the realm of physicists, economists, biologists and mathematicians. Until recently, the study of networks has been mainly theoretical and academic, because it is difficult to gather data about large and complex systems that is sufficiently reliable to support proper empirical application. In recent years, though, the Internet has given researchers the opportunity to study and test the mathematical descriptions of these vast complex systems. The growth rate and structure of cyberspace has allowed researchers to map and test several previously unproven theories about how links and hubs within networks interact with one another. The Web now provides the means with which to test the organisational structures, architecture and growth of networks, and even permits some limited prediction about their behaviour, strengths and vulnerabilities.

The main objective of this book is first and foremost to serve as an introduction to the wider legal audience to some of the theories of complexity and networks. The second objective is more ambitious. By looking at the application of complexity theory and

31. What I call the "Internet has Changed Everything" fallacy.

network science in various areas of Internet regulation, it is hoped that there will be enough evidence to postulate a theory of Internet regulation based on network science.

To achieve these two goals, Chapter 2 will look in detail at the science of complex networks to set the stage for the legal and regulatory arguments to follow. With the increase in reliability of the descriptive (and sometimes predictive) nature of network science, a logical next step for legal scholars is to look at the legal implications of the characteristics of networks. Chapter 3 highlights the efforts of academics and practitioners who have started to find potential uses for network science tools. Chapter 4 takes this idea further, and explores how network theory can shape Internet regulation.

The following chapters will analyse the potential for application of the tools described in the previous chapters, applying complexity theory to specific areas of study related to Internet Law. Chapter 5 deals with the subject of copyright in the digital world. Chapter 6 explores the issue of peer-production and user-generated content using network science as an analytical framework. Chapter 7 finishes the evidence section of the work by studying the impact of network architecture in the field of cybercrime, and asks whether the existing architecture hinders or assists efforts to tackle those problems.

It is clear that these are very disparate areas of study. It is not the intention of this book to be overreaching in its scope, although I am mindful that it covers a lot of ground and attempts to study and describe some disciplines that fall outside of my intellectual comfort zone. While the focus of the work is the Internet, its applications may extend beyond mere electronic bits. Without trying to be over-ambitious, it is my strong belief that legal scholarship has been neglectful in that it has been slow to respond to the wealth of research into complexity. That is not to say that there has been no legal research on the topic, but it would seem that lawyers, legislators and policy-makers are reluctant to consider technical solutions to legal problems. It is hoped then that this work will serve as a stepping stone that will lead to new interest in some of the theories that I describe.

SOME NOTES ON METHODOLOGY

As stated, this book has one overriding purpose, and that is to serve as an introduction to legal audiences to some of the topics explored by complexity theory and network science. I am painfully aware that this implies a need to explain concepts of physics and mathematics to audiences who may have no training in either. When writing *A Brief History of Time*, Stephen Hawking remarked that an editor had warned him that the inclusion of any equation would potentially halve the number of readers. Following that advice, this book will attempt to use non-mathematical explanations of the many concepts involved. This compromise is an attempt to inspire the legal reader to consider research that would otherwise be ignored because of the maths. The source material has in all cases been carefully cited to enable interested readers to access the original, replete with accompanying equations.

As the work purports to explain interdisciplinary studies, I am also conscious that in some instances I may have failed to convey the theories adequately. In those circumstances, the fault is solely mine.

2. The Science of Complex Networks

> Out of intense complexities intense simplicities emerge.
> Winston Churchill[1]

Kevin Bacon is in many ways an unremarkable movie star. From his cultural breakthrough in *Footloose*, to some of his forgettable roles in several 1990s romantic comedies, he has enjoyed critical success in films such as *Frost/Nixon*, *Apollo 13* and *JFK*. Nonetheless, he became part of Internet history as one of the first online memes when in 1994 he became the subject of the "Six Degrees of Kevin Bacon" game.[2] The game consists of trying to tie any randomly chosen actor to Kevin Bacon in less than six steps; the fewer steps the better. The origin of the meme is disputed, it could be because of his prolific acting career, or it could be that his name rhymes with separation, but the truth is that when one Usenet post made the claim that Bacon was the centre of the movie universe, and went on to try to prove it,[3] the Kevin Bacon game was born. The rest, as they say, is history.

The interesting thing about the Kevin Bacon game is that it serves to demonstrate a branch of studies into networks and complexity known as the small world phenomenon, which will be covered in detail later. A seemingly anodyne Internet meme has spawned a number of papers in reputable publications which describe the game, and go further into describing the phenomenon. It is through some of this research that we find that the network of actors is small enough that it ensures there usually will not be more than four

1. Churchill W, *The World Crisis, 1911–1918*, Nel Mentor ed, London: New English Library (1968), p75.
2. A version can be found here: http://www.thekevinbacongame.com/.
3. See the original thread at: http://tinyurl.com/bfg9mr.

connections between any given thespian;[4] we also learn that there are actors who are more connected than Kevin Bacon, such as Rod Steiger, Martin Sheen and Christopher Lee;[5] or that the short paths between actors characteristic of the Bacon game can be seen throughout other social clusters.[6]

The "Six Degrees of Kevin Bacon" is just a popular culture application of the wealth of research going into networks and complexity that has been experienced in the last decade. At their most basic level, several physical and social systems can be viewed through the study of links, nodes and hubs that constitute them. Complex network theory looks at how these operate, and offers valuable descriptive insights into their inner workings and development. Be it a relatively small social network such as that made up of screen actors and actresses, or vast computer networks such as the Internet, researchers have been finding some common denominators that help to analyse the behaviour of clusters. This chapter describes those studies and theories relevant to the rest of the book.

1. THE NETWORK SCIENCE REVOLUTION

In common parlance the word "network" is used to describe all sorts of phenomena where there is an interconnected plurality of individual elements. Therefore, we have telecommunication networks, social networks, transport networks, power networks, broadcasting networks, etc. In its daily usage, networks are consequently defined as "any netlike or complex system or collection of interrelated things".[7] While this is an adequate description of what networks are, the common usage of the word acquires a more precise meaning when looked at from a scientific standpoint. Mark Buchanan defines the scientific meaning of networks thus:

4. Adamic LA, "The Small World Web", 1696 *Lecture Notes in Computer Science* 443 (1999), p.444.
5. Durrett R, *Random Graph Dynamics*, Cambridge: Cambridge University Press (2007), p.7.
6. Gray E et al, "Trust Propagation in Small Worlds", 2692 *Lecture Notes in Computer Science* 239 (2003), pp.241–243.
7. *Oxford English Dictionary*, "Network", 2nd Edition (1989).

> The study of networks is part of the general area of science known as complexity theory. In an abstract sense, any collection of interacting parts –from atoms and molecules to bacteria, pedestrians, traders on a stock market floor, and even nations– represents a kind of substance. Regardless of what it is made of. That substance satisfies certain laws of form, the discovery of which is the aim of complexity theory.[8]

The understanding of how networks operate and interact with one another has been studied by physicists, economists and mathematicians for centuries. The birth of modern network theory can be traced to what is known as graph theory. In 1736, mathematician Leonard Euler published a classic paper answering what was known as the Königsberg bridge problem, which answered negatively the question of whether one could cross across the seven bridges of the Prussian city of Königsberg without having to cross the same bridge twice (Figure 2.1).[9] By applying a mathematical solution to this seemingly mundane problem, Euler established the methodological basis for the study of networks. The basis of the systematic study of networks is that at their basest form, they consist of individual elements known as nodes (or vertices), which connect to one another through links (or edges), typically in pairwise fashion, but they can also be unidirectional.[10] Graph theory can be used to chart paths through edges and vertices within any given network in similar fashion to that explained by Euler. Graph theory also provides the common convention to represent networks.[11]

---

8. Buchanan M, *Small World: Uncovering Nature's Hidden Networks*, London: Phoenix (2003), p.10.
9. Euler L, "Seven Bridges of Königsberg", in Newman JR (ed), *The World of Mathematics*, Vol. 1, Mineola, NY: Courier Dover Publications, (2000), pp.573–580.
10. Newman MEJ, Barabási A-L and Watts DJ, *The Structure and Dynamics of Networks*, Princeton, NJ, Oxford: Princeton University Press (2006), pp.2–3.
11. Ibid.

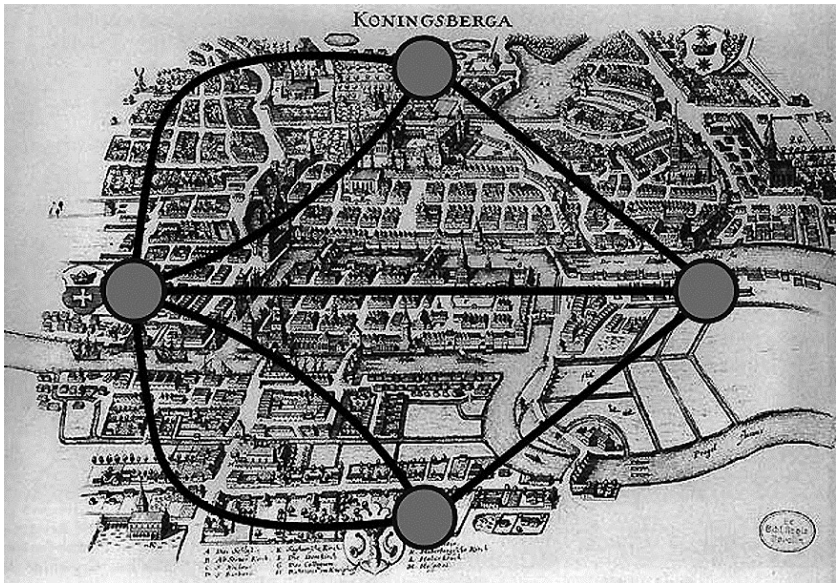


Figure 2.1 Graphical representation of the Königsberg bridge problem[12]

While the descriptive power of graph theory offered a powerful tool for mathematicians, its adoption to describe other networks was slow because its application was limited to static events. Eventually, scientists in other areas started to realise that one could look at several types of complex interactions using graph theory. One way of looking at it is to take the graphical representation of a travel across bridges, and replace it with the way in which information spreads through a social group, and then one can begin to see how graph theory describes other sorts of other interactive systems consisting of individual elements.[13]

However, charting static networks such as transportation hubs is one thing, but trying to chart random and dynamic networks involved levels of complexity that required a new frame of reference because the nodes and links are in constant movement. Using Euler's bridges again, it is relatively easy to create graphs that represent the possible

12. The seven bridges of Königsberg with superimposed graph solution to the problem. Created by the author from: http://commons.wikimedia.org/wiki/File:Image-Koenigsberg _Map_by_Merian-Erben_1652.jpg (original in the public domain).
13. Newman, supra note 10.

paths through static landmarks. But what happens if one is trying to create a graph that represents how information travels through dynamic networks? Take, for example, how a piece of gossip travels through a dinner party. Using each person as a node in the network, and linking who spoke with whom, one could construct a graph, but how would it be possible to chart whether or not the information was passed during any given exchange? If in this party A and B do not talk to each other, but the gossip eventually travels to B, is it possible to determine the path that the information took? (Figure 2.2).

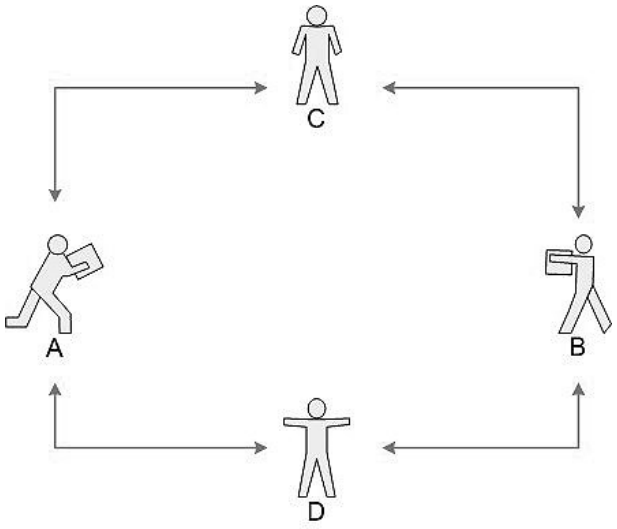


Figure 2.2 Random spread of information

In 1951 biophysicists Ray Solomonoff and Anatol Rapoport noticed the problem presented by dynamic networks when trying to chart data in biological systems, such as neurons and epidemics.[14] They found that these networks require a different set of analysis, but they just postulated the problem, they did not formulate solutions. By 1960, Hungarian mathematicians Paul Erdős and Alfréd Rényi had established a mathematical solution that accounted for some dynamic interactions by assigning random paths to the

14. Solomonoff R and Rapoport A, "Connectivity of Random Nets", 13 *Bulletin of Mathematical Biophysics* 107 (1951).

information;[15] this means that they would assign random number of connections to the nodes in a network. Going back to our gossip example, it does not really matter how the information gets from A to B as long as we know that the information goes through some of the intervening connectors; for explanatory purposes it is possible to assign a random path to the information, even if this is neither precise nor accurate – what matters is the end result. Erdős and Rényi's solution allowed the study of large-scale complex and dynamic networks, and it facilitated the further spread of graph theory as a useful model to analyse networks that had remained outside the grasp of graph theory. For example, models were presented that could try to chart information in social networks, or attempted to model social interactions.[16] For example, a study in 1978 tried to answer the question of how many people exert influence over others with whom they are in contact, and while the authors complained that they had generated more questions than they answered, they were able to produce valuable models of influence networks.[17] Thanks to the analytical tools provided by random graphs, network theory had grown into a veritable branch of economics and sociology, and had come of age.

It would be easy to overestimate the importance of network theory in the real world, but its importance has been continuously increasing. Once it migrated from the realm of mere mathematics to that of social studies, the application of graph theory to random networks had revolutionised the potential study of several dissimilar disciplines. The work of researchers like Erdős and Rényi had allowed the creation of a new branch of study that would cement theoretical principles for what was to become the modern discipline of network theory.

It is possible to imagine that if things had remained as they were, network theory may have remained an academic oddity. However, recent years have seen an explosion of

15. Erdós P and Rényi A, "On the Evolution of Random Graphs", 6 *Bulletin of the Institute of International Statistics* 261 (1961).
16. Rapoport A and Horvarth V, "A Study of a Large Sociogram" 6 *Behavioral Science* 279 (1961).
17. Sola-Pool I and Kochen M, "Contacts and Influence", 1 *Social Networks* 5 (1978).

research in the topic, prompting the creation of what some call the "new" science of networks. The aim of this field of study is explained thus:

> We argue that the science of networks that has been taking shape in the last few years is distinguished from preceding work on network in three important ways: (1) by focusing on the properties of real-world networks, it is concerned with empirical as well as theoretical questions; (2) it frequently takes the view that networks are not static, but evolve in time according to various dynamic rules; and (3) it aims, ultimately at least, to understand networks into just as topological objects, but also as the framework upon which distributed dynamical systems are built.[18]

This is a crucial point. Network science is not only a theoretical approach to complex systems, but it is concerned with practical application of the theory. One of the main events that have prompted the explosion of research into networks is the advent of the World Wide Web (WWW). There is little doubt that the Internet has given scientists the opportunity to study and test several of the pre-existing mathematical models of complex networks.[19] Although the Web is composed of billions of pages, its fast growth-rate and international reach allows researchers to map and examine several ideas about how networks interact. With a combination of the characteristics of online hyper-linking, and the help of spiders and web crawlers,[20] researchers have the means to test the organisational structures of the architecture and behaviour of networks.

Much of the current interest in networks can be traced back to a series of popular science books dedicated to publicising the latest developments in this area of research. Titles of note are *Linked* by Albert-Laszlo Barabási,[21] *The Tipping Point* by Malcom

---

18. Newman, supra note 10, p.4.
19. See for example: Broder A et al, "Graph Structure in the Web", 33 *Computer Networks* 30 (2000); Faloutsos M, Faloutsos P and Faloutsos C, "On Power-Law Relationships of the Internet Topology", 29 *Computer Communications Review* 251 (1999).
20. A web crawler is a computer program that browses the Internet in an automated and predetermined manner. See: Brin S and Page L, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", 30(1) *Computer Networks and ISDN Systems* 107 (1998).
21. Barabási A-L, *Linked: The New Science of Networks*, Cambridge MA: Perseus Pub. (2002).

Gladwell,[22] *Critical Mass* by Philip Ball[23] and *Six Degrees* by Duncan J Watts.[24] These "pop science" credentials could make those unfamiliar with the literature suspicious about the validity and reliability of network theories,[25] but this scepticism would be misplaced, as most of these books have sound peer-reviewed research behind them, and in most instances they have been written by the primary investigators themselves.

Network theory makes several conclusions and predictions that arise from empirical research and theoretical analysis. Some of these are more relevant to the present book than others; the ones that will be covered in one form or another later on will be described in more detail in the following sections.

2. NETWORK SCIENCE

2.1 Power laws

The modern understanding of networks begins with the study of statistical phenomena called power laws. A power law is a mathematical expression that happens "when the probability of measuring a particular value of some quantity varies inversely as a power of that value".[26] In other words, power laws are a mathematical concept that describes the divergence in the predictable and average value of an observable fact.

22. Gladwell M, *The Tipping Point: How Little Things Can Make a Big Difference*, London: Abacus (2002).
23. Ball P, *Critical Mass: How One Thing Leads to Another*, London: Arrow Books (2004).
24. Watts DJ, *Six Degrees: The Science of a Connected Age*, London: Vintage (2004).
25. It should be noted that network theory should not be confused with actor-network theory, see: Latour B, *Reassembling the Social: An Introduction to Actor-Network-Theory,* Oxford: Oxford University Press (2005). McLuhan also has something to say about networks, and is often cited as the father of network theory. See: Levinson P, *Digital McLuhan: A Guide to the Information Millennium*, London: Routledge (2001), pp.187–200. This work does not deal with these approaches.
26. See: Newman MEJ, "Power Laws, Pareto Distributions and Zipf's Law", 46:5 *Contemporary Physics* 323 (2005), p.323.

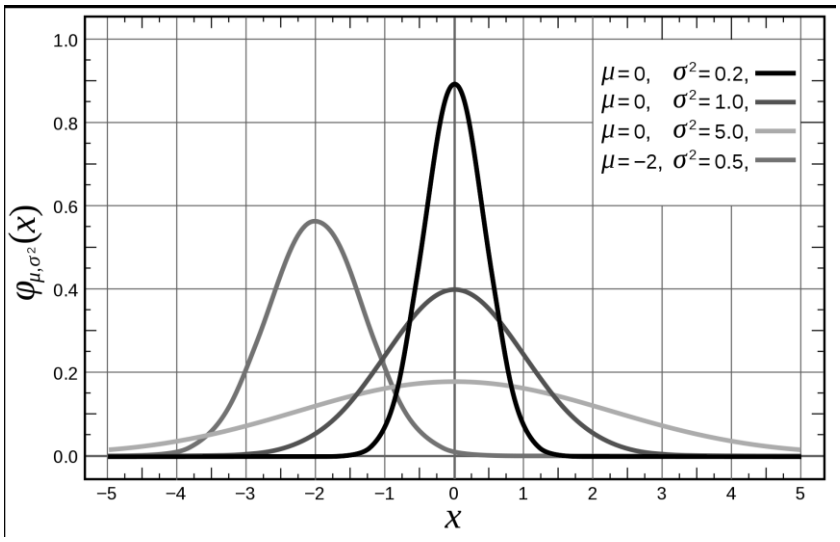


Figure 2.3 A selection of normal distribution probability curves[27]

In statistics, the normal distribution (also known as Gaussian distribution) is one where variables tend to concentrate along the average.[28] When plotting the number of occurrences along an X Y graph, this clustering towards the middle tends to produce a distinctive bell-shaped form because in normal distributions the largest number of instances is average (Figure 2.3). Most people are average height, although there are small numbers of both very short and very tall people; charting such distribution will provide a bell-shaped curve.[29] Power law distributions do not follow the normal trend; in them we find that there are a few remarkable occurrences that account for a very large number of instances of the studied event. Because of this, a power law distribution does not have a peak in the middle; a small number of occurrences account for a large part of the overall area of the chart, while given instances of an event tend to drop off sharply, which indicates the increased likelihood of extreme occurrences.[30]

---

27. From Wikipedia (released under public domain dedication), http://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg.
28. Weisstein EW, "Normal Distribution", *MathWorld* (2007), http://mathworld.wolfram.com/NormalDistribution.html.
29. Stigler SM, *Statistics on the Table*, Boston: Harvard University Press (1999), chapter 22.
30. Ball, supra note 23, p.295

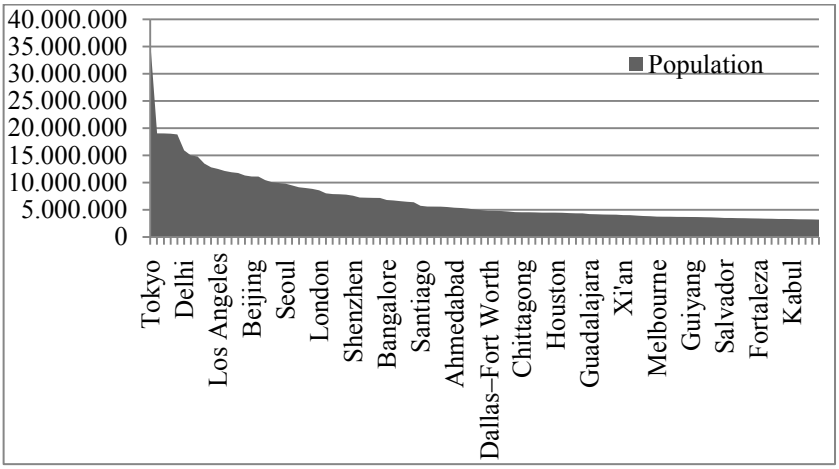


Figure 2.4 Power law distribution of city populations[31]

An example of power law distributions can be found in city populations. If we are counting all of the people living in cities around the world, we will soon discover that megalopolis like Tokyo, Mexico City, New York and Sao Paulo account for a disproportionate amount of the total city inhabitants. These cities generate tell-tale spikes in the data, accompanied by a long tail of smaller populations (Figure 2.4).

Power laws are useful statistical tools because not only do they serve to display distributions using a chart as displayed above, but they also provide the exponential factor with which the next given occurrence in a series either grows or decreases. Let us go back to city sizes in order to illustrate. In the United States there is a wide divergence in city size from the largest to the smallest; for example, Newman calculates that New York is 150,000 times larger than the smallest city.[32] A chart of city sizes would produce the characteristic graph displayed above. However, a power law also displays a constant exponential increase (or decrease depending on how you look at it) of one city to the next. This means that there is a constant rate in the way city sizes are distributed, so if

31. Wikipedia, "List of Urban Agglomerations by Population", http://en.wikipedia.org/wiki/World%27s_largest_urban_agglomerations.
32. Newman, supra note 26, p.324.

you knew the size of a city, you could make estimates of the size of the ones above and underneath it in a chart. When this is displayed as a logarithmic histogram, the end result is roughly a straight line, which is also characteristic of power laws (Figure 2.5).

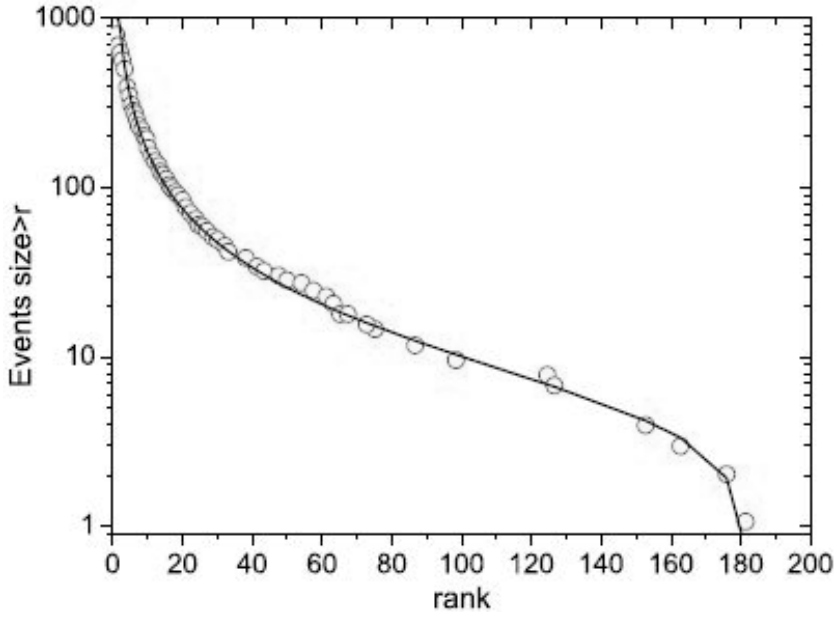


Figure 2.5 Logarithmic representation of power law in US cities[33]

It may be surprising that power laws seem to be found in all sorts of situations, from biological systems[34] to human mobility patterns.[35] Other places where these networks have been found are, according to Newman:

> In addition to city populations, the sizes of earthquakes, moon craters, solar flares, computer files and wars, the frequency of use of words in any human language, the frequency of occurrence of personal names in most cultures, the numbers of papers scientists write, the number of citations received by papers, the number of hits on web pages, the sales of books, music recordings and almost every other branded

33. Ibid.
34. Jeong H et al, "The Large-Scale Organization of Metabolic Networks", 407 *Nature* 651–654 (2000).
35. González MC, Hidalgo CA and Barabási A-L, "Understanding Individual Human Mobility Patterns", 453 *Nature* 779 (2008).

commodity, the numbers of species in biological taxa, people's annual incomes and a host of other variables all follow power-law distributions.[36]

While power laws are remarkable on their own merit, their presence is usually a good indication that we are faced with a specific type of complex system. As there is no law of nature that requires such an astounding correlation between completely disparate phenomena as earthquakes and web pages, power laws tell us that the systems that display them are responding to similar stimuli that shape them into predictable distribution curves. The apparent determinism occurs because of similar structural circumstances in all of the studied cases that display power laws. In other words, a power law distribution can tell us a lot about a specific system, because to display a power law, the system must behave in certain ways for it to appear. Average human height is not a power law; the number of connections in our brains is not a power law. But the distribution of proteins in some species display power laws,[37] as well as the protein interactions with viruses,[38] as well as the statistical significance of gene expressions.[39] This hints at a significant element in the study of power laws: if we understand how they work, we may be able to predict their appearance.

2.2 Scale-free networks

When applied to complex systems, power law distributions result in what is known as scale-free networks. In a normal distribution, there is little or no room for results that are considerably above and below the norm. To reuse the previous example regarding human average height, in any chart that displays people's heights in any given population one will expect to find that most people are average, with deviations towards both ends, thus forming a bell-shaped histogram. However, if heights behaved in a scale-

36. Newman, supra note 26, p.325.
37. Giot L et al, "A Protein Interaction Map of Drosophila Melanogaster", 302:5651 *Science* 1727 (2003).
38. Uetz P et al, "Herpesviral Protein Networks and Their Interaction with the Human Proteome", 311:5758 *Science* 239 (2006).
39. Ueda HR et al, "Universality and Flexibility in Gene Expression from Bacteria to Human", 101:11 *Proceedings of the National Academy of Sciences* 3765 (2004).

free manner, most people would be average height, while there would be some 30–50 metre giants walking around, and from time to time you could even encounter a person measuring hundreds of metres.[40]

It is called scale-free because the same distribution of relationships exists at any scale (forming a power law). If one was to look at some of the node and link structure in a scale-free network, then one would find the same degree of distribution of links and nodes. If we look at any random network and plot the links between nodes, and we isolate a small part of the network, no discernible pattern would be present. However, scale-free networks maintain the distribution of nodes and links at whatever level we want to look at. So, if the network is organised around hubs with certain number of connections, then it does not matter if we look at a few or at many nodes, this same degree of distribution will be present throughout.[41] This is akin to the concept of self-similarity where the system is exactly or approximately similar to a part of itself. This occurs in Mandelbrot sets[42] and other fractal topographies.

Power laws and scale-free topologies apply to large-scale complex systems in general, and networks specifically.[43] As stated earlier, networks are composed of nodes (vertices) and links (edges). Large-scale networks also have a third element, hubs, which are collections or clusters of nodes.[44] In a normal network distribution which displays a random topology, we would expect to find that nodes are distributed in an average manner, some with more links, and some with fewer links, which can be described through a typical random histogram. In a scale-free network, the vast majority of nodes and hubs have an average or small number of links, while very few hubs will have an exceptionally large number of links, forming super-nodes, or even super-hubs (Figure

40. Barabási, supra note 21, pp.67–69.
41. Newman, supra note 10, p.335.
42. Mandelbrot B, "How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension", 156:3775 *Science* 636 (1967).
43. Ravasz E and Barabási A-L, "Hierarchical Organization in Complex Networks", 67 *Physical Review E* 026112 (2003), p.1.
44. Ibid.

2.6).[45] When one reproduces these networks using graph theory representations, they also display very characteristic features. Random graphs tend to be chaotic, while scale-free graphs are organised around the hubs.

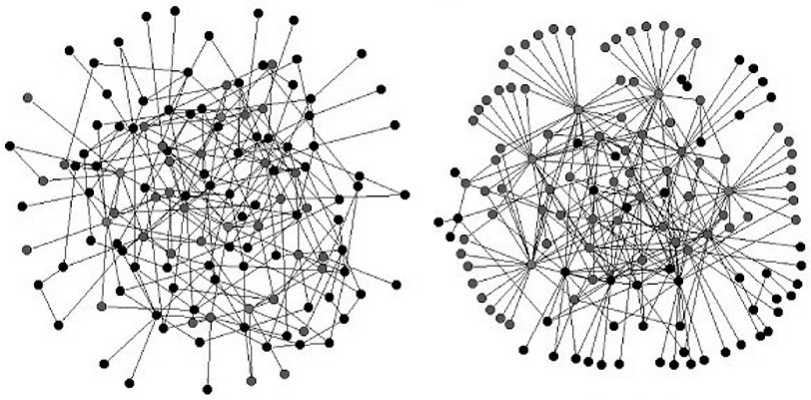


Figure 2.6 Random (left) and scale-free network (right)[46]

This way of looking at networks is particularly useful when analysing a large system such as the Internet. As mentioned earlier, the Web lends itself to the study of networks because of the potential ease with which it is possible to analyse link structure through search engines and autonomous agents. It is hardly surprising then that the Internet has been at the forefront of the resurgence in interest in graph theory and complex networks. There has been a wealth of innovative and informative research into the way in which the Internet works,[47] and its architecture is now understood enough to claim that it represents many of the inherent characteristics of scale-free networks and, as a result, it can be said that it responds to power laws. The topology of the Internet lends itself easily

45. Barabási, supra note 21, p.69–72.
46. Albert R, Jeong H and Barabási A-L, "Error and Attack Tolerance in Complex Networks", 406 *Nature* 378 (2000). Reproduced with permission.
47. For example, see: Dezso Z et al, "General Methods of Statistical Physics – Dynamics of Information Access on the Web", 73:6 *Physical Review. E* 69 (2006); Yook S-H, Jeong H and Barabási A-L, "Modelling the Internet's Large-Scale Topology", 99:21 *Proceedings of the National Academy of Sciences* 5 (2002).

as a ready-made tool for measuring connectedness. Spiders and other autonomous agents can be programmed to trawl the Web in order to gather information about its constituent pages, sites and links. This has allowed researchers to confirm the features of the Internet and understand its underlying architecture with an amazing degree of certainty.[48] Some of these features will be revisited later.

One would expect that a large network such as the Internet might exhibit random features instead of power laws. However, looking at how the Internet is organised, researchers have found that it exhibits scale-free characteristics in all of its components – namely page visits, incoming links, number of pages viewed on each visit, time spent on a site, popularity and architectural structure.[49] This predictability means that power laws are experienced and expected at all levels of granularity, whether one is looking at tens of thousands of pages, or just a hundred. Huberman comments that:

> The fact that the number of pages per site, and also the number of links per site, is distributed according to a power law is a universal feature of the Web. It holds throughout the World Wide Web, irrespective of the type of sites that one considers, from the smallest to the largest, and regardless of the nature of the site. The appearance of such a strong regularity out of a seeming random process is quite striking, and point to some kind of universal mechanism that not only underlies the growth of the Web, but also produces a power law distribution of its characteristics.[50]

This has allowed the charting of certain laws of the Internet: amidst the seemingly chaotic nature of the Internet, a hidden regularity emerges in every studied pattern. For example, websites under a domain seem to respond to power laws in the way in which pages are visited. The hub tends to be the home page, and subsequent links from the main site tend to decrease markedly into a power law distribution.[51] Similarly, web site popularity displays considerably few highly visible pages, with sharp drop-offs into a

48. For more about this, see: Huberman BA, *The Laws of the Web: Patterns in the Ecology of Information*, Cambridge, MA: MIT Press (2001), p.30.
49. Ibid, p.25.
50. Ibid, pp.29–30.
51. Ibid, p.30.

long tail of less visited sites.[52] The resulting clustering tends to produce an ecology dominated by hubs and super-hubs that act as the glue that binds and controls web traffic. This is why the Internet is not a random space, as the likelihood for an average user to visit a website responds to power laws.[53]

One of the main features of the Internet is that its growth responds to the expected accumulation of links, which is one of the trademarks of scale-free networks. Few websites accumulate staggering numbers of links, while the vast majority of sites have fewer links, which constitute a textbook example of a power law.[54] Not only is there a power law at work in Cyberspace, but the rate of accumulation of sites responds to how long they have been accumulating links, which serves to confirm its scale-free architecture.[55] This can be seen in the manner in which websites like Google, Bing and Yahoo act as hubs in the Web landscape.

2.3 Pareto distributions and Zipf laws

Another relevant feature of network science, and in particular with regards to power laws, is the existence of what is known as Pareto distributions,[56] which is a term used to describe large inequalities in data where most of the distribution is concentrated in a relatively small portion of a graph (Figure 2.7).

52. Ibid, pp.47–49.
53. Ibid, pp.23–25.
54. Albert R, Jeong H and Barabási A-L, "Diameter of the World Wide Web", 401 *Nature* 130–131 (1999).
55. Yook, supra note 47.
56. Reed WJ, "The Pareto, Zipf and Other Power Laws", 74(1) *Economics Letters* 15 (2001).

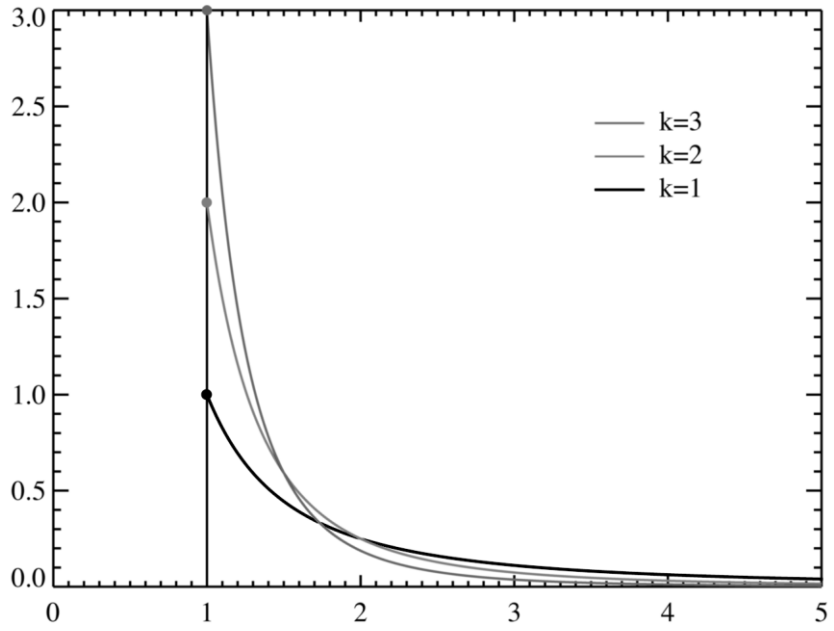


Figure 2.7 A typical Pareto distribution[57]

Italian economist Vilfredo Pareto was the first to establish this characteristic of power laws while studying property ownership. In 1906 he remarked that land ownership in Italy followed an 80/20 rule, that is, that 20 percent of the population owned 80 percent of the land.[58] Later, he made a similar discovery with regards to income distribution inequalities, remarking that roughly 2/3 of the wealth in Italy was concentrated in 1/3 of earners,[59] and also remarking that the number of top earners follow a power law. The 80/20 rule is remarkable because it has filtered through popular perception, so it is common to hear that 80 percent of the work is performed by 20 percent of the employees; or that 80 percent of the wealth is held by 20 percent of the population.[60]

In a strict economic sense, Pareto distributions are also known as the Pareto Principle, or Pareto's Law, and it is a function of size and rank, where the size is measured, for

57. Source: http://commons.wikimedia.org/wiki/File:Pareto_distributionPDF.png.
58. Pareto V, *Manual of Political Economy,* New York: Augustus M. Kelly Publishers (1971), p.45.
59. Mandelbrot B, "The Pareto-Lévy Law and the Distribution of Income", 1:2 *International Economic Review* 79 (1960).
60. Barabási, supra note 21, p.66.

example, by sales or wealth. The rank would be the percentage of the overall market share held by an individual. Pareto's Law predicts that in certain markets there will be a noticeable concentration of size and rank – in other words, fewer individuals will account for larger numbers of sales.[61] It is important to point out two things before continuing. Firstly, Pareto did not directly come up with this phenomenon, the name Pareto Law was first attributed to John Juran, a quality control engineer who first observed that Pareto's work on wealth applied to several other fields, and he is the one who described the Pareto Principle as the "vital few and trivial many".[62] Secondly, as Juran found out, Pareto's Law seems to be a universal law that is common to distribution of incomes, city sizes, prize returns on stock indices, meteor impacts and word frequencies.[63]

The implication of the existence of such regularity should be evident. There must be some self-organising principle in which certain disparate phenomena become organised in order to produce a large skew at the head of the chart. This cannot only be explained by selection bias or methodological similarity, there are just too many events that share these characteristics. Is nature deterministic at some basic level? We still do not know, although various economists have presented theories as to why such distributions are nearly universal.[64] What seems to be at work, however, is that whenever a network displays power laws, it will probably result in a plethora of other characteristics shared by scale-free topologies, of which Pareto inequalities seem to be one.

Zipf's law is a variation of Pareto distributions, named in honour of linguist George K. Zipf. While Pareto found a power law in income for top earners, he did not establish a specific rate for the invariance, but Zipf's law does this. Zipf was trying to put forward a specific view of society which stated that in any given social interaction people would

61. Giles DE, "Increasing Returns to Information in the US Popular Music Industry", 14:4 *Applied Economics Letters* 327 (2007).
62. Juran JM, "The Non-Pareto Principle: Mea Culpa", *Quality Progress* (1975), p.4.
63. Reed, supra note 56.
64. Ibid.

act in a manner that required minimal effort.[65] In order to support this observation, he made several empirical studies into various phenomena. In his most famous study, he found a power law in language when he discovered that in any given text corpus, a word's frequency is inversely proportional to the one next in rank.[66] So for example, in most English language texts the word "the" is most commonly used. Zipf provided evidence that in a studied corpus, "the" accounted for roughly 7 percent of the words, while the next in rank, "of", occurred half of that, and so on. In fact, only 135 words accounted for half of the studied corpus.[67] This is consistent with power laws.

There are two remarkable features of Zipf's law. One is that it is replicated in all sorts of other power law distributions, such as city sizes.[68] The other one is that Zipf's law is a common denominator of self-organised systems, where a chaotic environment becomes spontaneously ordered, a feature that will be dealt with in more detail in the last section.[69]

One possible explanation for the existence of Zipf's laws and Pareto distributions in large networks is what some researchers have termed "the rich get richer" effect.[70] As a network grows, popular nodes and hubs will continue to gather more links as time goes by; an effect that takes place because of the cumulative effect of the interaction between pre-existing links. The older a node is, the more likely it will be to have established links and to have been communicated to other nodes, while newer nodes will lack this advantage. This is caused by what is known as preferential attachment. The concept of preferential attachment in networks is a way to explain the way in which scale-free

---

65. Zipf GK, *Human Behavior and the Principle of Least Effort*, Cambridge MA: Addison-Wesley, (1949).
66. Ibid.
67. Wikipedia, "Zipf's Law", (2009), http://en.wikipedia.org/wiki/Zipf%27s_law
68. Ioannides Y and Overman HG, "Zipf's Law for Cities: An Empirical Examination", 33:2 *Regional Science and Urban Economics* 127 (2003). Another study replicating these findings is Rosen K and Resnick M, "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy", 8:2 *Journal of Urban Economics* 165 (1980).
69. Ball, supra note 23, pp.305–307.
70. Durham Y, Hirshleifer J and Smith VL, "Do the Rich Get Richer and the Poor Poorer? Experimental Tests of a Model of Power", 88:4 *The American Economic Review* 14 (1998).

networks grow, where nodes with previous connections are more likely to accumulate more links than newer nodes that do not have this advantage. For example, Newman[71] looked at two different scholarly collaboration networks, and measured the probability of a node acquiring new links as a function of its previous acquaintances and the number of previous collaborations. He found that there was a strong probability of the node acquiring links if it had both. Research into the development of the Internet bears out this effect; a study into the accumulation of links on any given site found that new nodes in the system were more likely to be attached to pre-existing nodes at a rate that responded to a power law.[72] Anyone familiar with web publishing will recognise this as anecdotally true.

However, the accumulation of links can lead to a collapse of node competition, where one node becomes the sole super-hub, a phenomenon known as the "winner-takes-all".[73] While this effect is rare, it responds to how similar complex systems act generally in physics, and specifically in gases, a phenomenon known as Einstein–Bose condensation. At normal temperatures gas atoms move and collide with one another at different speeds – the hotter the gas, the faster the atoms move, and vice versa. It is theoretically possible that at very low temperatures gases would stop moving completely, but this theoretical temperature is too low to happen naturally. Albert Einstein and Satyendranath Bose contributed separately to a framework that would allow gas condensation at higher temperatures, hence the name.[74] A very interesting finding from network theory is that the equations used to describe Einstein-Bose condensation in gases can be used to describe link accumulation in the World Wide Web,[75] which could serve as an explanation of the seemingly random runaway success of certain websites.

---

71. Newman MEJ, "Clustering and Preferential Attachment in Growing Networks", 64:2 *Physical Review E* 025102 (2001).
72. Krapivsky PL, Rodgers GJ and Redner S, "Degree Distributions of Growing Networks" 86(23) *Physical Review Letters* 5401–5404 (2001).
73. Barabási, supra note 21, pp.102.
74. Ibid, pp.97–100.
75. Bianconi G and Barabási A-L, "Bose–Einstein Condensation in Complex Networks", 86(24) *Physical Review Letter*s 5632-5635 (2001).

Pareto distributions and Zipf's laws are perhaps some of the most remarkable effects of the emergence of the science of networks, but they could even be taken as the precursors of psychohistory. Not only do they apply to widely diverging phenomena, they seem to be a set law as far as the Internet is concerned. Once these patterns about the World Wide Web are noticed, it becomes difficult to view complex networks in any other light.

2.4 Small worlds and social networks

The clustering of nodes present in scale-free networks described above explains one of the most publicised insights arising from the research into networks, and that is the phenomenon of small worlds, or the so-called six degrees of separation expounded by the Kevin Bacon game. This is the commonly-held knowledge that all of the people in the world are separated only by six connections from one another.

This belief originates from a study by psychologist Stanley Milgram, who tried to measure how many links there were between people in Kansas, Nebraska and one target in Massachusetts, which resulted in a surprisingly small number of intervening connectors.[76] While many letters did not reach their final destination, a total of 64 did, with an average number of 5.5 intervening links, hence the name "six degrees". Milgram had been inspired by some of the graph theory research conducted by Erdős and Rényi,[77] but his research was particularly informed by Sola-Pool and Kochen's aforementioned research into influence, which had left some unanswered questions about the length of social networks.[78]

While Milgram's experiment was limited both in execution and scope, it showcased one of the characteristics of social networks, and that is the importance of hubs to any complex system. The reason why there is a correlation between this hypothesis and scale-free systems is evident if one considers that there are certain hubs in social

76. See: Milgram S, "The Small World Problem", 2 *Psychology Today* 60–67 (1967).
77. Supra note 15.
78. Supra note 17.

networks that acquire more links than others. These hubs act as "connectors"[79] and, once a message has reached one of those, the chances are that it will offer a large number of links to other nodes in the system. This is highlighted in smaller social networks, such as the actor network in the Kevin Bacon game, or the scientific collaboration network. A lot of studies have been undertaken on the latter as it is easy to try to link two scientists using co-authored publications in scientific journals as a measure of connectedness.[80] An example involves Paul Erdős himself; in 1969 a paper suggested that Erdős had been so prolific that he could be used as a measure of academic author connectivity, hence establishing the Erdős number, which is the number of collaborators between any author and Erdős.[81]

The study of small worlds has been resurgent in recent years, which has coincided for obvious reasons with the growing interest in scale-free networks. Traditionally, small worlds can be defined as networks where the component vertices are clustered as to allow short paths between nodes.[82] Strogatz and Watts wanted to expand on this definition by testing whether other types of networks exhibited small world clustering between its components.[83] They first looked at the two most common graph models that were prevalent in literature at the time, a regular network with a steady number of connections, and a random network exhibiting complete disorder. They proposed that highly-clustered networks fell somewhere in between these two extremes (Figure 2.8).

79. Gladwell, supra note 22, pp. 34–64.
80. Newman MEJ, "The Structure of Scientific Collaboration Networks", 98:2 *Proceedings of the National Academy of Sciences of the United States of America* 6 (2001).
81. Goffman C, "And What Is Your Erdos Number?" 76:7 *The American Mathematical Monthly* 791 (1969).
82. Newman, supra note 10, p.286.
83. Watts DJ and Strogatz SH, "Collective Dynamics of 'Small-World' Networks", 393:6684 *Nature* 440 (1998).

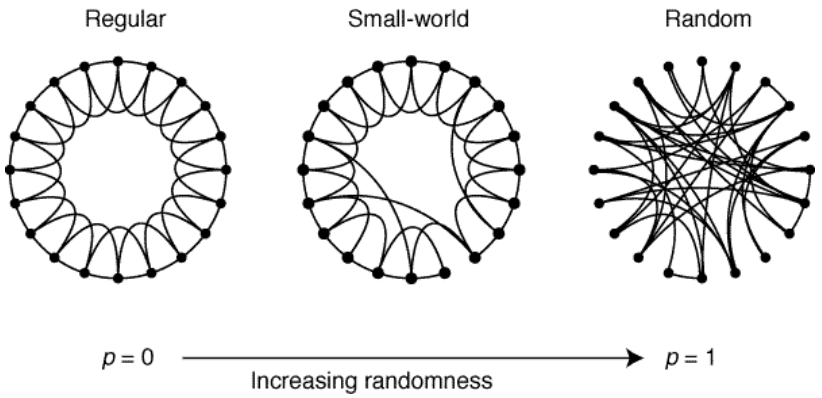


Figure 2.8 Small world network as compared to normal and random ones[84]

Watts and Strogatz set out to prove the model by testing it on three seemingly dissimilar networks. Firstly, they obtained information from the Internet Movie Database and assigned each actor as a node, then they looked at collaborations as links. Secondly, they looked at the Western power grid of the United States, the nodes in this network were the power stations and sub-stations, and the links the transmission lines. Thirdly, they looked at the neural network of the nematode worm *C. elgans*, the nodes were the neurons and the links were the synaptic connections between them. In each of these networks, they found that while the distance between nodes was similar to that encountered in random networks, the clustering coefficient was considerably higher than that which would be expected from any random number of connections, therefore proving the existence of small world networks.[85] This expanded the definition of a small world network as one where the distance between nodes expands logarithmically depending on the number of vertices in the system.[86] In other words, small worlds also display power laws.

While the study of small worlds may be interesting from an academic and social perspective, the question must be asked of whether they tell us anything about the real

84. Ibid, p.441.
85. Ibid. p.442.
86. Newman, supra note 10, p.286.

world. One critical reminder when looking at small world clustering in networks is that it is always important to know exactly what is being charted. While tenuous connections between collaborators can display interesting clusters between academic writers, the relevance of connectedness and clustering can be unearthed by first asking why it is that we study the connectivity between nodes in a network. The answer to this is that there are several circumstances where it is essential to learn the length of a pathway interconnecting individual vertices in a complex system. For example, biological networks such as food chains are very important to an organism's survival. What happens when we remove a species from the food chain? With anthropogenic extinction becoming a key issue at present, one team of researchers analysed what were the connecting paths between species, and surprisingly found that in nature most species are connected to one another by an average two degrees of separation, hinting at a more interconnected biological web than previously expected.[87] Similarly, research into food chains in the North Atlantic found that a catastrophic reduction of cod populations had a knock-on effect in 150 other marine species.[88] Small world clustering also serves to explain viral infections, and are being talked about as potential models of the spreading rate of highly-contagious epidemics.[89]

It is vital here to make a distinction about the type of analysis that is conducted within social networks, of which the small-world phenomenon is but one element. When we look at social networks from a network theory perspective, we are looking at two types of data, what Scott helpfully calls relational data and attribute data.[90] Social networks consist of individuals that interact with one another responding to social occasions, social meaning, individual motives, and cultural determinants. Relational data consists of the links themselves, "the contacts, ties and connections, the group attachments and

87. Williams RJ et al, "Two Degrees of Separation in Complex Food Webs", 99 *Proceedings of the National Academy of Sciences* 12913 (2002).
88. Buchanan, supra note 8, pp.148–151.
89. Carrat F et al, "A 'Small-World-Like' Model for Comparing Interventions Aimed at Preventing and Controlling Influenza Pandemics", 4 *BMC Medicine* 26 (2006).
90. Scott J, "Social Network Analysis", 22:1 *Sociology* 109 (1988).

meetings, which relate one agent to another and so cannot be reduced to the properties of the individual agents themselves".[91] The analysis of these relations is not concerned with motives and other cultural and social systems, and thus they lend themselves to study via network theory. Attribute data consists of the "attitudes, opinions and behaviour of agents",[92] and so lend itself to more traditional social science studies, such as economics, sociology, anthropology, etc. It is imperative to stress this point, because it must be remarked that the study of complexity in networks does not immediately erase the relevance of other areas of study. The study of links, pathways, the distribution of hubs in a social environment, and the number of intervening nodes required for information to travel from one node to another tell us some vital things about how social systems operate, but it does not erase the need of knowing why these things happen, or how the societies are organised one way or another.

It would be tempting to try to draw too many conclusions based on small world clustering in social networks. However useful the data is, it must be remembered that when looked at directly, social networks seem to be starkly divided by economic and ethnic sub-networks.[93] Nonetheless, there are several focal features of small worlds that make it potentially important for the subject of this book, namely that of Internet regulation. Firstly, small worlds are useful in measuring average path lengths in social networks, and particularly useful in charting the spread of information. This is significant to the analysis of online viral infections. Secondly, small worlds offer excellent tools with which one can analyse network architecture, which is a key feature of the regulation of the global network. Thirdly, small worlds could help explain the workings of vast networks within the Internet environment, such as criminal webs, copyright infringement applications, and other online features with a social component. All of these will be covered in detail in later chapters.

---

91. Ibid, p.110.
92. Ibid.
93. Kleinfeld J, "The Small World Problem", 39 *Society* 61-66 (2002).

2.5 Network resilience

There are two final characteristics of scale-free networks that are relevant to this work; those of robustness and cascading failures. First, scale-free networks are remarkably resilient and stable; that is, they tend to remain intact regardless of the removal of a node.[94] Strogatz explains that:

> …scale-free networks are resistant to random failures because a few hubs dominate their topology. Any node that fails probably has small degree (like most nodes) and so is expendable. The flip side is that such networks are vulnerable to deliberate attacks on the hubs. These intuitive ideas have been confirmed numerically, and analytically, by examining how the average path length and size of the giant component depend on the number and degree of the nodes removed.[95]

In other words, if one tries to attack a scale-free network randomly, the result will be that the attacked node will be unlikely to play any essential part in the way in which the network stays together. This is because hubs tend to be few, so the chances of hitting one randomly are very high. The Internet has proved to have inherited such robustness,[96] as virus attacks, and even Distributed Denial of Service (DDoS)[97] have not managed to bring down the entire network.

However, Strogatz also uncovers a potential vulnerability present in scale-free networks, which is that they are strong but not invulnerable. There are documented circumstances where scale-free systems have collapsed in spectacular fashion due to cascading failures. In 1996, a large blackout affected eleven states in the US and two Canadian provinces, which originated from the failure of one single line in Oregon.[98] Energy grids are typical examples of scale-free networks because they rely on a few key

---

94. Albert R, Jeong H and Barabási A-L, "Error and Attack Tolerance in Complex Networks", 406 *Nature* 378-382 (2000).
95. Strogatz S, "Exploring Complex Networks", 410 *Nature* 268 (2001), p.274.
96. Tu Y, "How Robust is the Internet?" 406 *Nature* 353 (2000).
97. A DDoS is an attack on a computer network by more than one system that causes loss of service to users, typically the loss of network connectivity and services by consuming the bandwidth of the victim. For more about DDoS, see: Edwards L, "Dawn of the Death of Distributed Denial of Service: How to Kill Zombies", 24:1 *Cardozo Law School Arts and Entertainment Law Journal* 23 (2006).
98. Barabási, supra note 21, p.119.

hubs in order to maintain distribution loads. If one of those hubs is removed, the entire system may collapse; an effect that spells the vulnerability of networks to random occurrences in hubs,[99] or even to targeted attacks against one.[100] This effect is often referred to as a cascading failure, because the removal of a hub will have knock-on effects on the nodes connected to it, and on the nodes connected to those, etc.

The relevance of robustness will become clearer later, but for now it is of great consequence to remark that this dual feature of scale-free networks offers one of the most interesting potentials for regulatory studies. Particularly, it could provide strategies for tackling illegal scale-free networks, such as P2P sharing sites. It could also provide tools to guard against large-scale hacking attacks against the Web's infrastructure.

3. COMPLEXITY AND SELF-ORGANISATION

3.1 Complexity

So far we have discussed some of the features of graph theory and network science to establish the framework for the later discussion into Internet regulation. There is a final branch of research that will be relevant, and while it can have a direct effect on networks and power laws, it can be classed as a different branch of study altogether, and these are the areas of complexity and self-organisation.

From reading some of the features of power laws and complex networks highlighted above, one cannot help but marvel at the order beneath the apparent complexity. This is noteworthy because there is clear evidence of a hidden order to seemingly random events, one that structures populations, websites, incomes, linguistics, biological organisms and all sort of unrelated complex systems. The fact that these events respond to a set of laws and principles cannot be a coincidence. One is perhaps tempted to re-

---

99. Moreno Y, Gomez JB and Pacheco AF, "Instability of Scale-Free Networks under Node-Breaking Avalanches" 58(4) *Europhysics Letters* 630–636 (2002).
100. Bollobás B and Riordan O, "Robustness and Vulnerability of Scale-Free Random Graphs", 1(1) *Internet Mathematics* 1–35 (2003).

examine Smith's Invisible Hand, and other such explanatory mechanisms to attempt to make sense of these findings. It is no coincidence that some have proposed Adam Smith as the first person who started the study of complexity in social phenomena.[101]

The systematic study of complexity, also known as complexity theory,[102] is a wide-ranging field encompassing mathematics, physics, chemistry, biology, economics, computer science and sociology.[103] Complexity can be defined as a large number of parts that interact to make up a whole which is independent of its environment.[104] Complexity theory consequently is the systematised study of such complex systems that attempts to find patterns in this complex behaviour. Anderson comments that:

> Modern complexity theory suggests that some systems with many interactions among highly differentiated parts can produce surprisingly simple predictable behaviour that is impossible to forecast though they feature simple laws and fewer actors. […] [N]ormal science shows how complex effects can be understood from simple laws; chaos theory demonstrate that simple laws have complicated, unpredictable consequences; and complex theory describes how complex causes can produce simple effects.[105]

Arguably, one of the most influential figures in modern complexity theory is biologist Stuart Kauffman, who can be credited as not only organising a revolution in biology with his study into self-organisation, but also is responsible for the manner in which his ideas have been transferred to the social sciences.[106] Kauffman initiated his study into complex systems while looking at genetic networks, and marvelled at the organising

---

101. Miller JH and Page SE, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, Princeton, NJ: Princeton University Press (2007), pp.4–5.
102. It is important to note that, although complexity theory is sometimes referred to as chaos theory, the latter can be described as one of its components. See: Alligood KT, *Chaos: An Introduction to Dynamical Systems*, New York: Springer-Verlag (1997).
103. The core fields of study in complexity theory are: complex-adaptive systems theory, non-linear dynamic systems theory, synergetics, and non-equilibrium thermodynamics. See: Goldstein J, "Emergence as a Construct: History and Issues", 1:1 *Emergence* 49 (1999), p.56.
104. Anderson P, "Complexity Theory and Organization Science", 10:3 *Organization Science* 216 (1999), p.216.
105. Ibid, p.217.
106. For example, he has been often invited to present at philosophy symposia, see: Kauffman SA, "The Sciences of Complexity and 'Origins of Order'", 2 *Proceedings of the Biennial Meeting of the Philosophy of Science Association, V Symposia and Invited Papers* 299 (1990).

interaction of genes. He asked a simple question, whether the organisation present in the genetic network was the result of spontaneous ordering, as opposed to the more gradual approach favoured until then.[107] When answering this question, Kauffman managed to revolutionise the study of complex systems.

Kauffman's theories explain organisation and complexity by looking at the way in which entities in a network respond to changes in neighbouring entities; changes that are eventually translated into spontaneous ordering of the overall system. Kauffman was puzzled by the way in which genes were able to influence one another within a dynamic network. Interestingly, this is a similar question to that posed by researchers into small worlds, who first noticed that there were patterns in the way in which people influence one another. Faced with nightmarishly complex dynamic systems where genetic interaction was not evident at first glance, he decided to make an assumption in order to study interactions; he postulated that genes would be regulated by two other random genes in the system.[108] While this is an artificial solution, it allows researchers to study interaction within a complex network much in the same way as one would study pathways of information in dynamic graphs, thus opening the door to the study of large complex systems. By assigning real numbers to large levels of complexity, Kauffman was able to measure fitness levels within networks. Fitness here should be understood in the strict biological sense, that is, it describes an organism's capability to reproduce. Kauffman proposed what is known as the NK model of fitness, where an organism has N number of genes, each with only two connections to randomly assigned genes, where K describes the level of complexity in a system.[109]

Kauffman's NK model allows the study of the reproductive success of genotypes by thinking of optimal reproductive states as mountains in a landscape. Imagining that any given genome is a landscape, Kauffman's model allocates fitness levels to different

---

107. Ibid, p.300.
108. Kauffman SA, "Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets", 22 *Journal of Theoretical* Biology 437 (1969).
109. Kauffman SA and Weinberger EW, "The NK model of rugged fitness landscapes and its application to maturation of the immune response", 141:2 *Journal of Theoretical Biology* 211 (1989).

states. Evolving organisms "climb" the mountain until they reach highly-stable points at the peak of the mountain; these stable environments allow faster reproduction of the genes, and produces what is known as fitness landscapes (Figure 2.9).[110]

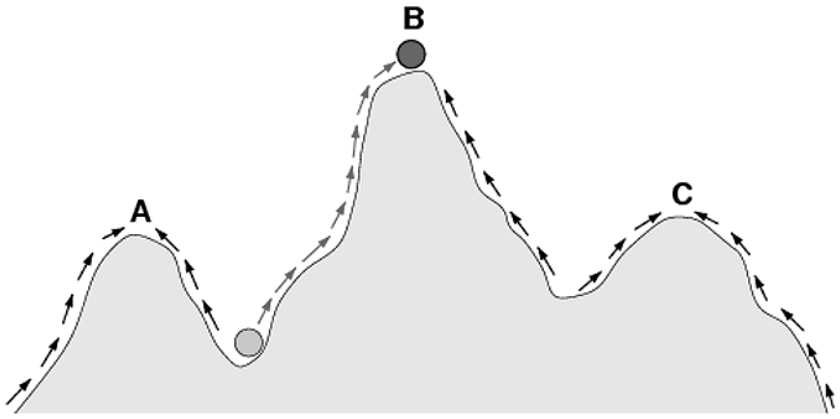


Figure 2.9 Fitness landscapes, where A, B and C describe fitness peaks[111]

By using fitness within a system as a measure of complexity and order, it is possible to extend the study of complexity to other fields. The idea of fitness levels and fitness landscapes has been influential with the examination of other networks. For example, Barabási uses a similar concept to analyse the interaction of nodes within complex networks such as the Internet. Using fitness as an analogy of an organism's capability to reproduce, he assigns fitness levels to web pages, therefore measuring a site's relative possibility to attract links.[112]

Another relevant area of study into complex systems is what is known as complex adaptive systems (CAS). Modern complexity theory studies two types of complex systems, predetermined systems, such as transport networks, and emergent systems, such as biological systems. In predetermined systems, there are steady connections

110. Kauffman SA, *At Home in the Universe: The Search for Laws of Self-Organization and Complexity,* Oxford: Oxford University Press (1995), pp.191–206.
111. From Wikipedia: http://en.wikipedia.org/wiki/File:Fitness-landscape-cartoon.png.
112. Barabási, supra note 21, pp.94–96.

between the elements in the system (much like Euler's bridges of Königsberg); whereas emergent systems consist of constantly interchanging pathways and interconnections. Complex adaptive systems fall into the former category, and can be defined thus:

> A Complex Adaptive System (CAS) is a dynamic network of many agents (which may represent cells, species, individuals, firms, nations) acting in parallel, constantly acting and reacting to what the other agents are doing. The control of a CAS tends to be highly dispersed and decentralized. If there is to be any coherent behavior in the system, it has to arise from competition and cooperation among the agents themselves. The overall behavior of the system is the result of a huge number of decisions made every moment by many individual agents.[113]

This is an important distinction to those systems that we have seen already. For example, scale-free networks are also decentralised and dynamic, but they tend to be more resilient and less responsive to patterns emerging from individual nodes. A telling characteristic of the presence of CAS is that while collective behaviour is crucial, small individual decisions may have a large effect on the overall system, what is usually illustrated in popular culture as the "butterfly effect".[114]

A good demonstration of CAS research within the social sciences is that of the standing ovation. Standing ovations are great ways to model complex adaptive systems, as they originate spontaneously in crowds of people. At some point one or two people get up, and there is an awkward moment where the crowd either follows, or it does not. Were one to make a mathematical model of standing ovations, one would have to take many things into consideration, such as the quality of the performance being such that it prompts people to get up and clap, the number of people who follow the lead, the topology of the auditorium, and similar considerations.[115] However one would like to make such a model, it would not really tell us much about how crowds behave, and what

---

113. Waldrop MM, *Complexity: The Emerging Science at the Edge of Order and Chaos*, New York: Penguin (1993), p.12.
114. Which receives its name from the idea that a butterfly flapping its wings may be the source of a hurricane; see: Boffetta G, Paladin G and Vulpiani A, "Strong Chaos without the Butterfly Effect in Dynamical Systems with Feedback", 29:10 *Journal of Physics A: Mathematical and General* 2291 (1996).
115. Miller and Page, supra note 101, pp.11–15.

the possibilities of getting a standing ovation are. However, by looking at the model using complexity models, it is more likely that one can reach a more accurate and useful description (and possibly even a prediction). Much in the same manner in which Kauffman assigns a value to gene connections, we can assign values to a random crowd. For example, by considering whether any friends of the performer are in the audience, and whether those have friends, and whether they are seated towards the front of the auditorium, one could begin to draw strong possibilities as these factors determine strongly whether one could turn an audience from one state (sitting) into another (standing). Think about your average opening night. People in the front seats are more likely to be friends of the performer because they have been allocated there; and the behaviour of the front row is more influential because they are more visible to the rest of the audience, so one could probably expect a standing ovation to be more likely during an opening night. A model of standing ovations that takes these factors into consideration will be more likely to produce accurate predictions of crowd management if one wanted to initiate a standing ovation. By following simple calculations into the shape of the auditorium, lines of sight, and number of people who react to one another, complexity modelling of social systems becomes a valuable analytical tool.[116] Similar research into crowd behaviour applies to other group phenomena, such as the Mexican wave.[117]

So, complex adaptive systems are a way to describe order, but also offer specific tools to predict the likelihood of a seemingly self-organising effect to occur out of complex situations.

3.2 Self-organisation

Self-organisation is a subset of complexity theory, just like graph theory and network science. While the science of self-organisation is relatively new, self-organisation itself

116. Ibid.
117. Farkas I, Helbing D and Vicsek T, "Social Behaviour: Mexican Waves in an Excitable Medium", 419 *Nature* 131 (2002).

has been written about by economists and philosophers for centuries, from Friedrich Engels[118] to John Stuart Mill.[119] Nonetheless, there have been several converging branches of study that have helped to shape our current understanding of complex systems.

Firstly, mathematician Claude Shannon developed what is known as information theory, which is a systematised method to quantify information.[120] Shannon was particularly interested in communication, and how information gets from one place to the other. He remarked that there needed to be a minimal unit to identify meaningful information, which he named a bit. Being able to measure information is useful, because Shannon was interested in finding the minimum number of bits required to transmit any given message. Shannon's information theory, together with other advances into the mathematical study of vast information, such as Kolmogorov complexity [121] and Turing's computational theory,[122] provide a strong analytical framework with which to analyse complex systems involving information.

Secondly, biologists have been at the forefront of the study of self-organisation due to the way in which biological systems organise themselves.[123] For example, ant colonies and beehives have been studied as some of the perfect examples of ordered complex systems; ant colonies' cemeteries and rubbish heaps are organised in a manner that optimises the distance from both the colony and each heap.[124] It is behaviour such as this that provides us with a working definition for self-organisation, not only in biological

---

118. Johnson S, *Emergence: The Connected Lives of Ants, Brains, Cities and Software*, London: Penguin (2002), pp.35–37.
119. Mill JS, *A System of Logic Ratiocinative and Inductive*, London: John W. Parker and Son (1872), p.371.
120. Shannon C E, "The Mathematical Theory of Communication", 27 *Bell System Technology Journal* 379 (1948).
121. Kolmogorov AN, "Combinatorial foundations of information theory and the calculus of probabilities", 38:4 *Russian Mathematical Surveys* 29 (1983).
122.Turing, AM, "On Computable Numbers, with an Application to the Entscheidungsproblem", 42:2 *Proceedings of the London Mathematical Society* 230 (1937).
123.For example, Hofstadter's self-assembling virus, Hofstadter DR, *Godel, Escher, Bach: An Eternal Golden Braid*, 20th-anniversary ed, London: Penguin (2000), p.542.
124. Johnson, supra note 118, pp.32–33.

systems, but that applies to other disciplines as well. Camazine et al define self-organisation as:

> …a process in which pattern at the global level of a system emerges solely from numerous interactions among the lower level components of the system. Moreover the rules specifying interactions among the system's components are executed using only local information, without reference to the global pattern. In short pattern is an emergent property of the system rather than being imposed on the system by an external ordering influence.[125]

In other words, self-organisation can be defined as any system that undergoes an organisation process due to internal elements present in the system, instead of responding to external output. In nature, Dynamic Systems (DS) theory is just another effort to explain self-organising behaviours.[126] In DS terms, self-organisation simply describes open systems that maintain themselves through the constant flow and dissipation of energy; chaotic systems where energy flows sometimes can adapt internally to form patterns that can be described as stable, yet not static.[127] Self-organisation as a result deals with ordering from within, and the order is a function of stability. This is where concepts such as Kauffman's fitness landscapes become useful.

The third branch that has been providing input into the understanding of self-organisation is the idea of emergent systems, which also forms part of complexity theory. Emergent systems display several characteristics attributed to the definition of self-organisation described above, but it is more a specific type of self-organisation that displays a qualitative distinction from the components that make up the whole. According to Goldstein:

> Emergence [...] refers to the arising of novel and coherent structures, patterns, and properties during the process of self-organization in complex systems. Emergent

125. Camazine S et al, *Self-Organization in Biological Systems*, Princeton NJ: Princeton University Press (2001), p.8.
126. Prigogine I and Stengers I, *Order Out of Chaos*, New York: Bantam (1984), p.13.
127. Ibid.

> phenomena are conceptualized as occurring on the macro level, in contrast to the micro-level components and processes out of which they arise.[128]

It must be pointed out that not all self-organising systems display emergence. As Corning points out, consciousness is emergent, but steam is not.[129]

How is it possible for a seemingly chaotic system to become organised and display behaviour that can be predicted? It seems counterintuitive to expect a complex environment to display self-organisation when one would expect the contrary. The answers can be found in several physical explanations that apply similarly to biological and physical phenomena. To understand this area of study first one needs to define what constitutes an ordered and a disordered system, a task that is not as easy as one may think. At the core of the idea of order, there is the concept of entropy, which in thermodynamics is the measure of the disorder within a system.[130] A system with high entropy is said to be more disordered than one with lower entropy. High entropic complex systems can be said to display stability; e.g. a pile of junk is a stable pile of junk, and something needs to happen for it to become something else. Normally, energy is one way in which a system can become ordered; e.g. a person going through a pile of junk may find usable parts to build a bicycle. Disordered systems become ordered all the time because of the application of energy; flowers grow due to the sun's energy; gases become less entropic by losing heat; computers use electricity to make computations. However, these systems are not self-organised according to the definitions used above. How does self-organisation emerge, if you pardon the play of words?

One common characteristic of self-organisation and emergence is that it can occur through the very interaction of the system's elements without outside influence. At the core of most research into self-organisation is how individual components somehow influence their neighbours, causing a chain reaction that will eventually result in the

---

128. Goldstein, supra note 103, p.49.
129. Corning PA, "The Re-Emergence of 'Emergence': A Venerable Concept in Search of a Theory", 7:6 *Complexity* 18 (2002), p.30.
130. Dugdale JS, *Entropy and its Physical Meaning*, 2nd ed, London: Taylor and Francis (1996), pp.5–15.

entire system becoming ordered.[131] There is a moment of change between the ordered and disordered system called phase transition, which takes place when a disordered system enters a period of criticality (known also as a meta-stable transitional period) after which the system becomes ordered rapidly (Figure 2.10). [132]

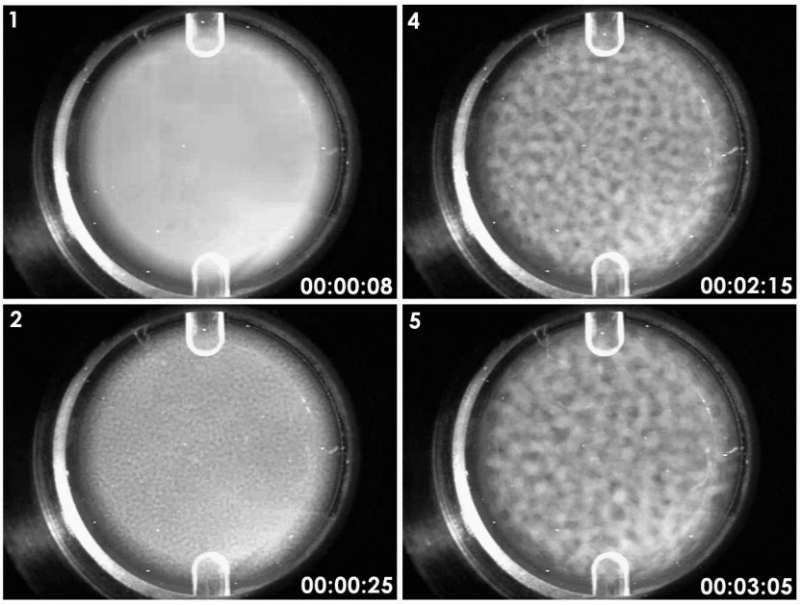


Figure 2.10 Phase transition of colloids in space[133]

An interesting feature of self-organisation has been discovered by conducting experiments into criticality of systems such as pendulums and sand piles. Researchers found that a stable system such as a pile of sand could undergo a critical transition at a stage when some more grains were added, creating an avalanche that released pressure.[134] While the event itself is not remarkable, as it is logical to assume that

131. Strogatz SH, *Sync: The Emerging Science of Spontaneous Order*, New York: Hyperion (2003), pp.42–45.
132. Ball, supra note 23, pp.100–103; and Strogatz, supra note 102, pp.230–251.
133. Still photographs taken over on the International Space Station to show the phase separation process of colloids, polystyrene polymers and solvents. http://mix.msfc.nasa.gov/ abstracts.php?p=2702.
134. Bak P, Tang C and Wiesenfeld K, "Self-Organized Criticality: An Explanation of 1/*f* Noise", 59 *Physical Review Letters* 381 (1987).

avalanches will happen if you add too much sand to a pile, the researchers found that the frequency with which this happened responded to a power law. In other words, the system organised itself into a state of criticality where avalanches could be expected at certain frequencies. This offers an interesting insight into self-organisation because if offers evidence that some systems will organise themselves into a state where phase transitions will occur, and that this organisation is somehow linked to power laws.[135]

Power laws, and particularly Zipf laws, are characteristic of self-organised criticality.[136] There could be a reason why this is so. Looking at scale-free networks, one could expect that hubs would have tremendous power to move a system in a given direction, causing self-ordering of the system.

There is a growing body of research into the way in which large groups of people make self-organising decisions in a seemingly spontaneous manner, particularly within game theory.[137] We see this type of spontaneous coordination in nature all the time, where flocks of birds move in one direction and another in coordinated fashion, but also responding as a whole to threats and sources of food.[138] Crowds of people tend to act in similar manner, any person familiar with crowd movements will notice inefficiencies in their behaviour, but also that at some point certain order emerges, particularly over time in what Surowiecki calls this "the wisdom of crowds".[139] But how does this self-organisation work?

The answer is surprisingly that crowds seem to behave just like gases. Henderson set out to study the way in which a crowd flows when walking through along a footpath.[140]

---

135. Ball, supra note 23, pp.296–302.
136. Ibid.
137. A typical and often cited example is El Farol Bar problem, where bar attendance is used to model uncoordinated behaviour, see: Arthur WB, "Inductive Reasoning and Bounded Rationality" 84 *American Economic Review* 406 (1994).
138. There are several computational models describing such behaviour, see: Reynolds C, "Flocks, Herds and Schools: A Distributed Behavioral Model", *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques* 25 (1987).
139. Surowiecki J, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few*, London: Abacus (2005).
140. Henderson LF, "The Statistics of Crowd Fluids", 229 *Nature* 331 (1971).

He discovered that plotting the speeds of people walking followed closely models used to describe the speed of gas particles in what is known as a Maxwell–Boltzmann curve. Further studies have managed to uncover more about crowd behaviour, discovering that crowds also display phase transitions from one state into another, just like the standing ovation described earlier.[141] The explanation for the emergence in the behaviour of crowds tells us a lot about the way in which social behaviour exhibits self-organisation characteristics. Apparently, people in crowds follow two simple rules, getting from A to B, but also doing so by keeping some personal space apart from other people in the crowd. By plotting these two rules into computer simulations, it is possible not only to get a very good idea of how crowds flow, but it is also possible to simulate how people react to obstacles along the way.[142] This gives us a hint of the self-organising power of elements in a complex system. By introducing a few rules, it is possible to explain and predict some kinds of human behaviour.

Perhaps the most common question asked whenever I talk to people about concepts of power laws and complex theory is "so what?" To point out that any given phenomenon occurs more frequently than any other is surely an exercise in stating the obvious, is it not? City sizes follow a power law, and so do avalanches, earthquakes and income distributions. So what? The answer to this question is that power laws do not only state that a city is bigger than another one, but that the frequency of occurrences follow specific and predictable ratios that hint at underlying causes that require an explanation. There is no reason why the letters in a corpus or the number of citations should follow a power law, but they do. Explanatory solutions such as emergence, self-organised criticality and phase transitions offer some explanations to why these things occur.

141. Ball, supra note 23, pp.162–165.
142. Helbing D et al, "How individuals learn to take turns: Emergence of alternating cooperation in a congestion game and the prisoner's dilemma", 8 *Advances in Complex Systems* 871 (2005).

3. Complexity and the Law

> It turns out that an eerie type of chaos can lurk just behind a facade of order---and yet, deep inside the chaos lurks an even eerier type of order.
> *Douglas Hofstadter*[1]

In 1970, popular science writer Martin Gardner published a column in *Scientific American* showcasing a game created by mathematician John Conway.[2] The game is called "The Game of Life", and it is played in a rectangular grid divided into squares. Each square can inhabit two states, alive or dead, and it replicates following four simple rules for each generation:

1. Each populated cell with one or no neighbours dies (isolation).
2. Each populated cell with four or more neighbours dies (overpopulation).
3. Each populated cell with two or three neighbours survives (survival).
4. Each empty cell with three neighbours becomes populated (reproduction).[3]

These rules allow for autonomous patterns to emerge from a limited set of starting variables. The player only needs to set the initial conditions, and each turn the cells undertake an automated game of life, death and births that is both chaotic and yet completely dependent on its initial conditions (Figure 3.1).

---

1. Hofstadter DR, *Metamagical Themas: Questing for the Essence of Mind and Pattern*, London: Penguin (1986).
2. Gardner M, "Mathematical Games: The Fantastic Combinations of John Conway's New Solitaire Game 'Life'", 223 *Scientific American* 120 (1970).
3. You can see an implementation of the game in Java here: http://www.bitstorm.org/ gameoflife/.

Conway's Game of Life has become famous in the study of complexity because it serves as a graphical representation of self-organising systems at work, where simple initial conditions can create ordered, dynamic and autonomous states.[4] It is no coincidence that the game has also been showcased in works dealing with evolutionary biology, as it also provides clear computational evidence of the power of self-reproducing automata, which offer obvious analogies to living organisms.[5] The reason for such interest should be evident when one looks at some of the complex and often beautiful designs that can arise from simple initial conditions. The basic set of rules, and the ease with which it can be converted into a computer program, means that the simplified model allows anyone to play with any given set of initial conditions. A lot of variations will die out quickly, while others will become stationary by reaching fitness peaks. Under certain circumstances, very complicated fractal effects can also emerge.[6]

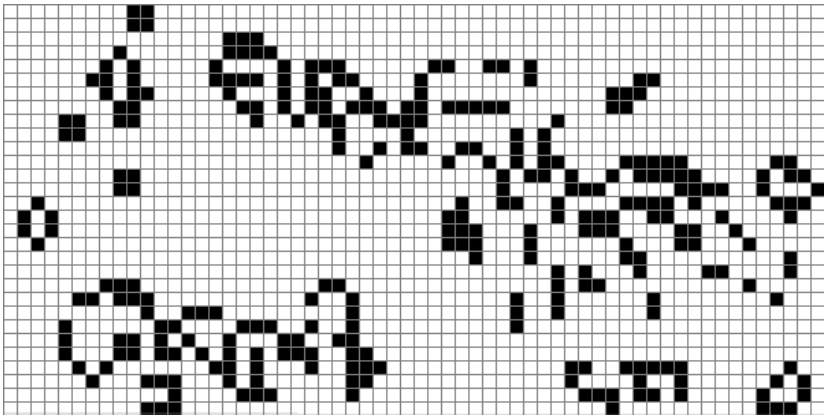


Figure 3.1 Conway's Game of Life.

The Game of Life, and other biological simulations describing complex adaptive systems, offers us interesting insights into the world of complexity described in the

4. Schulman LS and Seiden PE, "Statistical Mechanics of a Dynamical System Based on Conway's Game of Life", 19:3 *Journal of Statistical Physics* 293 (1978).
5. Dennett DC, *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, London: Penguin (1996), pp.167–171.
6. See this page for some animated patterns: http://radicaleye.com/lifepage/patterns/ javalife.html.

second chapter. The use of self-organising cellular automata serves as an illustration of precisely how phase transitions can occur in complex systems with very little initial input. However, what may be surprising is that applications of the game can be used to simulate social phenomena such as phase transitions in small world networks.[7] While it is clear that human beings do not operate in minimalistic conditions such as the rules in the Game of Life, the present work postulates that by studying such complex systems, it is possible to learn something about how larger networks operate.

The legal reader who has managed to wade through the previous chapter may be forgiven for asking the question of what it all means for the legal profession and research. Networks obeying certain rules and presenting specific architectures may be interesting to physicists, not to lawyers. Paraphrasing Leonard McCoy in Star Trek, "I'm a lawyer Jim, not a physicist". Nonetheless, this chapter will try to place the subject of complexity theory described in the last chapter in a legal context, attempting to prove that there is indeed something to be learned from the wealth of research into graph theory and networks.

1. NETWORK THEORY AND THE LAW

Before attempting to sketch a potential legal theory of complex systems, it is vital to understand whether there are any practical, philosophical and theoretical applications to network theory that may fit within a legal context.

At the time of writing, legal scholarship regarding the interaction between network theory and the law has been relatively scarce, but there is a growing body of literature on the subject.[8] This was perhaps inevitable as ideas of self-organisation that are tackled by

7. Huang S-Y et al, "Network-Induced Nonequilibrium Phase Transition in the 'Game Of Life'", 67:2 *Physical Review E* 026107 (2003).
8. Besides the works that will be showcased later, some works on the subject are: Hayes AW, "An Introduction to Chaos and Law", 60 *University of Missouri at Kansas City Law Review* 751 (1992); Matwyshyn AM, "Organizational Code: A Complexity Theory Perspective on Technology and

the various sub-themes in complexity theory offer powerful tools for social science as well as physical disciplines. In particular, economics has been the social science that has adopted such concepts the fastest, as evidenced by the existence of Pareto inequalities in all sorts of financial phenomena.[9] Sociologists have also been at the forefront of adopting several theories arising from the application of network theory, particularly its interest in small worlds in society.[10] Without meaning to sound too critical of my own profession, the law is just catching up, as is often the case.

Nonetheless, the number of legal scholars interested in the topic seems comparatively small given its prevalence in other social sciences. It is possible that the pervasiveness of mathematics and the technical nature of some of the papers may have dissuaded more attention to the topic. It is also possible that this is just part of the dichotomy between the physical and the social sciences that was highlighted in the Introduction. Network theory may be seen as nothing more than a formalistic and seemingly reductionist view of human society, an outlook that appears to erase complex social and political interactions, and replaces them with dots and lines on a chart. Nevertheless, if some better understanding of large systems has been made possible by complexity theory, then the law should take note and try to ascertain if there may be some legal issues worth exploring.

The possibility of following links and clusters of nodes and hubs means that the descriptive power of network theory can be easily tested in fields with pre-existing network-like characteristics. In the wider network research, a popular experimentation tool has been to chart citations between authors, or to discover small world

---

Intellectual Property Regulation", 11 *Journal of Technology Law & Policy* 13 (2006); and LoPucki LM, "The Systems Approach to Law", 82 *Cornell Law Review* 479, 480-82 (1997).

9. See for example: Simon HA and Bonini CP, "The Size Distribution of Business Firms" 48:4 *The American Economic Review* 607 (1958).
10. A notable example dealing with small worlds in the Broadway social scene is: Uzzi B and Spiro J, "Collaboration and Creativity: The Small World Problem", 111:2 *The American Journal of Sociology* 58 (2005).

characteristics of academic co-authorship networks.[11] It should come as no surprise then that a significant amount of existing literature charting the interaction between legal subjects and network theory is that of legal citation due to the availability of large datasets involving legal scholarship and case law, creating what Thomas Smith calls "the web of law".[12] Case law presents a valuable network subject matter because cases often cite existing precedent, the cases would be the nodes in the network, and the citations would be the links. Concentrating only on this network of cases and citations, Smith looked at the data from nearly four million US Federal rulings, and found a strong power law in the way in which cases cite one another. His study found that the vast majority of cases received few citations, while a significantly small number were cited a disproportionate number of times in a manner that clearly responded to power law exponential curves. If we recall some of the features of power law graphs, the mere appearance of inequality in occurrences is not enough to demonstrate the existence of power laws, but the difference between a case's rank in the chart will be inversely proportional to the following or preceding case. By charting these in logarithmic scale, the resulting histogram should be a straight line, which is precisely what happens with case citations (Figure 3.2). A similar power law was found when looking at legal scholarship citations, US Courts of Appeals, and the US Supreme Court cases.[13]

11. See: Redner S, "How Popular Is Your Paper? An Empirical Study of the Citation Distribution", 4:2 *The European Physical Journal B* 131 (1998); and Barabási A-L, Jeong H, Ravasz R, Néda Z, Vicsek T and Schubert A, "On The Topology Of The Scientific Collaboration Networks" 311 *Physica A* 590-614 (2002).
12. Smith TA, "The Web of Law", 44 *San Diego Law Review* 309 (2007).
13. Ibid, pp.331–335.

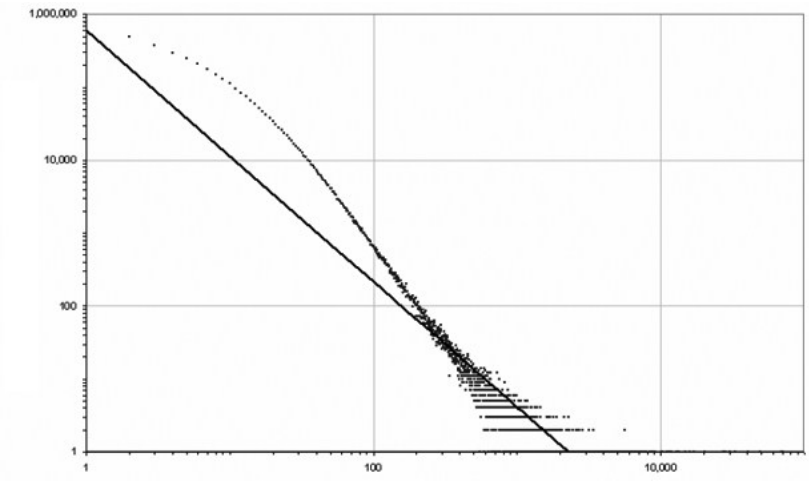


Figure 3.2 Power law distribution of US Federal cases[14]

Smith's results have been replicated in other studies looking at the network of cross-citations in US judicial decisions. Chandler conducted a study looking specifically at the way in which US Supreme Court decisions cite one other.[15] The study found that there is a scale-free topology at work as there are some decisions that are cited with disproportionate frequency. According to Chandler's research, the cases that act as the most cited hubs in this network of citations are older decisions regarding US Federal jurisdiction,[16] which may seem logical as this would be a legal area where precedent does not change that much, so the importance of precedent is transposed into more citations. It could be said that such a study may not be particularly enlightening, as it does not really say much about the actual nature of the rulings, but similar exercises could be of use for constitutional lawyers in different jurisdictions in order to recognise which cases are more likely to be encountered in future decisions, and also to determine the centrality of a case within the case law network.

14. Ibid, p.327. The *x* axis counts number of citing references, and the *y* axis counts the number of cases in log scale of nodes.
15. Chandler S, "The Network Structure of Supreme Court Jurisprudence", *International Mathematica Symposium 2005*, The University of Western Australia, Perth (August 2005).
16. The top two are *McCulloch v. Maryland* 17 U.S. 316; and *Gibbons v. Ogden* 22 U.S. 1.

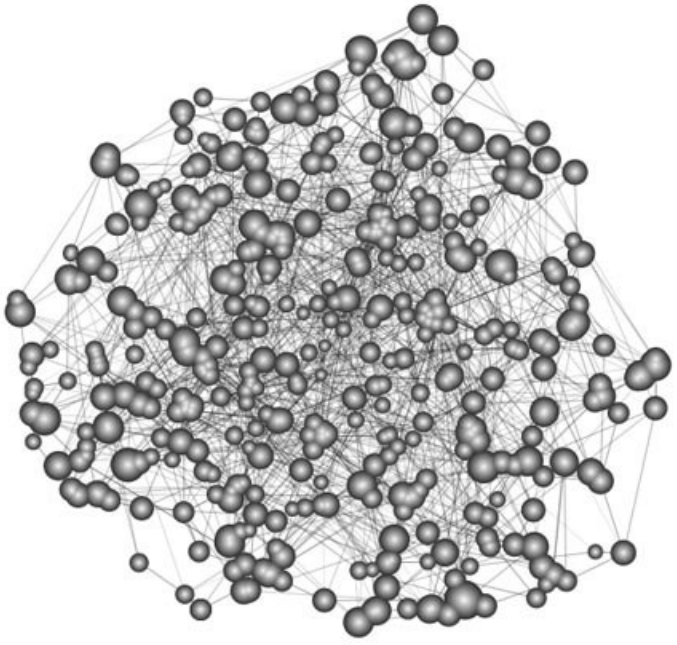
The clustering characteristic of legal scholarship and case law is just one area of potential usefulness to network theory to legal research. The cross-citation power laws discovered in legal citation could be extremely useful in patent law, where a tool that analyses the cross-citation of previously issued patents could prove to be an invaluable tool for patent examiners, attorneys and inventors. Strandburg[17] has conducted an excellent study looking at the clustering of citations in patents issued by the United States Patent and Trademarks Office (USPTO), which has demonstrated, amongst other things, that there seems to be increasing stratification in patent citeability since the 1980s. This means that a few patents are being cited with more frequency than in the past. Strandburg argues that this could be correlated with decreasing patent quality[18] experienced in the corresponding period. Another very interesting avenue of research explored in this paper is the possibility of an improved manner in which to classify patent claims. Currently, patent subject matter is assigned by examiners in an ad hoc fashion. Strandburg suggests that citation of previous patents may help in assigning the claim to a cluster, which would make its identification much easier. Citation clusters can be easily identified using network analytical tools as belonging to a small world community of patents, and so would be a better indication of whether it has been properly classified.

To illustrate Strandburg's findings, I took the liberty of downloading another patent citation dataset consisting of almost 3 million US patents granted between 1963 and 1999.[19] The links in the network are provided as citations made to those patents between 1975 and 1999, totalling over 16 million patents. I took the cross-citation by patent class, and applied small word modelling using a network analysis tool[20] to the data[21] and

17. Strandburg KJ, "Law and the Science of Networks: An Overview and an Application to the 'Patent Explosion'", 21 *Berkeley Technology Law Journal* 1293 (2007).
18. For more about patent quality, see: Jaffe A and Lerner J, *Innovation and Its Discontents*, Princeton NJ: Princeton University Press (2004), pp.61–69.
19. The dataset comes from: Hall BH, Jaffe AB and Tratjenberg M, *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools*, NBER Working Paper 8498 (2001); it can be downloaded from: http://vlado.fmf.uni-lj.si/pub/networks/data/patents/Patents.htm.
20. Network Workbench Tool, http://nwb.slis.indiana.edu/.

visualising it using small world graph software.[22] By doing this it became clear that there is high-clustering of patent class cross-citation very much in accordance with Strandburg's findings (Figure 3.3).

I will be the first person to admit that the aforementioned corroboration has to be taken with a large pinch of salt, but the experiment was done with one clear purpose in mind. The datasets are publicly available, and if one follows the methodology outlined in the cited works, it is possible even for someone without formal graph theory training to produce a graphic representation of any given dataset, particularly one that displays high clustering. The idea behind this experiment, unscientific as it may be, is that some of the basic tools for network analysis are available even for the uninitiated. All one needs is a healthy sense of curiosity and an unhealthy amount of time to test out the tools oneself.



21. Watts DJ and Strogatz SH, "Collective Dynamics of 'Small-World' Networks", 393 *Nature* 440 (1998).
22. Ham F and van Wijk J, "Interactive Visualization of Small World Graphs", *Proceedings of the IEEE Symposium on Information Visualization* (2004).

Figure 3.3 Small-world clustering in patent class citations[23]

Outside of citation, empirical application of network theory has been scarce, but again there are notable exceptions. Perhaps one of the most evident areas of study with regards to networks may very well be the regulatory arena. If we can understand a specific network that has given problems to regulators, then the potential for empirically-based research on how the network operates could provide clues as to how to regulate the troublesome area. A study has already attempted to look into the application of specific network theories to the telecommunications field.[24] Recognising that telecommunication networks operate as complex systems,[25] Spulber and Yoo hypothesise that the specific graphical representation of networks into hubs and nodes may be of use in trying to regulate emerging technologies such as access to broadband services and Voice-over-IP (VoIP) communications. This study has a narrow objective, as it relies only on the descriptive power of network science in order to provide regulators with the basis for charging for communications in complex telecoms networks. Is there room for a wider area of application?

Strahilevitz offers another empirical application to network theory by researching the legal implications of power laws and scale-free topographies in a ground-breaking analysis of the potential use of network science to the protection of privacy.[26] He uses the specific application of social network theories, such as small world distributions, to conclude that the scale-free nature of some social networks may provide us with tools with which we can measure the number of acquaintances that a member of the social system is likely to have. Then he proposes the fact that an individual involved in tort

23. Only the largest clusters show up in the visualisation.
24. Spulber D and Yoo CS, "On the Regulation of Networks as Complex Systems: A Graph Theory Approach", 99 *Northwestern University Law Review* 1687–1722 (2005).
25. The paper describes a complex system as "a system in which its elements interact in ways that transcend any organizing principles being applied to the network, allowing the network to evolve and adapt to environmental changes". Ibid, p.1694.
26. Strahilevitz LJ, "A Social Networks Theory of Privacy", 72 *University of Chicago Law Review* 919–988 (2005).

disputes about personal privacy may have the evidentiary means to measure the potential damage to his/her reputation and, therefore, a judge would be able to discern if there has been some actual damage done. He comments that:

> In a tort suit, courts are always called upon to examine causation: would the plaintiff have been harmed in the absence of the defendant's actions? Social networks theory provides a basis for evaluating that question when the plaintiff's injury stems from dissemination of previously private information. Courts simply need to ask themselves: was the widespread dissemination of this information inevitable, or did the defendant's actions materially affect the extent of subsequent disclosure?[27]

This is an elegant use of existing theories in order to provide a direct causal relationship to establish damages. However, one may be wary of establishing the causal link in the first place. If there is one thing that we have learned it is that scale-free networks predict that there will be super-connected nodes in a social network,[28] and we can easily expect individuals whose social interaction exceeds the average by various degrees. The person involved in the dispute could very well be one of those, and the calculation of actual damage could prove to be uncertain.

Another potentially valuable use of network theories in the law is in environmental policy-making. The life-sciences have had extensive experience in the use of empirical data in order to design policy in environmental and public health fields. The better understanding of complex environmental systems brought by some of the literature could be used in assessing risks posed by environmental threats, real or imagined.[29] Farber explains the use of power laws to design methods for assessing risks:

> The presence of statistical power laws supports the use of conservative methods of assessing risk. To be more specific, suppose that we are designing a procedure to

27. Ibid, p.975.
28. See for example: Kochen M, *The Small World*, Norwood, NJ:Ablex Publishing (1989), pp.147–158.
29. Farber DA, "Probabilities Behaving Badly: Complexity Theory and Environmental Uncertainty" 37 *U.C. Davis Law Review* 145 (2003), pp.156–161. For a less successful yet interesting attempt at marrying biotechnology and network science, see: Chen J, "Webs of Life: Biodiversity Conservation as a Species of Information Policy", 89 *Iowa Law Review* 495–608 (2003).

> identify any proposal posing a significant risk, with significance defined as some specific risk level such as one in ten thousand. [...] The only assumption is that among the relevant set of proposals, harmful effects follow a power-law distribution. If so, conservative test procedures may be warranted.[30]

In other words, in a scale-free environment we may expect harmful effects to occur, which are considerably higher than the average witnessed occurrences. If empirical research points towards the existence of power law distributions in a phenomenon that requires regulation, then conservative policies should be followed. This could certainly be useful if one considers that hurricanes appear to display scale-free characteristics.[31] Similar precautionary approaches could be taken in other life-science fields, particularly in public health policy. Pandemics like AIDS seem to follow scale-free behaviours,[32] where a few individuals can infect large numbers of people in a community by their role as connectors.[33] Public policy towards social pandemics like sexually transmitted diseases could be designed to look for these hubs and attempt to treat them first.[34]

These are just some examples of the growing body of legal scholarship tackling issues related to network theory, and hint at the potential that so far has been under-used, in my humble opinion. The current emphasis on the study of citation networks is probably caused by the fact that these are usually areas that are easy to data-mine, as there is an existing infrastructure to ascertain scholar impact and citeability. Nonetheless, as some of the empirical studies above demonstrate, this is a fertile ground for future research. Perhaps what is missing is a more widespread recognition of network theory's potential by existing legal theories. The following sections will attempt to continue making the case for doing just that.

---

30. Ibid, p.160.
31. Dessai S and Walter M, *Self-Organised Criticality and the Atmospheric Sciences: Selected Review, New Findings and Future Directions*, NCAR Extreme Events workshop, Boulder, CO (June 2000).
32. Dezso Z and Barabási A-L, "Halting viruses in scale-free networks", 65 *Physical Review E* 055103 (2002).
33. An example of this is the so-called patient-zero of the AIDS pandemic. See: Shilts R, *And the Band Played On: Politics, People, and the Aids Epidemic*, New York, NY: Penguin Books (1989).
34. Ayres I and Baker KK, "A Separate Crime of Reckless Sex", 72 *University of Chicago Law Review* 599 (2005), pp.610–614.

2. SELF-ORGANISATION AND THE LAW

2.1 Theoretical approaches

Besides the empirical applications to network theory described above, the study of emergent systems and self-organisation offer one of the most interesting areas of study to legal scholarship. Before trying to make a connection, we should revisit the concept of self-organisation studied in the previous chapter. For a system to be self-organised, it must contain an internal ordering process which does not respond to exogenous influences.[35] One way of looking at the law is to view it as an exercise in self-organisation within the complex system of society. The law fulfils one essential self-organising function, and it is to attempt to exercise order and control in society by various means. If we think of society as an emergent system, then the law in the shape of legislation, norms, regulation, case law and doctrine would constitute one of the internal sub-systems exerting an organising force.

Nonetheless, it is critical to distinguish what we are talking about when dealing with self-organisation in a legal context because there is room for confusion about the role of theories of emergence in legal context. Firstly, complexity theories of self-organisation can be used to explain how the law comes about and how it organises itself. This would be an internal theory of emergence of the law, and it would be concerned with the forces that shape the genesis and evolution of legal systems within its own system. Secondly, self-organisation can also be used to explain how the law works to shape other systems in self-organising fashion. As a result, the law itself is just another element in the wider complex societal system, and as such it helps to organise it. This would then be a meta-theory of the law as a self-organising element.

---

35. See Chapter 2, section 3.

The father of self-organisation studies in social systems is Niklas Luhmann with his influential theory of autopoiesis,[36] literally meaning self-creation. In its broadest sense, Luhmann's theory of autopoiesis follows the definition of self-organisation that has been used above so far as it describes social systems that respond to internal stimuli instead of relying on external elements.[37] Luhmann's autopoiesis theory rests on two significant assumptions with regards to social systems. First, he claims that social systems are real and not mere analytical abstractions; second, he postulates that social systems are self-referential as they produce their own meaning, or as he puts it, "everything that is used as a unit by the system is produced by the system itself".[38] According to Luhmann, highly ordered systems are not necessarily more complex than less ordered systems because the internal emergent factors need not be complex, and consequently the self-organising process can be a function of the interaction between these elements, and not a function of the complexity of the system.[39] This is undoubtedly consistent with Kauffman's theory of fitness landscapes;[40] the process needs only to produce fitness peaks for it to be ordered.

Luhmann himself saw the law as an autopoietic system part of the wider social network, but still differentiated and autonomous in its own merit. While it is informed by other sub-systems within society –such as politics, economics, religion, and education– it is self-referential, and for that reason self-organising. He comments that:

> [L]ike every autopoietic system, [the law] is and remains to a high degree dependent on its environment, and the artificiality of the functional differentiation of the social system as a whole only increases this dependency. And yet, as a closed system, the law is completely autonomous at the level of its own operations. Only the law can say what is lawful and what is unlawful, and in deciding this question it must always refer

36. While he did not coin the term, he is credited with using it to describe social systems. Autopoiesis first appeared in a discussion on biological self-creation: Maturana HR and Varela FJ, *Autopoiesis and Cognition: The Realization of the Living*, London: Reidel (1980).
37. Luhmann N, *Social Systems,* Stanford, CA: Stanford University Press (1995), p.22.
38. Luhmann N, *Essays on Self-Reference*, New York: Columbia University Press (1990), p.3.
39. Luhmann supra note 31, pp.34–36.
40. Kauffman S, *At Home in the Universe: The Search for Laws of Self-Organization and Complexity*, Oxford: Oxford University Press (1995).

> to the results of its own operations and to the consequences for the system's future operations. In each of its own operations it has to reproduce its own operational capacity. It achieves its structural stability through this recursivity and not, as one might suppose, through favorable input or worthy output.[41]

This is a very useful distinction between the two meanings of self-organisation described in previous paragraphs. Being part of the larger complex system of society, the law acts as one of its shaping elements in conjunction with other elements. However, the law is also its own autopoietic system, and thus one can study its own internal organising elements.

Autopoiesis has had tremendous influence in some legal theory circles.[42] It is not the role of this work to go into a detail description of the many works that it has inspired, even within the field of Internet regulation.[43] Relevant to this work, however, is the growing understanding of the regulatory power of self-organisation. Regulation itself is one of those tricky words that may mean different things in various contexts. In the strict legal usage, regulation can be defined as some form of external control that either restricts undesirable activities, or enables and facilitates others.[44] It is easy to see how the concept of autopoiesis is useful from a regulatory perspective, as it helps to explain how regulatory processes emerge, evolve and act as self-organising agents in society. Autopoietic regulation could be seen as an internal ordering force; organic, dynamic, and self-organising. This would contrast a more structured and hierarchical view of regulation known as "command and control" regulation,[45] where governmental bodies serve as the organising force exerting control in a top-down manner.

---

41. Luhmann N, "Law as a Social System", 83 *Northwestern University Law Review* 136 (1988), p.139.
42. Just to name a couple: Teubner G and Bankowski Z, *Law as an Autopoietic System*, Oxford: Blackwell Publishers (1993); and Baxter H, "Autopoiesis and the 'Relative Autonomy' of Law", 19:6 *Cardozo Law Review* 1987 (1998).
43. See for example, Savirimuthu A and Savirimuthu J, "Identity Theft and Systems Theory: The Fraud Act 2006 in Perspective", 4:4 *SCRIPTed* 436 (2007).
44. Baldwin R and Cave M, *Understanding Regulation: Theory, Strategy, and Practice*, Oxford: Oxford University Press (1999), p.2.
45. Sinclair D, "Self-Regulation versus Command and Control – Beyond False Dichotomies", 19 *Law & Policy* 529 (1997).

There is growing understanding of the role of self-organising complexity within the regulatory arena. Many authors have embraced Luhmann's autopoietic explanation of the law as a self-referential system and have adopted it as an explanation for self-regulating strategies where each element in the regulatory system is able to react within its environment and self-generate and reproduce internal solutions.[46] As it will be seen in more detail later, new technologies such as the Internet have provided particularly focal ground for the exploration of autopoietic regulation as it offers a clear contrast between command and control and self-regulation.[47]

Having said this, it is essential to note that there appears to be a clear split between the understanding of autopoiesis in legal systems and the concepts of self-organisation and emergence studied in the previous chapter. While Luhmann repeatedly uses examples from biology to describe autopoiesis, and his concept of self-organisation matches that used in the physical sciences, it is clear that autopoiesis is very much a social theory. With few exceptions,[48] the theoretical study of autopoiesis is devoid of the mathematical treatment and the wealth of evidence into self-organisation involving information theory, phase transitions and emergence described in Chapter 2. It is almost as if the social sciences and the physical sciences arrived at the same conclusion following entirely different paths. This could be caused yet again by the pervasive split between the social and physical sciences that is the common theme that runs through academia. This is unfortunate because both fields could use with some cross-pollination. Unfortunately, post-Sokal tendencies in the physical sciences appear to look at sociology with decided distrust.

That is not to say that there have been no attempts at bringing both worlds together, particularly in the existence of studies that not only offer an insight into the presence of

---

46. King M, "The 'Truth' About Autopoiesis", 20:2 *Journal of Law and Society* 218 (1993).
47. Murray has been particularly interested in drawing this distinction. See: Murray A, "Conceptualising the Post-Regulatory (Cyber)State", in Brownsword R and Yeung K (eds), *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes,* Oxford: Hart (2008), pp.287–315.
48. Most notably: Mingers J, *Self-Producing Systems: Implications and Applications of Autopoiesis*, London: Plenum Press (1995).

self-organisation in the law, but that are also aware of –and embrace – complexity theory.

Post and Johnson,[49] amongst others, have been at the forefront in bringing ideas of complex adaptive systems to the law. Some of the work on this area follows Kauffman's studies into complexity theory in biological systems outlined in the second chapter. According to the view, the law is an interconnected system which spontaneously arrives at a highly-ordered states. Post and Johnson use the idea of patching in complex systems to make both a descriptive and normative comment about the emergent faculties of the law. Patching is, according to Kauffman, a method of looking at the interaction of complex systems where its constituent elements are divided into quilt squares, or patches. Each patch tries to achieve optimal fitness regardless of what the other elements are doing, but in doing so they influence the overall state of the system as each patch encourages co-evolution into more ordered and efficient states.[50] Post and Johnson propose that the law can be seen in similar light by saying that:

> Legal institutions are (or should be) designed to solve problems defined over complex systems [...]. If we are to have effective problem-solving in this complex policy space, a central goal for the design of legal institutions is the formation of congruent, independently optimizing decision-making sub-groups.[51]

This is an interesting use of self-organisation, one that provides a useful theoretical framework to look at the way in which legal decision-making is achieved. The institutional self-organisation of legal networks seen is this light would perhaps provide answers about how certain legal decisions are made, but also about how the web of law, to use Smith's term, comes together into a coherent whole.

---

49. Post D and Johnson D, "Chaos Prevailing on Every Continent": A New Theory of Decentralized Decision-Making in Complex Systems", 73 *Chicago-Kent Law Review* 1055 (1999).
50. Ibid, p.262.
51. Post and Johnson, supra note 49, p.1084.

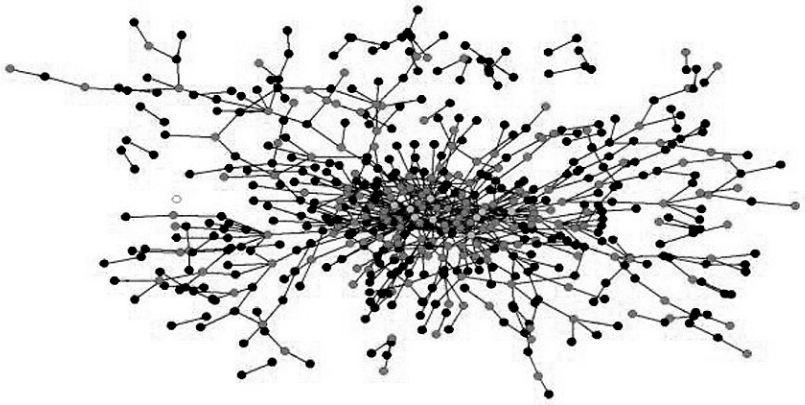


Figure 3.4 Social network structure of the US Federal judiciary[52]

Social network and small world theories may provide some help to try to determine the self-organising elements of the law. Following the patch example, self-organisation occurs when constituent nodes in the network come together and affect one another, consequently shaping the whole. If we think of the interaction between actors within the network, particularly from a social network perspective, then we could try to see how the legal web displays emergent features. For example, Katz and Stafford[53] collected data for 20,000 clerks working in the US Federal judiciary, and tried to paint a picture of the social structure within that vast network. They found high levels of clustering amongst clerks, which responded to scale-free topologies. Drawing from the complex systems literature, the authors found that the high clustering was in no way a directed phenomenon, and it displayed some form of self-organisation of the actors in the network that was dependent on the initial conditions of the network, suggesting that federal judicial actors self-organize at positions of criticality. Apparently, clerks who knew each other tended to cluster together at later stages of their careers, shaping the way in which the network organised. Unsurprisingly, some of the actors had more connections, and these tended to become essential parts of the whole. When visualising

52. Ibid.
53. Katz DM and Stafford DK. "Hustle and Flow: A Social Network Analysis of the American Federal Judiciary", 71:3 *Ohio State Law Journal* (2010).

the links of acquaintance and publication of the nodes, a familiar scale-free topology emerged (Figure 3.4).

2.2 Finding self-organisation in legal systems

From a theoretical perspective, autopoiesis and emergence in legal systems make a lot of sense. Nonetheless, there real usefulness of the theory exists if we can identify self-organising systems in the law.

Emison provides us with an invaluable test to recognise self-organising adaptive systems in a policy context.[54] In order to potentially make use of complexity theory in environmental policy, he set out to discover whether the subject of environmental regulation featured self-organisation. This is an key step, as he believed that environmental policy until then had been ineffective because it was attempting to regulate self-organising subject matter. However, Emison's concern may have been misplaced – as self-organisation and regulation are not exclusive, one just needs to think of the way in which crowds flow to realise that often architectural decisions may direct emergent forces in one direction, a subject that will be dealt with in more detail in the next chapter. Nonetheless, assuming that self-organisation is problematic, Emison argued that there are several characteristics of self-emergence for regulatory purposes:

1. Systems' components have a common purpose; the system's elements work together to move the whole to a different condition.
2. These systems are indeterminately complicated; the large number of interactions in a self-organising system means that it contains levels of complexity that cannot be pre-determined.
3. The systems are non-linear and dynamic; constant change in self-organising systems means that interactions and effects do not follow linear causality.
4. As systems change, new conditions emerge.

54. Emison GA, "The Potential for Unconventional Progress: Complex Adaptive Systems and Environmental Quality Policy", 7 *Duke Environmental Law & Policy Forum* 167 (1996).

5. The new conditions emerge towards patterns.
6. The patterns are self-similar; in other words, the system is exactly or approximately similar to a part of itself.

If the system under study exhibits these characteristics, then traditional "command and control" regulation will simply not suffice, and new approaches should be sought. This is a useful and tangible application of self-organisation for legal purposes.

Arguably, one of the most concise empirical applications of complexity theory and emergence in the legal field has been Trujillo's study on bankruptcy.[55] In this ground-breaking analysis, he took all of the bankruptcy corporate cramdown valuation decisions from 1970 to 1998[56] to try to discern patterns in valuation procedures at court level. One would expect that in any normal valuation process of bankrupt corporate assets would either follow no pattern, or there would be a split in valuation between what is being asked by the debtor and the creditor. However, Trujillo found that this was not the case, and he uncovered patterns in valuation that responded to "winner-takes-all" scenarios described by Bose–Einstein condensation.[57] He took this amazing finding, and other similar data, and showed that there was a pattern of self-organisation in legal decisions. He states that:

> The data offer preliminary support for the conclusion that some aspects of the U.S. bankruptcy legal system show a tendency to self-organize. Conclusive evidence of self-organizing dynamics in a legal system could have substantial jurisprudential significance. We know that simple deterministic dynamics do not explain the data we observe in legal systems. Since the decline of legal formalism, the dominant mode of explanation has been to attribute a randomness, or nondeterminism, to legal system dynamics and to suggest that any observable patterns are due to exogenous ordering—

55. Trujillo B, "Patterns in a Complex System: An Empirical Study of Valuation in Business Bankruptcy Cases", 53 *UCLA Law Review* 357 (2005).
56. "In a typical business cramdown, a bankrupt corporation (the debtor) has filed a plan of reorganization to which at least one class of creditors has objected. The debtor requests that the bankruptcy court confirm the plan despite the objection, in effect 'cramming the plan down' the throats of the objecting class of creditors". Ibid, p.359.
57. Ibid, p.371.

> such as decisionmaker bias—that affect legal ordering intersystemically. Evidence of self-organizing dynamics suggests the possibility that at least some of the patterns we observe are generated by deterministic dynamics operating intrasystemically.

Trujillo's research and analysis throws an interesting challenge to jurisprudential thought. Here we are presented with clear examples of self-organisation in legal contexts. As it will be made evident later in the book, his example is not the only one by far. If there is some form of self-organisation at the heart of legal decisions, what does that have to say about legal theory? Is there room for a complexity school of legal thought?

3. COMPLEXITY AND EXISTING LEGAL THEORIES

Writing about the possible application of complexity theory in discrimination cases, Di Lorenzo complained that judicial decisions and legislation in the United States had remained blind to the issue of complexity.[58] His grievance could be extended to the subject of legal doctrine and jurisprudence; although legal philosophy has been adopting an interdisciplinary approach involving economics and sociology, theories of complexity still do not fit easily into existing legal theories. As Trujillo remarked above, a deterministic outlook of legal processes may not fit well with prevalent doctrines more interested in randomness than in complexity. That is not to say that the subject has gone unnoticed,[59] and there are several attempts at placing the idea of complexity theory at the heart of jurisprudence.

---

58. Di Lorenzo V, "Complexity and Legislative Signatures: Lending Discrimination Laws as a Test Case", 12 *Journal of Law & Politics* 637 (1996), p.641.
59. Besides the research that will be highlighted later, some articles of note are: Goldberg DS, "And the Walls Came Tumbling Down: How Classical Scientific Fallacies Undermine the Validity of Textualism and Originalism", 39 *Houston Law Review* 463 (2002); Picker RC, "Simple Games in a Complex World: A Generative Approach to the Adoption of Norms", 64 *University of Chicago Law Review* 1225 (1997); Scott RE, "Chaos Theory and the Justice Paradox", 35 *William & Mary Law Review* 329, (1993); and Post DG and Eisen MB, "How Long is the Coastline of the Law? Thoughts on the Fractal Nature of Legal Systems", 29 *Journal of Legal Studies* 545 (2000).

The holdout on the recognition of the role of complex theory in law is twofold. Firstly, there is understandable reluctance to adopt seemingly mechanistic descriptive theories into complex social phenomena. Secondly, there is the fact that current legal theory tends to favour more socially oriented and less formalistic approaches to legal thought.

At first glance, the legal theory that could accommodate and explain some of the research highlighted in the previous sections is that of Formalism, which can be roughly described as a strict adherence to the letter of the law, without taking into account interpretation, or the social, political and anthropological circumstances that serve to understand legal decisions. [60] Perhaps unjustifiably, Formalism in legal theory has become equivalent with ideas of strict logical construction of legal decisions that are anathema to realism, utilitarianism and critical legal studies.[61] Complexity theory would appear to be formalistic as it offers descriptive tools that appear deterministic, suggesting the existence of pre-determined, unmovable and unavoidable rules of legal formation. However, a closer look at the basic components of complexity theory as it might apply to the law would serve to dispel this concern.

Take, for example, the ideas of self-organisation covered in the previous section. Emergent systems display a self-organising structure embedded in their own elements, the overall emergent system is dependent on the complex interaction of its component parts. If we look at the judicial system as an emergent system, then we are not making any comment about the actual decisions, but we could gain insight into how the decisions are made. Katz, Safford and Provins make this point by stating that:

> [J]udicial decision-making is decision-making in a judicial hierarchy. Agents across the institution consistently interact and those interactions undoubtedly consequence aggregate outputs. An important precursor to gaining leverage on the empirical implications of this revelation is an effort to develop a positive theory of judicial social structure. Much like the study of the pixels or the understanding of traffic systems, existing theories could benefit from modeling both direct and indirect

60. Stone M, "Formalism", in Coleman JL, Shapiro S and Himma KE (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law*, Oxford: Oxford University Press (2004), pp.170–171.
61. Ibid, pp.166–167.

> interactions between judicial agents. Along with factors identified by behavioral and strategic institutional scholars, we believe that a holistic model of judicial decision-making should account for the institution's self-organized social topology and its role in structuring the emergent and convergent outputs produced by the aggregate institution.[62]

Therefore, complexity and self-organisation could not be classed under Formalism as it is usually understood because the interactions of the components in the system are vital to the emergent patterns in doctrine, case law and legislation. But if complexity theory is not Formalism, then what is it?

A logical home for complexity theory would be one of the most interdisciplinary legal theories in existence, Law and Economics, which is the legal theory concerned with the application of economic analysis to legal systems.[63] As has been mentioned already, economists have been the quickest social science to adopt several complexity theory tools.[64] In particular, economists have seen in complexity theory potential explanatory tools for economic phenomena that have been a feature of the field since Adam Smith. The idea of complex adaptive systems and emergence fit well with economic themes such as the self-organising power of markets, the interaction between agents and the dynamic organisation of networks within the economy.[65] As is the case with other theories of self-organisation such as autopoiesis, Law and Economics has been greatly influenced by the biological sciences, in particular evolutionary theory.[66] In a ground-breaking article on the subject, Roe uses several chaos and complexity features of

---

62. Katz DM, Stafford DK and Provins E, "Social Architecture, Judicial Peer Effects and the 'Evolution' of the Law: Toward a Positive Theory of Judicial Social Structure", 24 *Georgia State University Law Review* 977 (2008), p.985.
63. Ronald Coase is usually considered to be the father of this discipline. See: Coase R, "The Problem of Social Cost", 3:1 *The Journal of Law and Economics* 1 (1960).
64. See for example: Beinhocker ED, *The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics*, Boston MA: Harvard Business School Press (2006); and Arthur B, Durlauf S and Lane DA, "Introduction: Process and Emergence in the Economy", in Arthur B, Durlauf S and Lane DA (eds), *The Economy as an Evolving Complex System II*, Reading, MA: Addison-Wesley (1997).
65. Ibid.
66. For example, see: Hirshleifer J, *The Dark Side of the Force: Economic Foundations of Conflict Theory*, Cambridge: Cambridge University Press (2001), pp.183–197.

evolutionary biology to explain corporate organisation and self-regulation.[67] Nonetheless, while it is tempting to try to shoehorn complexity theory within the wider field of Law and Economics, there could be several problems with this approach. Chiefly amongst those, Law and Economics has been a central feature of the Chicago School of Economics[68] where it could be seen as an ideologically-driven field of study concerned with the supremacy of the market, which explains its interest in evolutionary biology and the prevalence of the use of the idea of survival of the fittest. While complexity theory has strong implications for self-organisation, it would be wrong to allocate an ideological explanation to such undertakings.

In a similar vein to that of Law and Economics, Ruhl offers an interesting attempt to draw analogies between competing legal theories and Darwinian evolutionary theory by making one-on-one comparisons between legal theories and their "equivalent" in evolutionary biology. According to his novel take on the subject, Formalism is akin to the more discredited forms of Social Darwinism, Realism is equated to punctuated equilibrium,[69] Critical Legal Studies are equivalent to ecosystem nonequilibrium[70] and so forth.[71] While this exercise feels rather forced at times, Ruhl's contribution serves to further disentangle the jurisprudential web of complexity by making it clear that issues such as fitness, emergence and self-organisation do not fall easily into existing legal theories, particularly in theories such as Realism and Critical Legal Studies.

There appears to be an overarching feature in most of the literature dealing with jurisprudential thought of complexity theory. Many authors are clearly aware that there is something of great consequence happening in the physical sciences, so they attempt to

---

67. Roe MJ, "Chaos and Evolution in Law and Economics", 109 *Harvard Law Review* 641 (1995).
68. Mackaay E, "History of Law and Economics", in Bouckaert B and De Geest G, *Encyclopedia of Law & Economics*, Cheltenham UK; Northampton, MA: Edward Elgar (2000), pp.65–117.
69. For more about punctuated equilibrium, see: Gould SJ and Eldredge N, "Punctuated Equilibrium Comes of Age", 366 *Nature* 223 (1993).
70. See: Phillips JD, "Divergence, Sensitivity, and Nonequilibrium in Ecosystems", 36:4 *Geographical Analysis* 369 (2004).
71. Ruhl JB, "The Fitness of Law: Using Complexity Theory to Describe the Evolution of Law and Society and its Practical Meaning for Democracy", 49 *Vanderbilt Law Review* 1407 (1996).

tie in some of the ideas of chaos, complexity and networks into legal phenomena.[72] There is a clear danger then that the study of potential application of complexity in legal phenomena could be regarded as simply a fad initiated by scholars who have read a couple of books on chaos theory or graph theory and decided to apply it to their research. However, as evidenced by the growing body of relevant and informative research into networks highlighted in the first section, there is definitely a tangible role of complexity theory in legal scholarship. The existence of power laws in legal doctrine and case law should give us a hint as to the presence of an emergent feature embedded in legal systems. However, this feature does not fit well into existing legal theories. Is it perhaps time for a new legal theory?

4. A NEW LEGAL THEORY OF COMPLEX SYSTEMS?

Assuming that complex theory cannot be housed in existing legal theories, it could be time to formulate a new theory. While it is not the remit of the present work to undertake such a task, this could be a good opportunity to sketch what a legal theory of complex systems should look like.

The first task is to try to set the bar high as to what constitutes a valuable and viable application of complexity theory to the law. Complexity theory in general and network theory specifically, are broad fields of study that encompass a large range of sub-theories and practical applications. So if one is to seek a valuable contribution from these fields to jurisprudence, one has to initially try to filter what constitutes viable input from these fields. Just because a set of data displays a power law, it does not mean that it should be immediately worthy of study. The first rule of a potential theory should be relevance; in order to be of interest to legal theory, the application of complexity theory

72. For example, an excellent study on computational complexity can be found here: Kades E, "The Laws of Complexity and the Complexity of Laws: The Implications of Computational Complexity Theory for the Law", 49 *Rutgers Law Review* 403, 452 (1997).

should tell us something crucial about the law. As stated by Geu, the role of a new theory of legal complexity does not need to have immediate practical application, but it could serve to arrive to an adequate level of confidence about the predictive power of the application of complex theory to legal problems.[73]

The second element that needs to go into a new legal theory of complexity is to determine exactly what is being studied. Legal theory is generally concerned with endogenous and exogenous elements of study. Amongst the exogenous topics one can find the study of how the law operates within society, and what the external elements that shape it are. Endogenous topics include theories that study the interaction of law formation, regulation, decision-making processes, legislative power, case law, policymaking and enforcement. Exogenous topics include the role of law in the economy, the political process and society as a whole; but also cover external influences into the endogenous processes.

Complexity theory can serve to explain both endogenous and exogenous topics. A couple of examples can serve to illustrate this. Looking at exogenous elements that shape legal processes, Luhmann has commented that the study of self-organisation in legal systems is being shaped by the application of extraneous disciplines such as computing, information theory, robotics and autopoiesis.[74] On the endogenous side, Di Lorenzo has been looking at the analytical power of complexity theory in order to study the dynamics of the legislative process with emphasis on the indeterminacy created by competing and dynamic elements present in the legislative decision making.[75] In both instances, we have scholars interested in bringing in tools that are more easily found in biology and physics textbooks, and applying them to legal topics both at internal and external levels.

---

73. Geu TE, "Chaos, Complexity, and Coevolution: The Web of Law, Management Theory, and Law Related Services at the Millennium", 65 *Tennessee Law Review* 925 (1998).
74. Luhmann, supra note 37, p.178.
75. Di Lorenzo V, "Legislative Chaos: An Exploratory Study", 12 *Yale Law & Policy Review* 425 (1994).

It is the endogenous aspects of interaction that are more relevant to legal theory as a whole. When we look at legal processes through the eyes of complexity theory we see a vast networks of norms: networks of legislators interacting with lobbyists and stakeholders; networks of case law interacting at different hierarchical levels; networks of legal citation; networks of interpretation of norms; networks of enforcement and networks of regulation. Just looking at each one of those elements we see emergent features that hint at underlying ordering rules. Paraphrasing Geu, the law "oozes complexity".[76]

This could of course be completely irrelevant to legal theory as a whole, but if the patterns tell us something about how decisions are made, and how policies are adopted and enforced, then complexity and network theories have a valid place in jurisprudence. Traditionally, we like to think of the law not as a chaotic system, but as the result of decisions by players according to established constitutional and legislative rules. Those norms are set in place by legislative powers that are also not considered to be chaotic. However, some aspects of complexity theory do account for individual actions and their subsequent effects in complex systems, particularly in the theories of self-organisation. Would it be useful to look at the law much in the same way as we look at the phenomenon of standing ovations covered in the second chapter? Complex adaptive systems in particular are modelled to take into account the action of prime movers, and to measure the effect these have in the system as a whole. Understanding the law as an autopoietic system allows actors to make decisions that will shape the network as a whole.

Under this light, it might be possible to establish a nested hierarchy of decisions that generate phase transitions in the legal sphere. Think for example of the most basic type of legal sources, that of social norms. Sunstein has remarked how many social norms do not seem to originate from rational behaviour, and often respond to arbitrary, inefficient

76. Geu TE, "The Tao of Jurisprudence: Chaos, Brain Science, Synchronicity, and the Law", 61 *Tennessee Law Review* 933 (1994), p.989.

and irrational choices by people that eventually get turned into generally accepted social precepts.[77] This behaviour could be viewed from strictly sociological and economic perspectives, but it could also be seen under the eyes of complexity theory. If we assign fitness levels to norms, one could say that under some circumstances some decisions enter a phase transition that eventually becomes widely accepted, hence reaching a peak in fitness landscapes and solidifying its wide adoption. In fact, these become so widely accepted that it becomes very difficult to shift social behaviour, even if the norm is shown to be irrational.

To illustrate this point, we need only look at the way in which norms are created and enforced in society. Milgram[78] conducted several social experiments in New York by getting some of his students to challenge deeply ingrained social norms, such as jumping queues and asking people for their seat in public transport; he found that while people seemed discomforted by the blatant breaking of rules, they were more likely to acquiesce if the person was firm and assumed an air of authority. In some instances, Milgram found that when someone blatantly violates a social norm such as jumping a queue, very few people protested vocally; and about half of the people asked to give up their seat actually did so. But perhaps one of the most interesting parts of the experiment was just how reluctant his students were to actually undertake the braking of established social norms. One could say that these norms have reached a peak in fitness landscapes, and moving the behaviour towards different norms could prove akin to the act of speciation in evolutionary biology.

Similar considerations can be undertaken in the hierarchy of norms, and the role of a legal theory of complex systems could be to establish more examples amongst legal sub-systems, just like many scholars have been doing.

As stated in the first section, one of the most useful roles of a legal theory of complexity could be its application in the area of regulation and policymaking. Take

77. Sunstein C, "Social Norms and Social Roles", 96 *Columbia Law Review* 903 (1996).
78. Milgram S, *The Individual in a Social World*, New York: McGraw-Hill (1992), pp.19–33.

Baldwin and Cave's definition of regulation as the sustained and focused application of control over social activities, be this governmental, market-driven or social norm.[79] Given this definition, theories of complexity can prove useful when trying to determine and exercise regulatory control, as better understanding of group behaviour can elucidate regulatory puzzles.

To illustrate this, let us look at crowd flow again. Under the more inclusive definitions of regulation, the attempt to direct pedestrian or vehicular traffic is indeed a type of regulation inasmuch as a public authority is trying to exercise control over a human activity. This is an area where complexity theory has become invaluable as the use of computer simulations based on complexity principles can be used to devise better, safer and more efficient manners of traffic control. A notable example of this has been the building of pedestrian bridges in Saudi Arabia to control the flow of pilgrims at the Hajj. We are all familiar with the news stories of hundreds of people dying due to trampling and overcrowding of poorly designed pathways; the worst of these took place in 1990, where 1426 people died in a tunnel leading out from Mecca.[80] After two serious incidents in 2004 and 2006,[81] plans went underway to redesign the access bridges to the stoning pillars where most incidents took place, but this time taking into account simulations based on complexity theory. Particularly, researchers from the Swiss Federal Institute of Technology looked at footage from previous disasters, and were able to come up with a model that not only accurately described and replicated what happened, but also would be used in the building of the new access ramp to avoid future incidents.[82]

---

79. Baldwin and Cave, supra 44 note p.2.
80. Asser M, "Hajj Perils, Ancient and Modern", *BBC News* (5 March 2001), http://news.bbc.co.uk/1/hi/world/middle_east/1203697.stm.
81. "Hundreds killed in Hajj stampede" *BBC News* (12 January 2006), http://news.bbc.co.uk/1/hi/world/middle_east/4606002.stm.
82. Helbing D, Johansson A and Al-Abideen HZ, "The Dynamics of Crowd Disasters: An Empirical Study", 75 *Physical Review E* 046109 (2007).

Traffic control is not only a significant empirical regulatory application of complexity theory, but it also provides a valuable analogy for regulatory approaches as a whole. Often, regulatory power is exercised without taking into account how people actually behave in any given situation. When I have presented in conferences about this topic, I usually finish my slides with a picture of a footpath located at the University of Stuttgart (Figure 3.5). In it, one can clearly see a designed pathway in an open park connecting several buildings. However, it is clear from the picture which is the preferred route taken by pedestrians, which is unsurprisingly the shortest and most efficient route across the park, clearly ignoring the beautifully designed walkway. Would-be regulators may be well served by keeping images such as this in mind when trying to deploy regulatory solutions dealing with complex human behaviour.

Figure 3.5 Regulatory failure[83]

With these starting points in mind, I propose that a legal theory of complex systems should recognise at least a couple of corollaries arising from network and complexity theory.

83. Ball P, *Critical Mass: How One Thing Leads to Another*, London: Arrow Books (2004), p.169.

1. *The law is a self-organising system*. If we take for granted the both the definition of both self-organising systems and of autopoiesis covered already, then it becomes evident that the law is a complex system that contains endogenous elements which function towards the emergence of self-organising characteristics. Out of seemingly chaotic elements, an ordered and coherent system seems to arrive, seemingly out of nowhere.
2. *The law is a dynamic network*. Taking also the definition of network described in the last chapter, the law can be seen as a network consisting of interactive nodes. These interactions are more akin to the dynamic and rapidly changing environment found in biological systems, instead of steady networks such as Euhler's bridges. As a result, the law as a network can be charted using graph theory, and these interactions will display either random behaviour, or scale-free topologies.

What to call this legal theory? Professor Jim Chen and other scholars have started a blog called Jurisdynamics, which "describes the interplay between legal responses to exogenous change and the law's own endogenous capacity for adaptation".[84] This is a catchy name, and could perfectly describe the interaction between the law and complexity theory.

Hopefully, this chapter has served to make the case that there are potential legal applications to network theory. The next chapters will look at the specific case of Internet regulation and cyberspace law in order to describe in more detail some of the issues that can be explored using complexity theory.

84. Chen J, "Introducing Jurisdynamics, a New Blog on Law", *Jurisdynamics Blog* (July 14, 2004), http://jurisdynamics.blogspot.com/2006/07/introducing-jurisdynamics-new-blog-on.html.

4. Internet Architecture and Regulation

> Every other medium is somewhat responsibly regulated. The Internet is the only one that is being left alone in the name of informational freedom. People say that they want the Internet to be free and they want to make sure that no one controls it. The idea that no one controls the Internet is laughable. Whoever controls the delivery systems controls the Internet. And these people aren't doing it out of philosophical enlightenment or out of charity; their intention is to have control over the market. They're really sucking people in with this thing. They're making intelligent people believe that the Internet is a force for freedom and democracy. But it can be used for anything.
> Caleb Carr[1]

In 1993 author Julian Dibbell published a remarkable article entitled "A Rape in Cyberspace".[2] In it he recounts the happenings of a virtual world called LambdaMOO,[3] a text-based environment with roughly 100 subscribers where the users adopted assumed personalities (or avatars) and engaged in various role-playing scenarios. Dibbell tells the story of Mr Bungle, a clown avatar who programmed a routine into the virtual environment called a voodoo doll, which had the function of taking another person's avatar and manipulating it to follow the controller's orders. Mr Bungle used his voodoo doll to describe various unsavoury sexual encounters with other characters against their will. The incident became a scandal in LambdaMOO; the fact that this was not real was irrelevant to the affected users. They felt violated, and demanded some form of action from the community. What followed was a complex discussion regarding social norms

1. Interviewed by: Offman C, "Fight the Future!" 8:12 *Wired Magazine* (2000), p.102.
2. Dibbell J, "A Rape in Cyberspace", *The Village Voice* (21 December, 1993), http://www.juliandibbell.com/texts/bungle_vv.html.
3. A MOO is a recursive acronym that describes an object oriented Multi-User Dungeon (MUD). http://en.wikipedia.org/wiki/MOO.

in virtual environments, punishment and enforcement. The community decided that some form of penalty was warranted, while the designers of the space, the so-called wizards, proclaimed that they would implement whatever judgement was passed by the community. Discussion ensued, arguments and counter-arguments came, but in the end one of the wizards decided to "toad" the Mr Bungle character, that is, it was deleted from the system. Judgement had been passed and enforced in the most terminal way possible in virtual worlds, character deletion.

The story of LambdaMOO has become a classic in Internet regulation literature, and has been pondered and retold in seminal works such as Lessig's *Code*[4] and Goldsmith and Wu's *Who Controls the Internet*.[5] It is a testament to Dibbell's powerful writing that the story of the virtual misconduct of an avatar during the early days of cyberspace could have such an effect on the body of work dealing with the regulation of the Web; at the time it came out, the virtual capital punishment of Mr Bungle seemed like a perfect example of self-regulation and governance of the online world. Since it was written, we have become used to much more serious online offences –Internet trolling occurs on a daily basis – and regulation theories have much more to worry about than users taking their online role-playing games way too seriously. Nonetheless, the story of LambdaMOO still resonates because it brings us back to crucial questions that have been the subject of literature, philosophy and jurisprudence for centuries. How does a community organise itself? Is external action needed, or does self-regulation work? What constitutes regulatory dialogue? How does regulatory consensus arise? And most importantly, who enforces norms?

While this work does not pretend to answer these age-old questions, it is clear that new technologies have been proving fertile ground to enrich existing theories of regulation. The emergence of the World Wide Web, and the growing numbers of people who use it in critical facets of their lives, have served to examine some of the

4. Lessig L, *Code: and Other Laws of Cyberspace,* New York, NY: Basic Books (1999), pp.74–78.
5. Goldsmith JL and Wu T, *Who Controls the Internet? Illusions of a Borderless World*, Oxford: Oxford University Press (2006), pp.14–17.

assumptions about how we regulate these spaces. When presented with previously uncharted legal issues, some of the assumptions of how the law responds to new challenges are being asked.

This chapter will serve three main purposes. Firstly, it will cover some of the basics about the underlying architecture of the Internet. Secondly, it will discuss some of the research from complexity theory, and in particular from network theory, about how the Web works. Lastly, the chapter will look at the regulation of cyberspace with emphasis on the application of network theory to current regulatory ideas in order to make the case for its relevance. This will set the stage for later chapters, where specific regulatory challenges will be analysed using tools learnt from complexity disciplines. It is one of the stated hypotheses of this work that network science is of particular interest to the regulation of the Internet, and it will be put forward that network theory may prove to be a valuable descriptive and normative tool to help us understand these technologies.

1. THE INTERNET

Almost every book which deals with Internet regulation in one form or another contains a short history of the Internet. In danger of falling into redundancy, this work will not be an exception, although emphasis will be given to those aspects of the history of the Internet that are relevant to network theory.

In its most basic form, the Internet is a communications network made up of hardware and software which connects computers that fall under two types, hosts and routers. A host is simply any computer connected to the Internet via a modem, cable or a local-area network (LAN). There are two types of hosts, on the supply side we have servers, which are computers that have software designed to deliver content on demand, be it web pages, files, music, images, streaming video, email, etc. On the reception-side we have terminals and workstations; these are computers that have an Internet connection, but also software capable of receiving content, known as a client (e.g. browsers, mail clients,

instant messaging). In between these types of hosts there is a vast array of intervening gateways (or routers), whose main function is to route the information from the servers to clients. Right away, one can see the relevance of network theory to this set-up; the computers that make up the network are vertices, and the intervening connections are edges. The presence of the classic node and link structure of networks is indeed one of the reasons why the Internet is such a great space for studying network theory.

It might be superfluous to define the Internet; after all, we all use it on a daily basis and take it almost for granted. Nonetheless, these technical definitions serve the purpose of framing the Internet within network theory. The Internet Engineering Task Force (IETF) has defined the Internet as a network which contains several defining architectural characteristics.[6] Chief amongst these is the understanding that the Internet is "a network of networks", which means that it is made up of a vast array of sub-networks interconnected to one another through a global infrastructure, but most importantly, where all of these networks communicate using standard protocols.[7] These sub-networks are known as autonomous systems (AS) because they are in many ways self-contained, yet interact with the wider network through gateways.[8] Because it is a network consisting of millions of computers,[9] there is an inherent complexity in the way in which it is organised, a complexity that is managed through a system of routers. Another of its main characteristics is that it must tolerate network-wide variation, which means that it is also a dynamic network independent of changes in the intervening computers.

---

6. Internet Engineering Task Force, *Requirements for Internet Hosts: Communication Layers*, RFC 1122 (1989), http://tools.ietf.org/html/rfc1122.
7. Ibid, p.7.
8. Mahadevan P et al, "The Internet AS-Level Topology: Three Data Sources and One Definitive Metric", 36:1 *Computer Communications Review* 17 (2006).
9. By December 2008, the CIA World Factbook calculates that there are more than 500 million Internet hosts, that is, computers that serve some form of content, see: https://www.cia.gov/library/publications/the-world-factbook/rankorder/2184rank.html.

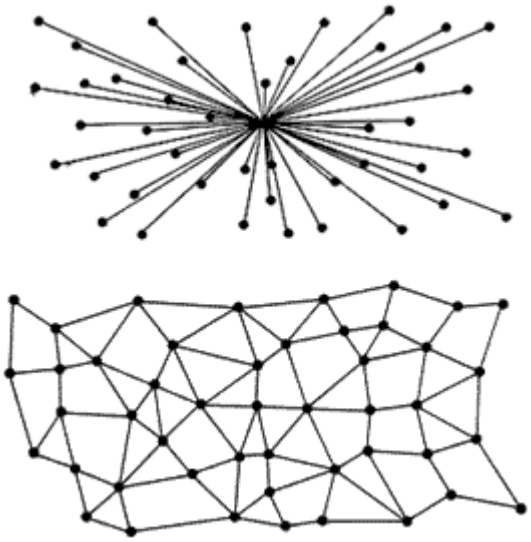


Figure 4.1 Central (top) and distributed (bottom) networks

This complexity has not grown overnight, and it is the result of decades of development. Most communication networks rely on centrality of communications in one form or another; for example, the telephone network is a good example of a system that relies on central connecting points from one end to another, known as exchanges. The reliance on centrality in these exchanges make them vulnerable to targeted attacks on the central hubs. During the Cold War, the US military was concerned about relying on such central communication networks, so researchers at the RAND Corporation came up with the idea of distributing intervening nodes in order to make the system less likely to suffer in case of a nuclear strike.[10] The idea was to break down messages into packets of information, and send those packets through a distributed network that would forward them to a number of nodes in the system instead of going through a central node (Figure 4.1). By building deliberate redundancy in the system, the network would rely on protocols capable of putting together the packets at the receiving end. This basic architectural solution, known as packet switching, explains some of the resilience

10. RAND, *Paul Baran and the Origins of the Internet*, http://www.rand.org/about/history/ baran.html.

exhibited by the modern Internet. Paul Baran, one of RAND's researchers, came up with most of the early ideas of packet switching and node distribution, and by 1962 there was a large military network in place which applied these principles.[11]

The next stages of the history of the Internet took place in academic institutions, although supported by military research funds. The modern Internet had its first application in 1969 through the connection of four institutions into what was known as the Advanced Research Projects Agency Network (ARPANET).[12] ARPANET had some interesting features that make it a direct predecessor of the global network; it used some of Baran's ideas about packet switching and node distribution but, most importantly, the information was routed through the network using computers known as Interface Message Processors (IMPs), which are precursors of the Internet routers that act as the backbone of the modern Web.

By the early 1980s there were hundreds of computers connected in this manner, but there was still one element missing, and that was the existence of common communication protocols that would allow information to reach from one point to another. Between 1984 and 1988, the European Organization for Nuclear Research (CERN) achieved a critical stage in the development of the Internet by implementing the Internet Suite, consisting of a collection of protocols such as the Transmission Control Protocol (TCP) and the Internet Protocol (IP), known collectively as TCP/IP.[13] TCP/IP is what makes Internet communication possible by ensuring the existence software packages that facilitate both network resilience and speed by allowing the existence of multiple paths from one point on the Web to the other.

The Internet Suite describes a number of applications, tools and layers that constitute what we know as the Internet, and it can be better understood as a collection of network

11. Naughton J, *A Brief History of the Future: The Origins of the Internet*, London: Phoenix (2000), pp.102–105.
12. Ibid, pp.213– 215.
13. Ibid.

layers that operate at all stages of transmission and reception. These layers are set out in official documents by the IETF[14] and are:

1. *Application Layer*: This is the top communication level made up of protocols for user applications such as sending mail (Send Mail Transfer Protocol SMTP), sending files (Hyper Text Transfer Protocol HTTP); it also includes protocols used for system support, such as that which identifies servers in the system (Domain Name System DNS).
2. *Transport Layer*: This provides end-to-end protocols for communication between hosts in the system, such as the TCP and the User Datagram Protocol (UDP).
3. *Internet Layer*: Because the Internet is a network of networks, every computer connected to it has to be able to find its components. The Internet Protocol fulfils this function, and is differentiated from the application and transport layers by the fact that it does not consist of instructions to reach a destination, or is used to make the actual communications, but it allows data packets to reach their destinations by allowing identification of participating computers based on their IP address.
4. *Link Layer*: The link layer consists of protocols that allow connection of a host with gateways and routers within a network, usually a large area network (LAN) (e.g. Ethernet protocols).

Of these layers, perhaps the most central components are the Domain Name System and the Internet Protocol. These are what allow a computer to know where to go when the address "www.google.com" is entered into a browser. Every computer connected to the Internet has a numerical Internet Protocol address. Web servers are no exception, these are computers which store and serve files, and have domain names assigned to that address. As well as connecting to an Internet service provider, a computer has access to a domain name server (DNS) which stores information of which domain name is

14. IETF, supra note 6, pp.8–10.

assigned to each address, allowing people to type these domains in their browser.[15] If you knew a server's address, it would possible to connect directly without having to type its domain name, but this would make the entire system unwieldy. The Domain Name System allows ease of use because it assigns specific IP addresses to a domain name, and the DNS servers hold the information and route communication requests accordingly. There is a hierarchy of authoritative DNS servers, at the top sit a number of computers known the Root Nameservers, which are housed by 13 top level institutions which propagate all information of who is who online. At the next level sit the top level domain names (.com, .org, .gov); then top level country domain names (.uk, .de, .fr), and then each internet service provider and sub-network usually has its own DNS servers.

With the adoption of the Internet Suite as the standard set of communications by the end of the 1980s, most of what we know today as the Internet infrastructure was already in place, and the only thing left was its wider public adoption. It is a common misconception that the Internet was invented by Tim Berners-Lee and Robert Cailliau at CERN in 1990,[16] what they did was to make use of the existing infrastructure and protocols and suggested the creation of pages of hypertext stored and distributed in hosts, which would be viewed in client software called a browser. What Berners-Lee and Cailliau invented was the World Wide Web, which is just one of the many Internet applications, although perhaps the most visible one. The WWW became extremely popular even in early days, yet it achieved mainstream recognition in 1994 with the launch of the Netscape graphical browser. The rest, as they say, is history.

2. THE LAWS OF CYBERSPACE

15. Mockapetris PV and Dunlop KJ, "Development of the Domain Name System", 25:1 *ACM SIGCOMM Computer Communication Review* 112 (1988).
16. Berners-Lee T and Cailliau R, *WorldWideWeb: Proposal for a HyperText Project*, internal CERN memorandum (1990), http://www.w3.org/Proposal.html.

Having a better idea of the basic underlying architecture of the Internet only tells us part of the story. By reading the above description, it is easy to see why the Internet lends itself to analysis through graph and network theory. The Internet is undoubtedly a complex dynamic network, and while grasping how data gets from one computer to another is crucial to gain an understanding of the regulatory solutions that apply to it, network theory has the potential of uncovering much more about its inner workings.

One of the first tasks when analysing Internet architecture through network science is to define terms. As discussed in the last section, the Internet is made up of three basic elements: hosts, gateways and the communication protocols between these. Translating this structure into graph theory, Internet hosts and gateways would be nodes, and the communication protocols would be links. Under some circumstances, vital hosts and gateways can operate as hubs in the network; an importance that will be wholly dependent on whether it is a server, or a central interconnecting router. This basic set-up allows us to "map" the Internet by looking at the interconnection nature of hosts. By doing so, the Internet takes on an almost organic look (Figure 4.2).

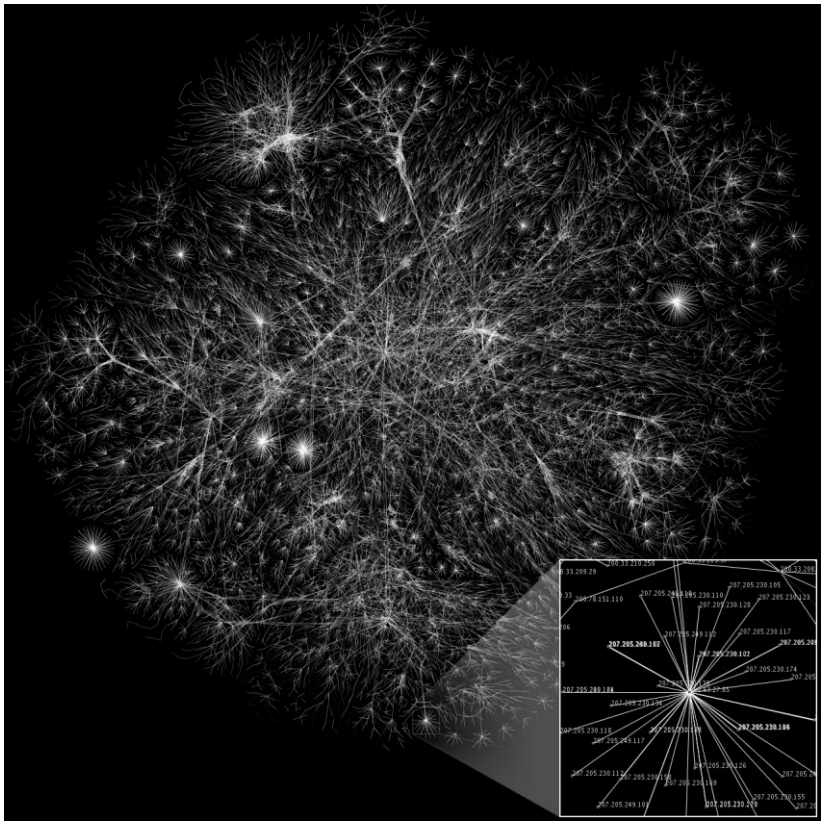


Figure 4.2 Map of the Internet[17]

When we map the Internet in this way, some things become apparent. Firstly, it is clear that the Web's architecture makes the centrality of some nodes crucial to the whole. Secondly, the Internet appears to be a scale-free network and not just a random network; this is because in a random network one would expect nodes to accumulate links and connections without apparent patterns, while in a scale-free network it is expected that some nodes will have considerably more links than others, as is the case in this picture.

17. The image is a partial map of the Internet using 2005 network data. Each line represents a link between two IP addresses, and the length of the line represents the network delay between those points. The images are from a visualisation program called the OPTE Project, by Barrett Lyon (released under a CC licence). For higher quality colour images, see: http://www.opte.org/maps/.

Does research corroborate this superficial reading? As it has already been suggested in Chapter 1, the answer is yes. Physicist Albert-László Barabási has been at the forefront of research into this area, and has become one of the main exponents of the growing interest in network theory partly due to his observations about the Internet. In 1999 Barabási published an influential article in *Nature* with Réka Albert and Hawoon Jeong detailing some of their findings in charting the network structure of the World Wide Web.[18] They programmed an autonomous agent that collected outgoing links from the indexed pages and reported the data back for analysis. They set out this robot expecting that the Web would display random distribution of links between nodes following the Erdős and Rényi random model of dynamic networks that had been prevalent until then. However, they found that the distribution of links followed a strong power law; in other words, the degree of distribution of incoming and outgoing links would be the same if one was to look at 20,000 pages, or just 20.[19] According to them:

> The power-law tail indicates that the probability of finding documents with a large number of links is significant, as the network connectivity is dominated by highly connected web pages. Similarly, for incoming links, the probability of finding very popular addresses, to which a large number of other documents point, is nonnegligible, an indication of the flocking nature of the web. Furthermore, while the owner of each web page has complete freedom in choosing the number of links on a document and the addresses to which they point, the overall system obeys scaling laws characteristic only of highly interactive selforganized systems and critical phenomena.[20]

This was a remarkable finding because in one swoop it identified the Internet as a scale-free network, and also conveyed that it had self-organising characteristics. Something in the Internet's architecture was organising web designers from around the world to link to more popular pages. They attribute the existence of power laws in the

---

18. Albert R, Jeong H and Barabasi A-L, "Diameter of the World-Wide Web", 401:6749 *Nature* 130 (1999).
19. Huberman BA, *The Laws of the Web: Patterns in the Ecology of Information*, Cambridge MA: MIT Press (2001), p.25.
20. Albert et al, supra note 18.

link structure of the Web to the "rich get richer" phenomenon, whereby older nodes in the network are more likely to accumulate links.[21] Moreover, when one thinks of cyberspace using Kauffman's fitness landscape theory, it becomes evident that popular nodes exhibit more fitness, and then we end up with fitness peaks where the more popular sites tend to accumulate more links, and the ones with less have limited growth. While other networks exhibit similar power laws, it seems like the "rich get richer" phenomenon is particularly suited for online growth, as link accumulation does not cost anything.[22]

In the same issue of *Nature*, another paper hinted at the answer of why the Web behaved in this way.[23] Adamic and Huberman were interested in the Internet's growth, and remarked that it appeared to follow power laws. Instead of looking at links, they looked at the number of pages at any given site. They remarked that one could accurately predict the number of pages in any random site without having to exhaustively use search engines in order to mine the information. Initially, they remarked that taking any random site and allowing for random accumulation of new pages did not produce a power law. However, when one entered into the equation a node's age in the network, the result did follow a power law. While this seems to be evidence of the "rich get richer" model exhibited by scale-free networks, more tweaking is needed in order to explain the power laws exhibited by the Internet. In another article responding to Albert and Barabási's "rich get richer" theory,[24] Huberman and Adamic commented that age was not enough to predict incoming link fitness within the network, as evidenced by the meteoric rise of popular sites regardless of their age in the system, such as Google.[25] This is a key point, and one that can be seen in Barabási's own

21. Barabási A-L, *Linked: The New Science of Networks*, Cambridge MA: Perseus Pub. (2002), pp.79–89.
22. Watts DJ, *Six Degrees: The Science of a Connected Age*, London: Vintage (2004), p.113.
23. Huberman BA and Adamic LA, "Growth Dynamics of the World-Wide Web", 401:6749 *Nature* 131 (1999).
24. Albert et al, supra note 18.
25. Huberman BA and Adamic LA, "Power-Law Distribution of the World Wide Web", 287 *Science* 2115 (2000).

admission in later works that Google did not seem to obey the same power law model as other sites did.[26] As Huberman and Adamic put it, "not all websites are created equal". One could call this a modified "rich get richer sometimes" model, whereby older sites generally accumulate links faster, but growth rates are not uniform. This offers a more accurate explanation of the power law features of the Internet. Huberman postulates that "a simple assumption of random multiplicative growth, combined with the fact that sites appear at different times and/or grow at different rates, leads to an explanation of the power law behaviour so prevalent on the Web".[27]

There is a wealth of research which seems to corroborate the scale-free characteristics of the Internet. Faloutsos, for example, took three snapshots of the Internet at router level, and uncovered scale-free characteristics between these central connecting nodes.[28] Similarly, Vespigniagni and Peracci conducted a survey between 2000 and 2002 of round-trip time around nodes by using PING (Packet InterNet Groper) data, similar to sending sonar signal that measures average Internet distance by sending packets of information to a destination and measuring how long it takes for them to get back to the sending machine. They found scale-free behaviour of these pings, which they considered surprising as they were not expecting it at all levels, as they found.[29]

In the spirit of fairness, there have been some criticisms from computer scientists against the emphasis on physics and mathematics in current network theory analysis of the Internet, with some authors claiming that the theoretical approaches should be followed by "real network" experiences.[30] Li et al are particularly scathing about the over-hype in scale-free modelling of Internet phenomena, and attribute potential methodological

---

26. Barabási, supra note 21, p.94.
27. Huberman, supra note 19.
28. Faloutsos M, Faloutsos P and Faloutsos C, "On Power-Law Relationships of the Internet Topology", 29:4 *ACM Computer Communications Review* 251 (1999).
29. Percacci R and Vespignani A, "Scale-Free Behavior of the Internet Global Performance", 32 *European Physical Journal B* 411 (2003).
30. Willinger W, Alderson D and Doyle JC, "Mathematics and the Internet: A Source of Enormous Confusion and Great Potential", 56:5 *Notices of the American Mathematical Society* 586 (2009).

biases that produce scale-free results.[31] Having said this, critics are in the minority, and the preponderance of evidence seems to lead us to a consensus which is overwhelmingly in favour of the scale-free Internet.

Another characteristic of the Internet is that is displays small world clustering. To refresh some of the concepts in Chapter 1, small world networks are those where any random vertices in the network can be reached through short intervening paths. If the Internet is a scale-free network where some nodes have disproportionate number of connectors, then one would assume that it does indeed display small world connectedness between links. Huberman[32] conducted a series of experiments trying to obtain the average links from among 64,826 sites. According to his findings into node length online, the average path between two random websites is as small as 4.22 links. The reason for the small path length is attributed to high clusters of individual websites that are connected to one another. While there are some websites with high-connectivity acting as hubs, once such vertices are reached then the paths are considerably reduced.[33] Interestingly, not only does the Internet itself displays small world clustering, but the social actors in the network are also responding to such grouping behaviour. Internet tools such as social networking offer some insights into small world communities of users. For example, a study into "friend lists" in the social network site Myspace found that a user's number of friends roughly responded to a power law.[34] This seems to indicate that both the predetermined element of the network, such as the backbone architecture, and the emergent elements, such as users, display high levels of clustering.

A third important aspect of the Internet with regards to network theory is that it is resilient. Cohen et al conducted an analysis of the Internet's connectivity trying to ascertain if the random removal of nodes from the network would have a knock-on

31. Li L et al, "Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications", 2:4 *Internet Mathematics* 431 (2005).
32. Ibid, pp.41–54.
33. Watts, supra note 22, pp.79–81.
34. Thelwall M, "Social Networks, Gender, and Friending: An Analysis of Myspace Member Profiles", 59:8 *Journal of the American Society for Information Science and Technology* 1321 (2007).

effect and disrupt the network as a whole.[35] Because the Internet is a scale-free network, they discovered that although the interconnection between nodes on the Internet would become more diluted as nodes were randomly removed, the network would remain essentially connected even approaching 100 percent node breakdown. However, Callaway et al explored directed attacks to the network, and discovered that although the Internet is highly resilient to random attacks; it becomes very fragile with the targeted removal of the most connected nodes.[36] These findings are perhaps straightforward when one understands the nature of scale-free networks and the role played by highly-connected hubs in the system, but they are vital for understanding the network's architecture as a whole, and are of particular relevance for various legal subjects that will be explored later.

Another interesting characteristic of the Internet unearthed by network theory has been the way in which information travels within the distributed nodes, and particularly how viral infections spread and remain in the system. The Internet has been designed with high connectivity between nodes in mind, which might explain both the small world clustering and its resilience to random attacks. However, the high connectivity relies on the prevalence of hubs – as we have seen these are hosts that are at the higher-end of the curve and could be said to be the glue that connects the network together. Vespignani and Pastor-Santorras looked at computer virus epidemics for a period of 50 months from 1996 to 2000, looking at the spreading and the survivability of the most virulent infections.[37] They found that the Internet is highly susceptible to fast viral spread because of the same high level of interconnectivity between nodes and, more importantly, infections are susceptible to a global pandemic if highly-connected hubs are infected. More worryingly perhaps, they found that despite the existence of antivirus

35. Cohen R et al, "Resilience of the Internet to Random Breakdown", 85 *Physical Review Letters* 4626 (2000).
36. Callaway DS et al, "Network Robustness and Fragility: Percolation on Random Graphs", 85:25 *Physical Review Letters* 5468 (2000).
37. Vespignani A and Pastor-Santorras R, "Epidemic Spreading in Scale-Free Networks", 86:14 *Physical Review Letters* 3200 (2001).

software and the deployment of updates designed to tackle specific outbreaks, viruses will remain in the network for an unlimited amount of time. While this model is particularly useful in describing the viral properties of computer viruses, the model can be applied to the spread of other types of information, such as viral marketing, or the existence of viral videos.[38]

Further research has been producing more relevant facts about the self-organising nature of the Internet's architecture. Broder et al conducted a massive crawl of 203 million nodes on the Internet.[39] Besides finding that the distribution of nodes followed a power law, they also discovered that apparently there was a core of highly connected nodes within the network, that they named the "giant strongly connected component" (GSCC) of the World Wide Web, consisting of 56 million pages. These were pages that could be reached by one another using directed paths. However, the surprising part of their research was that the rest of the studied pages consisted of pages that could be reached from the SCC but not reach the SCC; then there were those that could not be reached from the SCC but could reach it; and finally there were pages that could do neither and were isolated from this core.[40] What this tells us is that there is an inherent centrality of connected nodes that is of great consequence to the way in which information flows online, and also is relevant to some regulatory aspects of the Web.

All of this research gives us a better understanding of some of the governing laws underneath the structure of the Internet. It must be remarked that with the exception of the network's resilience, these architectural characteristics are not the result of conscious planning on the part of the Web's designers. It seems clear that the Internet exhibits self-organising features, but also that nodes and actors within the network are often presented with some unmovable and unchangeable inherent features within the system. Does the deterministic nature of cyberspace have regulatory repercussions?

---

38. Boase J, "A Plague of Viruses: Biological, Computer and Marketing", 49:6 *Current Sociology* 39 (2001).
39. Broder A et al, "Graph Structure in the Web", 22 *Computer Networks* 309 (2000).
40. Ibid, p.310.

3. REGULATING STRATEGIES

3.1 Technocracy

Who controls the Internet? Just by looking at the output from network theory, one would be tempted to answer that nobody does. Yet the existence of an ordered cyberspace cannot be denied – there are common communication protocols, a vibrant exchange of information between millions of servers and clients, and there is a tangible infrastructure of satellite connections, phone lines, fibre optics and local networks that allows computers from around to world to have access to a shared pool of data. While one can argue that the Internet has some self-organising characteristics, it would be disingenuous to claim that such order happens in entirely spontaneous fashion. There is a kernel of decision-making bodies that have shaped some of the architecture that is under study now.

At the heart of the Internet's governance, that is, who makes decisions about its architecture, we find a rather haphazard assembly of standard-setting organisations. The IETF is one such body, whose stated purpose is "to make the Internet work better".[41] Anyone can join the IETF, yet it operates through a complex network of workshops, thematic and regional working groups. Most of the executive technical work is performed by tighter and more exclusive groups of experts, amongst these are the Internet Engineering Steering Group (IESG),[42] the Internet Architecture Board (IAB),[43] the Internet Society (ISOC)[44] and the Internet Assigned Numbers Authority (IANA).[45] These groups have the collective responsibility of setting out new standards, tweaking and modifying existing ones, and proposing changes to the overall Internet architecture.

---

41. Alvestrand H, *A Mission Statement for the IETF*, RFC3935 (2004), http://www.ietf.org/rfc/rfc3935.txt.
42. http://www.ietf.org/iesg.html.
43. http://www.iab.org/.
44. http://www.isoc.org/.
45. http://www.iana.org/.

The World Wide Web Consortium (W3C) is another standard-setting organisation, but it deals specifically with the WWW. The W3C is a much more formal institution, it is in charge of the Web's standards, but it also issues technical guidelines for the management of the network. However, the W3C is not a legislative institution and it cannot compel members or states to adopt its recommendations. The W3C is a consortium of organisations, made-up mainly by multinational technology and telecoms corporations such as British Telecomm, AT&T, Adobe Systems, Microsoft and Nokia. It also has membership from governments and academia.[46]

The third institution with a large say over the Internet's architecture is the Internet Corporation for Assigned Names and Numbers (ICANN), which controls the DNS system.[47] ICANN controls and co-ordinates the domain name system by holding top-level control of the root nameserver system. ICANN is also responsible for accrediting the domain name registrars, which are the ones that operate each country's top level domain name system. ICANN is unique amongst the Web's governing institutions because it is constituted as a non-profit public-benefit corporation based in California, and was initially established by the US Department of Commerce. This state of affairs has led to some protests from other countries about what they see as excessive control of one of the central Internet governing bodies by one country, and has led to an attempt to overhaul the system through the UN World Summit of the Information Society (WSIS).[48] The summit failed to wrestle control from ICANN, but managed to set up yet another institution, the Internet Governance Forum (IGF), which handles mostly capacity-building and digital divide issues in developing countries.

What seems to come out from the regulatory picture at the governance level is that the Internet is not organised in a centralised manner, and that its operation is determined by a complex regulatory apparatus – one could even call it a technocracy. Bowrey

46. W3C, *About the World Wide Web Consortium*, (2007), http://www.w3.org/Consortium/.
47. See: Mueller M, *Ruling the Root*, Boston, MA: MIT Press (2004).
48. Hamelink CJ, "Did WSIS Achieve Anything at All?" 66:3-4 *International Communications Gazette* 281 (2004).

adequately comments that this structure of Internet governance seems more concerned with engineering and less with concepts such as accountability and responsibility.[49] Zittrain also remarks on the technocratic nature of the Internet's architecture as a result of financial constraints, but also as a conscious design effort built in the system to reflect the sensibilities of the system's creators.[50] Perhaps this is not such a bad thing if we believe the Internet to be simply a connection of nodes in a network, and perhaps the resulting governance layer that controls much of the network's architecture is simply trying to provide technical solutions to technical problems.

Nonetheless, the emerging picture of technocratic regulation of the Internet is one that is consistent with network theory. The basic architecture is set by the governance groups, but these have no real control about the actual growth of the network, or how it operates at a basic level. So despite the existence of some level of control, we are then left with a self-organising network with emergent characteristics, where vertices and edges cluster together following power laws and small world topologies. Complexity reigns supreme, but is cyberspace really an uncontrolled technical anarchy?

3.2 Cyber-libertarians

Given its technocratic origins, some of the earliest theories on Internet regulation advocated low intervention by external regulators. Particularly in the early 1990s, regulators were slow to respond to the challenges, and were very much taken aback by the potential of the new technology and the appearance of a global communications network that was completely unregulated and, most importantly, seemed to be immune from regulation. In an often cited work on the topic, lyricist John Perry Barlow wrote his famous (or perhaps infamous) *Declaration of Independence of Cyberspace*, in which he set out to attack government intervention in cyberspace, favouring a quasi-libertarian self-regulated approach. He wrote:

---

49. Bowrey K, *Law and Internet Cultures*, Cambridge, UK; New York, NY: Cambridge University Press (2005), p.50.
50. Zittrain J, *The Future of the Internet: And How to Stop It*, London: Allen Lane (2008), p.28.

> Cyberspace consists of transactions, relationships, and thought itself, arrayed like a standing wave in the web of our communications. Ours is a world that is both everywhere and nowhere, but it is not where bodies live. [...] Our identities have no bodies, so, unlike you, we cannot obtain order by physical coercion. We believe that from ethics, enlightened self-interest, and the commonweal, our governance will emerge.[51]

While it was not his intention, Barlow may have been talking about self-organisation, albeit a rather naïve version of it. He believed that Internet communities would be able to exercise self-regulatory control because governments would not be able to intervene. How wrong he was.

Barlow was eventually joined by other commentators and scholars who believed that it would be difficult to subject the Web to traditional regulatory methods. Other authors proposed similar theories which tried to explain that the Internet could not be controlled in any effective manner, and so proposed several models of self-regulation that would be able to organise the network in some coherent fashion.[52] Of note amongst these theories is Post and Johnson's Net Federalism.[53] In it, they argue that cyberspace is a separate entity with clear borders from the physical world, and consequently it should be treated as an independent regulatory sphere for all legal purposes. Because the Internet would still require some form of regulation, they argued that the Web should be able to assemble its own legal institutions in a manner similar to the creation of federal states brought together under a unifying ideal. These self-regulated federal states would generate their own sets of rules consistent with practice in that part of cyberspace. The most remarkable thing about this theory is that it is informed greatly by the author's

51. Barlow JP, *A Declaration of the Independence of Cyberspace*, (1996), http://homes.eff.org/~barlow/Declaration-Final.html.
52. A comprehensive review of most of these ideas can be found in Greenleaf G, "An Endnote on Regulating Cyberspace: Architecture vs Law?" 21:2 *University of New South Wales Law Journal* (1998), http://www.austlii.edu.au/au/journals/UNSWLJ/1998/52.html.
53. Johnson DR and Post DG, "Law and Borders: The Rise of Law in Cyberspace", 48 *Stanford Law Review* 1367 (1996).

work on complexity theory showcased in Chapter 2. Particularly, they see the emergence of self-regulatory spheres as a prime example of fitness landscapes, where norms emerge in self-organising patches of order. They comment that:

> We have suggested elsewhere that the Internet calls for a higher degree of deference to rulemaking within a-geographical, decentralized, voluntary associations, and we believe that [chaos theory] provides normative underpinnings for this view. Allowing individuals to define the boundaries of their own, a-geographical patches by voluntary movement into, and out of, decision-making bodies that have little, or even no, tie to particular physical location – what we might call "self governance" – may allow both more rapid, and more "congruent," responses to shifts in spillover patterns.[54]

While this is a persuasive use of complexity theory to try to reach a comprehensive solution to the perceived problems of Internet regulation, Post and Johnson completely underestimated the regulatory push from governments and international organisations that would take place just after they had written their ideas.[55] Even back in the late 1990s, several authors criticised the cyber-libertarian ideas of unregulated spaces. Boyle[56] in particular seems to have understood that the premise behind the theories of the impossibility of exercising any credible governance over cyberspace were not only wrong-headed, but rested on completely untested hypotheses. In his view, cyber-libertarianism was blind to the many avenues of control available to public regulators.

Nonetheless, not all cyber-libertarian ideas were proved wrong. In 1993, John Gilmore, cyber-activist, programmer and one of the founders of the Electronic Frontiers Foundation (EFF), was quoted in Time Magazine as saying that "The Net interprets censorship as damage and routes around it".[57] This seemingly innocuous quote has

---

54. Post and Johnson, supra note 49, p.1087.
55. Just to name a few in the area of copyright: The 1996 World Intellectual Property Organization (WIPO) Copyright Treaty; the US 1998 Digital Millennium Copyright Act (DMCA); and the Directive 2001/29/EC of the European Parliament and the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.
56. Boyle J, "Foucault in Cyberspace: Surveillance, Sovereignty, and Hard-Wired Censors", 66 *University of Cincinnati Law Review* 177 (1997).
57. Elmer-Dewitt P, "First Nation in Cyberspace", 49 *Time Magazine* (3 December, 1996), http://www.chemie.fu-berlin.de/outerspace/Internet-article.html

probably been one of the most remarkable aspects of online information, and it still holds true to this day as evidenced by the existence of the so-called Streisand Effect. In 2003, actress Barbara Streisand sued a photographer to try to remove aerial pictures of her home, which resulted in more people visiting the offending site and copying and republishing the picture than would be normally expected had she not initiated legal action.[58] The effect has been proved time and time again. For example, in December 2008 the Internet Watch Foundation (IWF)[59] blacklisted a Wikipedia article featuring the cover art for the Scorpions 1976 album Virgin Killer because it was flagged as child pornography. The result was that customers of several Internet Service Providers (ISPs) in the UK could not properly access the online encyclopaedia. The censorship attempt backfired spectacularly; the Virgin Killer page averaged 500 visits during the months previous to the event, but at the peak of the scandal, the page had received 126,000 views in a single day.[60]

The reason why this piece of cyber-libertarian lore is relevant to complexity theory is that it is a clear example of the resilience of scale-free networks. Isolated attempts to bring down a node are likely to fail. However, there is a more crucial link to complexity, as any given node on the Internet has N number of incoming connections, so it is likely that the information contained in that node will have been replicated and spread along the network quickly. Moreover, concerted efforts to shut down one node may prompt it to acquire more links, and so the chances that the information contained in the node will increase exponentially as a function of the incoming links. There is research that supports this assumption. For example, Wu et al have been looking at how information spreads in scale-free networks, particularly in closed circles of acquaintances, and have discovered that there are certain thresholds after which a given link can be said to have

58. Bernoff J and Li C, *Groundswell: Winning in a World Transformed by Social Technologies*, Boston MA: Harvard Business School Press (2008), p.7.
59. The IWF is a UK industry watchdog that identifies objectionable content and passes IP addresses to be filtered by UK service providers.
60. Guadamuz A, "Censorship UK", *TechnoLlama Blog* (8 December, 2008), http://www.technollama.co.uk/censorship-uk.

gone viral and spread rapidly amongst closed groups.[61] In the case of filtered information, any incoming link will increase the chances of that information being spread through the network.

It would then be possible to postulate a model for the Streisand Effect that goes something like this: any average page has an average number of incoming links; a specific attack on that node will prompt others to pay attention, increasing as a result the number of incoming links; at some point the number of incoming connections enters a phase transition, and the replication will increase following a power law.

3.3 Architecture and Code

Needless to say, other than the example of the Streisand Effect, cyber-libertarianism seems destined to languish as an interesting footnote in the history of Internet regulation. By the turn of the century, new regulatory explanations had come up to replace the libertarian approach. At the forefront of many of these studies has been Lawrence Lessig. In his influential book *Code and Other Laws of Cyberspace*[62] he postulates that there are four main types of regulation in an online world: markets, norms, law and architecture (Figure 43).[63] Most theories of regulation up until then accounted for the first three. Lessig's breakthrough came in the way in which he rightly identified the prevalence of architectural regulation in technological settings. Lessig argued that the Internet itself is highly dependent on the technological architecture that sustains it, the "code" in which it is written, the connectivity layers between domains, the protocols used in order to distribute information from one computer to another, the functional layers of the said protocols, the domain name server system that indicates one computer's location in the system, and so on.[64] Whether the Internet can be subject to regulatory control will depend entirely on its underlying architecture. For example, some

61. Wu F et al, "Information Flow in Social Groups", 337:1-2 *Physica A: Statistical and Theoretical Physics* 327 (2004).
62. Lessig, supra note 4.
63. Ibid, p.88.
64. Ibid, pp.100–102.

of the constituent code of the Internet is open, that is, it can be inspected, copied and modified by all sorts of people. This code could not be subject to government regulation. However, the protocols and communication tools that make up the online world are more critical than the underlying code because they are needed for connectivity to take place. So whoever controls the underlying "pipeworks", and the protocols, controls the Internet.

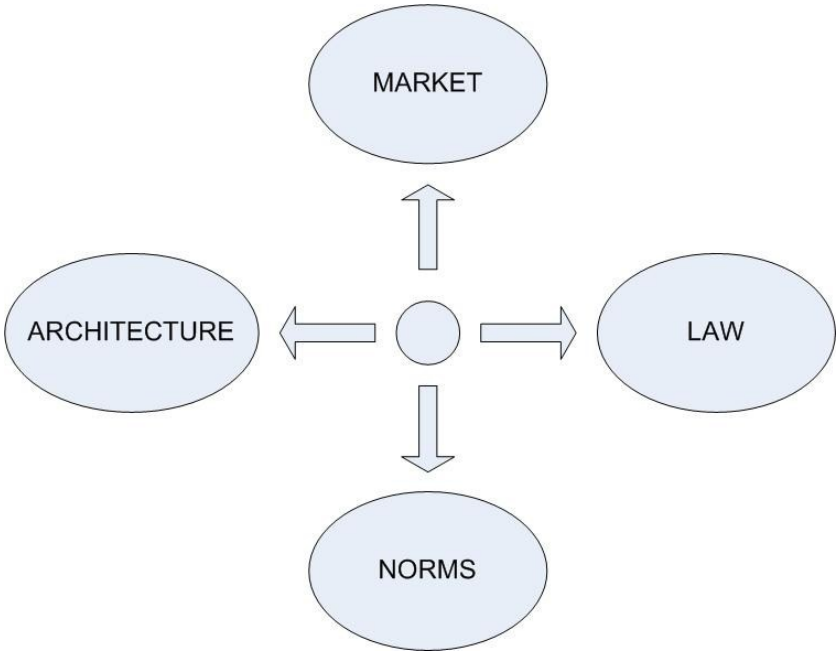


Figure 4.3 Lessig's regulatory matrix[65]

Although he does not go into self-organisation and complexity as such, Lessig's architectural regulation suggests the existence of some form of self-ordering mechanism. He identifies that there is an invisible hand of cyberspace that exerts an ordering force into the architecture of the Internet:

65. Ibid (image released under a Creative Commons licence).

> Control. Not necessarily control by government, and not necessarily control to some evil, fascist end. But the argument of this book is that the invisible hand of cyberspace is building an architecture that is quite the opposite of its architecture at its birth. This invisible hand, pushed by government and by commerce, is constructing an architecture that will perfect control and make highly efficient regulation possible.[66]

Nonetheless, Lessig's version of the invisible hand of cyberspace is limited, as he believes that it is shaped by code. So, programmers, regulators and policymakers can make conscious decisions that shape what the underlying architecture will look like, hence exercising real control over the shape of the Web.[67] This version of self-organisation is as a result limited by conscious decisions, and while cyberspace may reach its own efficient regulation, it can be subject to change.

It is possible to think of a modified version of Lessig's Code that responds better to what we are beginning to understand about the Internet. In this model, programmers, regulators and policymakers do make conscious decisions that shape what the underlying architecture of the Internet, but these decisions are in turn limited by the underlying laws of the Web described by authors like Barabási and Huberman. In other words, programmers set the rules, but the web self-organises around these rules.

The basic structure of the Internet expressed in the Internet Suite gives us a clear case of a conscious architectural decision. However, the resulting characteristics of the Internet, such as its resilience, the existence of small world pathways, and the almost universal presence of power laws at all levels of the network are not a result of conscious decisions. They happen as a result of the architecture, but their existence does not stem from the will of its designers. The network is created, but it responds to network theory because of deterministic reasons.

---

66. Ibid, p.7.
67. Ibid, pp.106–108.

3.4 Regulating the gateways

While Lessig's Code is a prime example of what could be considered a golden age in the study of the regulation of the Internet,[68] as the technology matured, so did regulatory solutions. The rise of Napster in 1999, and the later emergence of peer-to-peer (P2P) file-sharing networks,[69] served as clear reminders of the difficulties of enforcing the law in the digital domain. The almost interminable source of illicit materials online, coupled with the widespread availability of infringing content, gave the public the impression that as far as the Internet was concerned, everything went. Nonetheless, despite the glaring failure in shutting down file-sharing networks, the early years of the 21st century witnessed the deployment of relatively successful regulatory approaches by many national governments.

The Internet regulatory landscape up until around 2000 was a mixture of cyber-libertarianism, half-hearted legislative solutions and code. The Internet was a global, distributed and borderless network because it had been designed like that. It also displayed scale-free resilient characteristics because its origins as a military network favoured the rerouting of damage to one node by distributing communication throughout its backbone. Castells describes this as "architecture of openness".[70] Vint Cerf, one of the fathers of the modern Web, went as far as stating that the Internet traffic was "totally unbound with respect to geography".[71] However, as Goldsmith and Wu rightly point out, this initial architecture was not entirely set in stone, and unsurprisingly, it soon became clear that national governments were attempting to draw borders in cyberspace.[72] The

68. Other works of note are: Reidenberg J, "Lex Informatica: The Formulation of Information Policy Rules through Technology", 76 *Texas Law Review* 553 (1998); and Boyle J, *Shamans, Software, and Spleens: Law and the Construction of the Information Society*, Cambridge, MA: Harvard University Press (1996).
69. See: Smith S, "From Napster to Kazaa: The Battle over Peer-to-Peer Filesharing Goes International", *Duke Law & Technology Review* 8 (2003).
70. Castells M, *The Internet Galaxy: Reflections on Internet, Business, and Society*, Oxford: Oxford University Press (2001), p.26.
71. As cited by Guernsey L, "Welcome to the Web. Passport, Please?" *New York Times* (15 March 15, 2001), http://www.nytimes.com/2001/03/15/technology/welcome-to-the-web-passport-please.html.
72. Goldsmith and Wu, supra note 5, p.58.

most successful attempt to do just that is the segregation of the Internet into national intranets. While the Internet was supposed to be globally distributed, several countries started redesigning the entry points into their national networks in order to impose screening mechanisms that would allow them to filter out undesired content if necessary.

This state of affairs is a logical result of the manner in which the Internet grew. While the global architecture of the Internet as a distributed network still holds true because of the existence of routers and distributed protocols, the actual physical Internet is often centralised. In the early days of the Internet, a lot of information was spread through the telephone network, which ensured its high distribution ratio; albeit it was rather expensive.[73] Later, a high-speed backbone had to be built to accommodate larger amounts of information being spread throughout the system, as the network relied on cables and satellite in order to operate, and later on optical cables.[74] The end result was a more centralised Internet than was originally envisaged (Figure 4.4), as the router distribution worked within connected nodes. This can be explained best using Britain as an example: the country has a large number of roads, but not being connected to continental Europe, it relies on ports and airports as communication hubs. The modern Internet looks something like that, with physical connections akin to ports where most of the information comes through, and then it is distributed using routers and hosts in the manner in which it was intended. What many countries have been doing is to reduce the number of physical entry points to their countries, sort of creating chokepoints on the Internet. If a government controls these gateways, then it will be easier to exercise control over the Internet in that particular country as a whole.

---

73. Some early networks, such as Fidonet, even sent all of their packets during cheap-rate calling times, see Naughton supra note 11, p.190.
74. Leiner B et al, *A Brief History of the Internet*, Internet Society Paper (2000), http://www.iicm.tugraz.at/thesis/cguetl_diss/literatur/Kapitel02/References/Leiner_et_al._2000/brief.html.

Figure 4.4 Map of the global Internet backbone[75]

The best example of this is the so-called Great Firewall of China, known in China as the Golden Shield Project. The Great Firewall is a multi-layered technological solution that takes advantage of the fact that the Chinese government controls the few Internet gateways into the larger Chinese Internet. This allows them to impose effective filtering restrictions to incoming Internet traffic by various means. The most crucial is the filtering of IP addresses originating from blacklisted services, which range from Blogger to Sex.com.[76] While this is in no way a perfect system, it does allow the Chinese government a level of influence that was thought would not be possible with the distributed architecture. The Great Firewall works by deploying hardware routers at each of the entry points into the country. These routers are given lists of banned IP addresses, so when an Internet host within China makes a request to access a banned site, the router does not forward the request to the target host, so the site appears not to exist, and returns a network error message to the client.[77]

It has become clear then that the most effective regulatory solution to online content is to exercise control at the access points. This regulation model has been replicated in

75. This image shows the global backbone from one bandwidth provider, Verizon (reproduced with permission): http://www.isp-planet.com/resources/backbones/vz_business.html.
76. Goldsmith and Wu, supra note 5, p.92.
77. Ibid, pp.92–94.

many other countries,[78] proving that the Internet is decreasingly distributed, and looks more like self enclosed city states with some intervening connecting ports.

It must be pointed out that the regulation at the gateway level has a lot of relevance for network and complexity theories. The first interesting effect is that the growing balkanisation of the Web has resulted in a networked federation reminiscent of Post and Johnson's Net Federalism, but this is not a self-regulated utopia as envisaged by them, but a tightly controlled collection of regulatory patches that have achieved stable fitness landscapes. Secondly, these national webs tend to exhibit large clustering characteristics which make them more likely to exhibit small world topologies. Research seems to validate this idea; Zhou, Zhang and Zhang conducted a study into the Chinese Internet at the AS level, and discovered that the internal topology of the sub-network mirrored that of the wider Internet, which hints at the presence of a fractal or self-similar Internet where the component sub-networks have the same characteristics as the whole.[79] In fact, the study also found "rich get richer" characteristics, as well as small-world path lengths, which seem to further the idea that these laws of the Internet are universal.

It might be easy to miss the monumental importance of this finding. Here we have evidence that points towards the existence of universal rules that apply to the network at all levels, one of the very definitions of scale-free topologies. Moreover, these similarities are replicated even behind national firewalls. It is possible that the distributed nature of the Internet protocols favours the prevalence of scale-free characteristics at all levels. The relevance for regulation theories is that whenever a government tries to cut-off and/or filter the network, what it is doing is simply creating a small version of the wider network with the very same characteristics of the larger one. The Internet is indeed fractal, a fact that seems to be ignored by regulators all over the world.

---

78. For a comprehensive survey, see: Zittrain J et al, *Access Denied: The Practice and Policy of Global Internet Filtering*, Cambridge, MA: MIT Press (2007).
79. Zhou S, Zhang G-Q and Zhang G-Q, "Chinese Internet AS-level Topology", 1:2 *IET Communications* 209 (2007).

3.5 Complex regulatory networks

It would seem that of the Internet regulatory strategies that have been proposed in recent years, it is the architectural solutions such as coding regulation at the gateways that have gained prevalence in recent years. But what about the regulatory bodies themselves? Is it possible that they constitute a complex system that could be analysed through network theory?

Andrew Murray has given some valuable input to the literature dealing with cyberspace governance by suggesting that regulation theories should concern themselves with the actors in the regulatory landscape. One of his main ideas is to draw a matrix of regulatory relationships which paint a picture of complex regulatory networks. Murray proposes that we look at how regulatory systems evolve due to internal and external forces, suggesting that such evolution represents a complex system.[80] Starting from Lessig's dot at the centre of regulation, Murray turns it around and paints the dots as the regulators, the actors in the system. He then draws association lines between each of the actors in order to illustrate the complex relationships that shape specific regulatory landscapes. Murray comments:

> Thus where regulators vie for regulatory acceptance they do not act in a regulatory vaccum, any action by any one member of the regulatory matrix (either as regulator or regulatee) has an effect on the actions of the others. This is because all regulators and regulates form part of an environmental system and a change in any one aspect of this environment affects all who participate in that environment […] At each point in the regulatory matrix, a regulatory intervention may be made, but the complexity of the matrix means that it is impossible to predict the response of any other point in the matrix.[81]

---

80. Murray A, "Conceptualising the Post-Regulatory (Cyber)State", in Brownsword R and Yeung K (eds), *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes*, Oxford: Hart (2008), pp.288–289.
81. Ibid, p.301.

Using ICANN as an example (Figure 4.5), A would be the US Department of Commerce, which unilaterally created ICANN. The World Intellectual Property Organisation (WIPO) is B, domain name owners are C, the European Union is D, and WSIS is E. All of the actors interact with one another trying to exercise influence over the regulatory matrix through various actions. The matrix becomes exponentially more complex as new actors and new interactions are added, which paints an accurate picture of just how complex the regulatory system has become.[82]

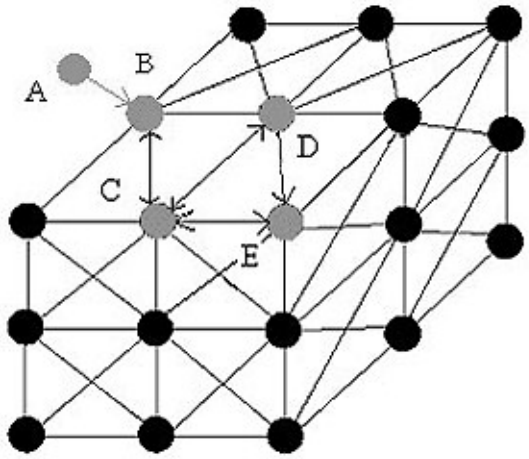


Figure 4.5 Murray's ICANN regulatory matrix[83]

This regulatory complexity does not mean that regulation is not possible, but that it features levels of interaction that are not explained by other models, such as the solitary dot in Lessig's *Code*. This interaction creates a dot community of regulators and subjects, where the actors can be one or the other interchangeably.

Something that is quite striking in Murray's regulatory matrix is that it is reminiscent of similar dynamic systems that are studied both by complex and network theory. For

82. Murray AD, *The Regulation of Cyberspace: Control in the Online Environment*, Milton Park, Abingdon UK; New York, NY: Routledge-Cavendish (2007), pp.23–237.
83. Ibid, p.236 (reproduced with permission from the author).

example, Erdős and Rényi were faced with similar complex dynamic situations when looking at information paths, and Solomonoff and Rapoport also encountered complex systems in biological networks.[84] More importantly, complex regulatory networks are reminiscent of the problems encountered by Kauffman when looking into genetic networks.[85] In all of these situations, the solution to unravel the complexity arising from dynamic interactions is to make assumptions as to the number of connections that any given node has, as well as the level of influence exerted by such networks. Kauffman's NK model could be used to try to describe the level of complexity in any given regulatory system, where N is the number of actors, and K the measure of the complexity in the system. Because we do not know exactly what the relationship paths within the regulatory network are, it would be feasible to assign values to each of the nodes, hence describing the overall complexity of the system in numeric form.

There is another model dealing precisely with influence in complex systems, that of Sola-Pool and Kochen.[86] They were trying to arrive at estimates of how many people are influenced by one another, and provided several computational solutions that eventually ended in high levels of clustering within social networks. The relevance to regulatory networks can be found in their conclusions, which describe similar network matrices than those theorised by Murray. For example, in a random group of 1000 people, they started with a small cluster of friends, A, B, C and D. Starting only with B, they assigned random links by assuming that B met with 100 number of people in f days, but including A. They repeated the operation with a number of people chosen from the wider pool of 1000, and discovered that there were few people with only one acquaintance in common with A, and that there were even fewer with many acquaintances in common with A. This is expected from later research into small worlds, but the relevance for regulatory networks is that while it is not possible to chart precisely how many contacts a regulatory entity has, it would be possible to try to determine the average number of

84. See Chapter 2.1.
85. See Chapter 2.4.
86. Sola-Pool I and Kochen M, "Contacts and Influence", 1 *Social Networks* 5 (1978).

interrelations within the network. By assigning random numbers of connections between the actors in the regulatory matrix, it would be possible to untangle its complexity by explaining that the actors will probably have fewer connections in the overall network.

4. TOWARDS A REGULATION THEORY OF THE SELF-ORGANISING INTERNET

One of the main hypothesis presented in this work is that the Internet is a complex network that displays self-organising characteristics. This seems to be an incontrovertible fact if one reads all of the evidence coming out of the existing research highlighted above, and while some of the details can be argued over, readers will have to forgive this categorical statement. Nobody set out to organise the Internet in a manner that would display small world pathways between the billions of pages and links that make up the Web; no single organisation designed the Web in a way that it would show scale-free characteristics. While it has been remarked that resilience was built into the system in order to withstand attacks, this seems to be the only truly conscious feature of the Internet; all of its other architectural traits have been shaped by the invisible hand of cyberspace, as Lessig calls it. For that reason, the theories of regulation that have been described have to be seen in the light of this self-organising reality. With that in mind, one could pose a theory of Internet regulation, the self-organising Internet. This is hinted at by other theorists – Post and Johnson[87] are probably some of the theorists that seem closer to it, but their version requires some tweaking.

In order to refresh some of the concepts of self-organisation studied in Chapters 2 and 3 relevant to the issue of Internet regulation, it should be pointed out that complex adaptive systems tend to become stable due to internal features within the system that allow organisation to occur. Self-organisation arises as a stabilising force that turns chaos into order because complex systems favour stability. The Internet becomes

87. Post and Johnson, supra note 53.

organised because of the interaction of its parts favours clustering and stability in order to manage complexity. But what are the parts of this global telecommunications network? On the one hand, we have the technical components: the nodes, hubs and links made up of computers, servers, protocols, links and connections. On the other hand, we have the social part of the network: the actors that design pages, the decision-makers and the users. All of these come together into a self-organising force with human and machine elements that resembles a cyborg. In complex adaptive system terms, the technical network is the predetermined system of steady connections, while the social element is the emergent system consisting of interchanging and dynamic connections.

The social (emergent) element exhibits self-organising attributes because there is no centralised body that directs the eventual stability of the system. Granic and Lamey explain the human element of this self-organising force thus:

> Who runs the Net? Who or what organizational body is responsible for maintaining its various nodes and improving its efficiency? The answer to these questions points to one of the most interesting aspects of self-organization: complexity emerges spontaneously from the interactions of the simpler components of a system. There is no 'central control station'. The Internet is a vast, coherent system not as a result of some brilliant inventor's design or some governing body's regulations, but because of the critical mass of millions of users who electronically interact daily, setting the conditions for the spontaneous creation of a higher-order complexity. This is the same decentralized, emergent order exhibited by flocks of birds, colonies of ants and angry mobs.[88]

On the other hand, we have the technical (predetermined) element of the self-organising picture. In the words of Andersen, the Internet is a "technical autopoietic system"[89] where the computerised elements consisting of links, servers, computers and networks replicate and organise themselves despite human interaction. It is, however, essential to emphasise that while this is a technical complex system there are imperative

88. Granic I and Lamey V, "The Self-Organization of the Internet and Changing Modes of Thought", 18:1 *New Ideas in Psychology* 93 (2000), p.98.
89. Andersen PB, "WWW as a Self-organizing System", 5:2 *Cybernetics & Human Knowing* 5 (1998), p.38.

human elements, what Fuchs calls a socio-technical system where the technical nature simply enables the self-organising nature of human interaction.[90]

Whatever its nature, simply describing the Internet as a self-organising systems is of little use to theories of regulation unless we can understand better how this self-organisation takes place, and how it is relevant to the question of how to regulate the Internet. Otherwise, we are simply stating the obvious without adding any analytical insights into how we regulate complex systems.

This is a trickier challenge than might appear at first glance. It is tempting to remain descriptive when it comes to the Internet, and to assume that if it features self-organising characteristics, then there is little else we can do to change it. Whatever decisions are made about online networked environments, the hidden organising elements within the system will work against all of our efforts to regulate. Just relax, sit back and watch the Internet do its thing.

While tempting, such an approach seems both unimaginative and cowardly, but may prove to be realistic. When talking about autopoiesis in regulation and governance structures, Luhmann, for example, did not believe that it is possible to exercise governance in autopoietic systems. He saw such attempts as futile exercises because an autopoietic system organises itself in order to reduce internal complexity, and thus regulatory efforts are doomed to fail.[91] However, it is also possible to take another view, one that believes that regulation is possible even in such systems and that self-organisation is simply an obstacle to work around; thus regulation can be reactive or proactive to the autopoietic organising force.[92]

These are what I call the deterministic and the optimistic views of self-organisation in regulation theories. Regardless of which one of these two views one favours, the first step has to be taken in recognising the self-organising nature of cyberspace, and to

---

90. Fuchs C, "The Internet as a Self-Organizing Socio-Technological System", 11:3 *Cybernetics & Human Knowing* 57 (2005), p.58.
91. Luhmann N, *Social Systems*, Stanford, CA: Stanford University Press (1995), p.67.
92. Engel C, *Governing the Egalitarians from Without: The Case of the Internet*, Max Planck Preprint Paper 2003/10 (2003), p.40, http://ssrn.com/abstract=462485.

identify the areas that are more likely to display some of the characteristics of complex adaptive systems. Therefore, the very act of describing the Internet as a self-organising system is in itself a principal regulatory insight. I can only hope that this step has been fulfilled already. So, what next?

Here it all depends on whether one is a determinist or an optimist. If one is a determinist, then there is not much more that a theory of the self-organising Internet can do other than to describe how the Internet operates. If one is an optimist, then the task is more difficult. The first step is to answer some simple questions about why and how we regulate cyberspace. Do we want complete control over the Internet? Do we want architectural control over the technologies and standards that make it? Or do we want to inform policymakers so that they can better deploy their regulatory tools? It will be assumed that complete control is out of the question. This is another categorical statement, but it is hoped that the evidence already presented, and that which will be explored in further chapters will serve to make this point stand on its own merits.

The answer of what constitutes a useful theory of regulation then may rest on informing programmers and policymakers in better ways so that they can deploy better architectural solutions, an informed code if one may. This information will allow for better legislation and more realistic policies, and will allow better understanding of regulatory decisions that take into account the complex and adaptive nature of the Internet.

Take digital copyright for example. Copyright in the online environments has become one of the hottest topics surrounding internet regulation, and governments have been making serious attempts to curb piracy by deploying legislation. Content owners have been similarly interested in developing technical solutions in the shape of technological protection measures that restrict copying of digital works.[93] By knowing how copyright networks operate, content owners and legislators can develop better solutions, taking the

93. Westkamp G, "Digital Rights Management, Internet Governance and the Autopoiesis of Modern Copyright Law", 7:4 *Contemporary Issues in Law* 317 (2005).

view of course that regulation is possible. Similar solutions can be deployed in other areas subject to Internet regulation, such as privacy in social networks, the growth of user-generated content, network neutrality and cybercrime, just to name a few. The main task ahead is to take an optimistic approach and state that it is possible to regulate online environments despite self-organisation. The following chapters will attempt to do just that.

Before moving to other topics, it is important to make a quick distinction about what we are talking about with regards to online regulatory structures. The topic of self-organisation has not bypassed theorists of regulation. As stated above, Lessig, Post and Johnson, and Murray heavily hint at self-organising features of online environments. Similarly, Benkler,[94] with his concept of peer-production, and Zittrain,[95] with his concept of generativity, have been providing impressive theoretical frameworks that deal with the very same self-organising phenomena that have been suggested in this chapter. While some of their ideas will be dealt with in more detail later, it is important to point out that some of the approaches to self-organisation have been dealing mostly with the description of features inherent to Internet content. While their importance will become relevant later, what we are trying to do here is to frame the regulatory aspects of the Internet from a network theory standpoint.

94. Benkler Y, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven, CT ; London: Yale University Press (2006), p.256.
95. Zittrain, supra note 50, p.7.

5. Copyright Networks

> You can't shut us down! The Internet is about the free exchange and sale of other people's ideas!
> Futurama[1]

Pirates have become an unavoidable feature of popular culture. Just how cut-throat mercenaries, thieves and scoundrels were turned into romantic swashbuckling heroes is hard to determine, but from Long John Silver to Errol Flynn and Johnny Depp, the figure of the hardened-yet-lovable rascal is a powerful archetype in our collective minds. How precisely piracy was turned into an equivalent term to describe copyright infringement is much easier to follow. The use of the word dates back to almost 300 years, when French authors began using the term to describe those pillaging their work.[2] The term was already in widespread use when Mark Twain used it to fight those who were copying his works, and were engaged in what he described "pure robbery".[3]

Whichever its origin, copyright piracy has been at the forefront of copyright owner's concerns for centuries; a worry that has spawned the wide-ranging system of copyright enforcement that we know today. However, copyright has not been faring well on the Internet. As Nicolas Negroponte stated:

---

1. *I Dated a Robot* (2001).
2. Febvre L and Martin H-J, *The Coming of the Book: The Impact of Printing, 1450–1800*, 2nd Edition, London: Verso Classics (1997), p.293.
3. "TWAIN'S PLAN TO BEAT THE COPYRIGHT LAW; Will Run Autobiography in New Editions of His Old Works TO PUT PIRATES TO ROUT His Task as a Lobbyist Finished, So He Will Return to New York To-day", *New York Times*, (12 December, 1906), http://bit.ly/9bYh1G.

> In a digital world, the bits are endlessly copyable, infinitely malleable, and they never go out of print. Millions of people can simultaneously read any digital document - and they can also steal it.[4]

The potential and reality for widespread copyright infringement online has been named as a cause for the alleged drop in sales experienced in some content industries in recent years, particularly as claimed by the music industry.[5] It is in response to this perceived threat that a wide-ranging legislative effort has been deployed in order to curb piracy.[6] These have been comprehensive attempts at trying to regulate copyright in digital environments, and the reasoning behind such legislative solutions has been rarely challenged in policy-making circles. But despite these efforts, piracy is not only rife,[7] but it appears to be immune to legal challenges.

The pervasiveness of online file-sharing can certainly be attributed to the fact that it is difficult to compete against free products, and many Internet users will prefer to obtain content by downloading from peer-to-peer (P2P) networks instead of purchasing works protected by copyright. But this alone cannot explain the astounding resilience of file-sharing networks. Over and over again these services are defeated in court,[8] but as soon as one service falls another one is waiting to pick up its users.

Perhaps the explanation for this seemingly regulatory failure rests on some of the issues explored in the last chapter. As the Internet is a complex network, would it be possible that central elements in the system, such as content, copyright regulation and

4. Negroponte N, "A Bill of Writes", *Wired 3.05* (May 1995), http://web.media.mit.edu/~nicholas/Wired/WIRED3-05.html.
5. British Phonographic Industry, *Impact of Illegal Downloading on Music Purchasing*, BPI Paper, http://www.ifpi.org/content/library/The-Impact-of-Illegal-Downloading.pdf.
6. Just using treaties and laws affecting the United Kingdom, in the last years there has been the WIPO Copyright Treaties 1996; the Information Society Directive 2001/29/EC; The Copyright and Related Rights Regulations 2003 SI No. 2498 and the Digital Economy Act 2010.
7. Using industry figures again, the International Federation of the Phonographic Industry (IFPI) claims that 21 percent of people living in the top European markets (21 percent) "are engaged in frequent unauthorised music-sharing". See: IFPI, *IFPI Digital Music Report*, (2010), p.19, http://www.ifpi.org/content/library/DMR2010.pdf.
8. Most prominently in *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 125 S. Ct. 2764.

file-sharing networks, are actually responding to the self-organising nature of scale-free networks? Is the resilience displayed by P2P networks explained by network theory? Is it possible that business models in the content industries are actually designed with a different type of network in mind?

This chapter will try to answer these questions in two ways. Firstly, it will look at the copyright industries from a network theory perspective. Then, it will look at the P2P networks in the same light. It is hoped that by looking at the shape of the networks, some answers will begin to emerge. Particularly, the puzzling question of why regulators have been so unsuccessful in attempting to control online piracy may have an answer in network theory. This serves as the first in-depth case study that tries to demonstrate the importance of network theory for Internet regulation.

1. PARETO AND THE SUPER-STAR EFFECT

Intellectual property law in general, and copyright law in particular, have been drafted, promoted and perpetuated with the idea of the creator as a struggling individual who requires protection in order to make a living.[9] It is no coincidence that authors have been at the forefront of copyright policy and reform since its inception; Jonathan Swift[10] and Mark Twain[11] are just two names that prove this trend. In modern times, whenever there is talk of copyright reform, musicians are brought out to make impassioned arguments

9. The case for this idyllic idea of copyright rhetoric is masterfully made here: Coombe RJ, *The Cultural Life of Intellectual Properties: Authorship, Appropriation, and the Law*, Durham: Duke University Press (1998); and Woodmansee M and Jaszi P, *The Construction of Authorship: Textual Appropriation in Law and Literature*, Durham; London: Duke University Press (1994).
10. Deazley R, *On the Origin of the Right to Copy: Charting the Movement of Copyright Law in Eighteenth-Century Britain (1695–1775)*, Oxford: Hart (2004), p.128.
11. Litman J, *Digital Copyright: Protecting Intellectual Property on the Internet*, Amherst, NY: Prometheus Books (2001), pp.4–15.

about their dwindling coffers in order to try to garner public support for more protection.[12]

Together with the narrative of the lone author, international copyright law has been mostly influenced by the interests of what can be known as the large copyright industries, namely publishing, music recording, film-making, and recently the software and games industry.[13] To a lesser extent, more individual creative pursuits have had less representation, but still command some influence – these are artistic fields such as photography, art, sculpture, drama, etc. This has created an interesting chemistry in which the larger, more visible collective industries manage to maintain the status quo by relying on the "lone author" narrative, while the truly individual creative industries are under-represented.

Because the large copyright industries are profitable economic endeavours, it has always been easy to sell copyright reform by putting forward the argument that changing the law will have positive economic effects downstream. However, something that is less explored is that the economics of the copyright industry have until recently relied on income distributions that are better understood under the terms of power laws. Here is where Pareto comes in.

To refresh some of the concepts explained in Chapter 2, Pareto distributions,[14] named after economist Vilfredo Pareto, are used to describe large inequalities in data, where most of the distribution is concentrated in a relatively small portion of overall instances. This is popularly known as the 80/20 rule, following the perception that 80 percent of the work is performed by 20 percent of the employees; or that 80 percent of the wealth is held by 20 percent of the population.[15] In the content industries, the Pareto distribution

12. For an unintentionally comical piece that demonstrates this trend, see: Hucknall M, "Fundamental socialism", *The Guardian* (23, November 2006), http://www.guardian.co.uk/commentisfree/2006/nov/23/comment.music.
13. Boyle J, "The Second Enclosure Movement and the Construction of the Public Domain", 66:1 *Law and Contemporary Problems* 42 (2003).
14. Reed WJ, "The Pareto, Zipf and Other Power Laws", 74:1 *Economics Letters* 15 (2001).
15. See Barabási A-L, *Linked: The New Science of Networks,* Cambridge, MA: Perseus Pub. (2002), p.66.

would translate into a situation where 80 percent of the profits come from only 20 percent of creators.

It is easy to see why this is relevant to the content industries. If Pareto's Law is correct, then one would expect to find similar income inequalities in the creative sectors protected by copyright law. Most of the sales would go to a small number of individuals or firms, the "vital few and trivial many" as the Pareto principle states. This is perhaps the first hurdle of the science of networks with regards to Internet regulation. If something as universal as Pareto distributions do not occur in copyright markets, then the potential usefulness of network theories would be severely diminished, as so many of the debates regarding regulation in recent years has been centred precisely on this topic. Thankfully, most evidence seems to point towards a strong presence of Pareto's Law in the entertainment sectors. Most research into the economics of the content industries clearly displays Pareto distributions of wealth, exemplified by the often-commented phenomenon that most copyright earnings go to a comparatively small number of people.[16]

Let us look at some historic examples to get a picture of the evidence to support this statement. In a study of musicians' earnings in 1981, Rosen established an analytical framework that described the emergence of the "superstar" in order to explain anecdotal evidence pointing towards a disproportionate skew in earnings from a few creators at the top of the best-selling lists, followed by sharp drops in sales outside of a small number of artists.[17] Unfortunately, Rosen seems to have emphasised quality of performance in his analysis, which does not really explain the popularity of superstars. Quality is very subjective, and while one may argue that virtuoso performers do relatively well in some

16. See: Towse R, *Creativity, Incentive and Reward: An Economic Analysis of Copyright and Culture in the Information Age*, Cheltenham UK; Northampton, MA: Edward Elgar (2001), pp.80–86; and Towse R, "Copyright Policy, Cultural Policy and Support for Artists", in Gordon W and Watt R (eds), *The Economics of Copyright: Developments in Research and Analysis*, Cheltenham UK; Northampton, MA: Edward Elgar (2003), pp.66–81.
17. Rosen S, "The Economics of Superstars", 71 *American Economic Review* 845 (1981).

fields, popularity is more fickle than that. One is tempted to name several examples of popular works that have dubious quality credentials.[18]

Some of the most striking evidence with regards to Pareto inequalities comes from the music industry. Connolly and Krueger[19] conducted a survey of ticket sales in the United States between 1981 and 2003, and found that the top 1 percent sellers accounted for a disproportionate amount of the overall market (Figure 5.1). Not only did the top 1 percent creators outperform their competitors, but there was a marked increase over time of the superstar effect, in "1982, the top 1% of artists took in 26% of concert revenue; in 2003 that figure was 56%". This seems to respond not only to Pareto distributions, but at least in ticket sales we see also the "rich get richer" effect taking place, which is also something to be expected in complex networks.

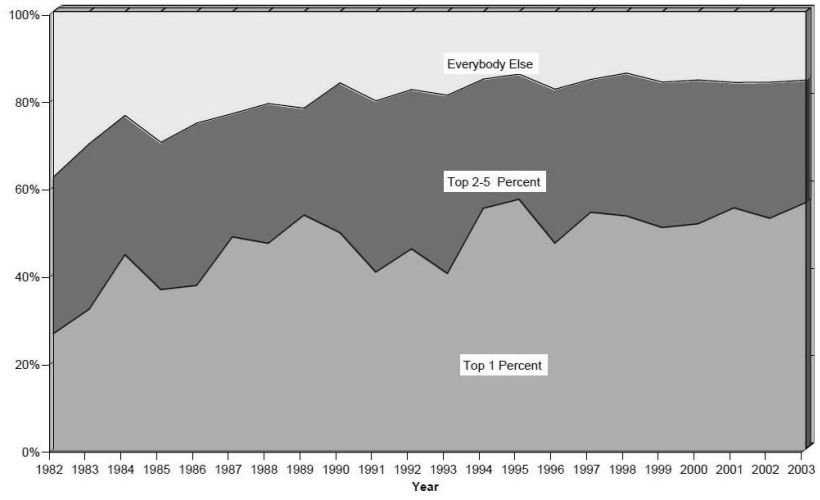


Figure 5.1 Share of total ticket revenue accruing to top performers, 1982–2003

18. On a very personal note, one could mention every film made by Michael Bay as evidence that quality and box office receipts are not correlated.
19. Connolly M and Krueger A, "Rockonomics: The Economics of Popular Music", in Ginsburgh VA and Throsby D, *Handbook on the Economics of Art and Culture*, Amsterdam: Elsevier (2006), pp.667–719.

This is a phenomenon that is encountered in other copyright works, such as in films. De Vany and Walls[20] undertook a survey of cinema ticket sales during a period of 13 years in North America, and found a strong Pareto distribution, where 78 percent of all movies lose money, and only 22 percent are profitable. Not only does this seem to accommodate almost exactly the 80/20 rule, but looking at profitable movies films produced a similar skewed result; for example, just 35 percent of profitable movies earn 80 percent of total profit.[21]

Similarly, the publishing industry seems to exhibit comparable skewed results. While conducting a search on price sensitivity in the online book market, Chevalier and Goolsbee found that "a tiny fraction of books in print account for most book sales".[22] Moreover, this phenomenon is replicated in other intellectual property industries, particularly in research and innovation.[23]

The first corollary of the existence of Pareto distributions with regards to earnings, profits and royalties may very well be that most creators cannot expect to make a living from copyright, and only a minority of works will be successful. For example, in the UK the most effective collecting society is the Mechanical-Copyright Protection Society (MCPS), which has more than 18,000 members; in 2004 it distributed £219 million GBP amongst them. Even if those profits were distributed equally, the average would be approximately £11,000 GBP.[24] This displays staggering levels of inequality of distribution.

---

20. De Vany A and Walls W, "Motion Picture Profit, the Stable Paretian Hypothesis, and The Curse Of The Superstar", 28 *Journal of Economic Dynamics and Control*, 1035 (2004).
21. Ibid, p.1040.
22. Chevalier J and Goolsbee A, "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com", 1:2 *Quantitative Marketing and Economics* 203 (2003), p.208.
23. Scherer FM, "The Size Distribution of Profits from Innovation", 86:49/50 *Annales d'Économie et de Statistique* 495 (1998); and Scherer FM and Harhoff D, "Technology policy for a world of skew-distributed outcomes", 29:4-5 *Research Policy* 559 (2000).
24. See: Mechanical-Copyright Protection Society, *Directors' Report and Accounts* (2004), http://www.mcps-prs-alliance.co.uk/aboutus/.

The evidence for the existence of Pareto's Law in the copyright industries is overwhelming, but what does it tell us about copyright policy? Is this another example of network theories telling us things we already knew?

There are several reasons why Pareto distributions in this area are of the utmost importance. The first and obvious conclusion is that copyright policy must have been informed by the existence of such inequalities. The large copyright industries are profit-making activities, so it is to be expected that they are organised to respond to the Pareto principle. If that is the case, then legislators must also have responded to the state of affairs and must favour the large earners as worthy of protection. While intuitive, it is more difficult to determine with certainty if this has been the case. While some scholars have attempted to rationalise copyright law in economic terms,[25] the exploration of the impact of Pareto's Law in the content industries has not been the subject of much scrutiny.

There is strong evidence, albeit indirect, that copyright law favours the superstars. In markets with strong Pareto distributions, one would expect to find that copyright law is drafted to protect top earning industries. One only needs to look at the copyright history of the last couple of decades to notice that there has been a strong push towards maximalism and stronger protection enshrined in the following pro-copyright owner provisions present in recent copyright policy:

a) Longer terms of copyright.
b) Legal protection of technological protection measures.
c) Criminalisation of some copyright infringement.
d) Erosion of fair dealing and fair use provisions.

25. A seminal work attempting to do just that is: Landes WM and Posner RA, "An Economic Analysis of Copyright Law", 18:2 *The Journal of Legal Studies* 325 (1989).

e) Creation of new exclusive rights or expansion of existing ones (such as making a work available to the public).[26]

Some copyright legislation specifically mentions that the goal of copyright protection is to incentivise creativity.[27] Given the prevalence of Pareto's Law in the copyright industries, this goal takes a secondary role, and it seems clear that the objective of copyright protection is to maximise profits, which means maximising protection for the superstars.

The obvious question to ask here is whether or not the status quo of protecting the superstar sellers affects other people involved in the content industries. The answer to this question lies in another economic idea based on Pareto, that of Pareto efficiencies. Pareto efficiency happens when the reallocation of resources makes someone better-off at the expense of making someone worse-off,[28] in other words, this happens in goods that are rivalrous in nature. For example, sharing a limited amount of funds would feature Pareto efficiency because giving more to one person would leave less money to be shared. Copyright works can be both rivalrous and non-rivalrous. For example, in the MCPS example cited above, there is a limited amount of money collected to distribute amongst copyright holders, which would create a Pareto efficient situation. On the other hand, copyright works are also non-rivalrous because it is possible to make a copy without negatively affecting others. However, when talking specifically about copyright protection drafted to protect superstars, one should argue that this does not affect in principle those who do not sell that well, as in theory the market can accommodate more sales.

Even if there is an inherent unequal distribution of profits in the copyright industries, there is little that copyright law can do to alleviate this situation. It is easy to complain

---

26. For more about these, see: Boyle J, *The Public Domain: Enclosing the Commons of the Mind*, New Haven, CT; London: Yale University Press (2008).
27. Particularly, the US Constitution, Art. I, § 8, cl. 8.
28. Greenwald B and Stiglitz JE, "Externalities in Economies with Imperfect Information and Incomplete Markets", 101 *Quarterly Journal of Economics* 229 (1986).

about the unfairness of it all, but there is really nothing to be gained bemoaning such inequalities; the law is simply responding to the universal presence of Pareto distributions in the creative markets. The superstars get more protection because they are the ones who sell more. The only possible way to redress this would be to create a distributive copyright system where profits are shared amongst a wider number of people, but this seems both impractical and unfair.[29]

There is, however, something to be learned from the prevalence of Pareto's Law, and it is that it serves as counter-evidence against the myth of the lone author described above, as copyright policy is based on a system that benefits a small minority. This should prompt future policymakers to look twice at setting policies that may have larger effects on the public, as a cost–benefit analysis of the current situation should attempt to benefit users and consumers, and not only a minority of stakeholders. Nonetheless, if Pareto distributions are almost inevitable when it comes to measuring copyright earnings, it is understandable that for many years policy has been skewed towards benefiting those who make profits from content.

Nonetheless, the universal prevalence of Pareto's Law in the creative sector is a result of the analogue world. When we look at what has been happening with the advent of digital markets and the Internet, a different picture emerges.

2. THE LONG TAIL

2.1 The rise of the long tail

Something interesting has been happening in recent years with regards to the allocation of profit in the copyright industries. As explained in the previous section, under classic Pareto distribution, high-earners take the larger slice of the profits, and sales drop off sharply. Nevertheless, a more detailed analysis of the copyright markets taking into

29. For an excellent argument against the use of distributive justice in copyright markets, see: Benoliel D, "Copyright Distributive Injustice", 10 *Yale Journal of Law & Technology* 45 (2007).

consideration electronic commerce and new media tend to produce a different story; large amounts of sales accumulate at the head of the graph, and while there is a drop-off point, the earnings accrued by smaller participants in the market tends tail off into the distance (Figure 5.2). The resulting graph shows a slightly different world to that of Pareto, that of the increasing returns, or what is also known as a "long-tailed distribution". This has turned into what is known as the theory of the long tail. In the word of Chris Anderson, its creator:

> The theory of the Long Tail is that our culture and economy is increasingly shifting away from a focus on a relatively small number of "hits" (mainstream products and markets) at the head of the demand curve and toward a huge number of niches in the tail. As the costs of production and distribution fall, especially online, there is now less need to lump products and consumers into one-size-fits-all containers.[30]

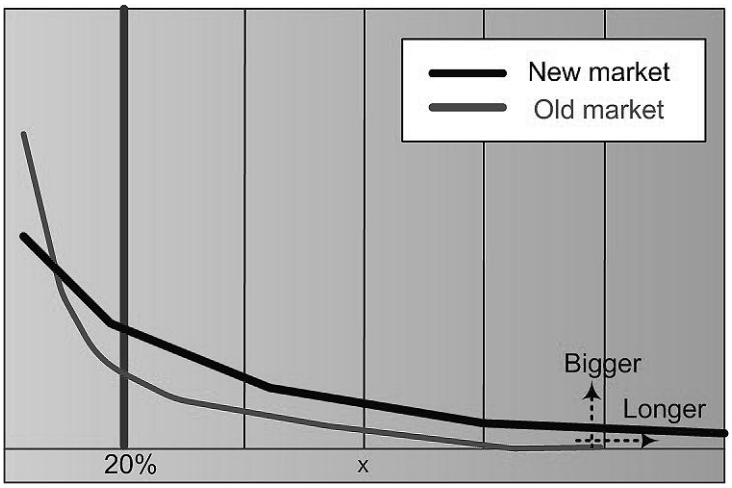


Figure 5.2 Long tail versus Pareto[31]

While the long tail does indeed look like the traditional Pareto distribution, there is a surprising addendum when one looks at how sales charts behave when one adds into the equation Internet data. In traditional brick-and-mortar creative industries, the retail

30. Anderson C, *The Long Tail FAQ*, (2005) http://www.thelongtail.com/about.html.
31. Source (released under a CC licence): http://commons.wikimedia.org/wiki/File:Longtail.jpg.

sector is specifically designed to respond to Pareto inequalities. Hits are given prevalence in shelf space all over music stores, bookshops or DVD rental locales.[32] However, something strange is happening to these inequalities online. Electronic retailers still experience the occurrence of a few massive hits and a long tail of less fortunate sellers, but when you factor out the need for limited shelf space, the tail keeps going, and does not seem to disappear.[33]

Anderson offers several examples that help to explain this remarkable find. Retail giant Wal-Mart shelves an equivalent 55,000 tracks in an average store, while digital music service Rhapsody has 1.5 million tracks. One would normally expect to see sales figures to respond to Pareto distributions. This happens still in "brick-and-mortar" retailers, but the remarkable find is that Rhapsody's entire inventory has sold at least one copy.[34] In e-commerce giant Amazon, one third of total sales come from books that are outside of the top 100,000 list, and 57 percent of all book sales come from titles that are not stored in high-street book retailers.[35] The long tail recognises that traditional media responds to power laws as profits go to a small cluster of entities. However, the Internet has provided a varied number of opportunities for those who did not have a chance to profit previously.

Further research into long tail economics appears to corroborate Anderson's findings. For example, an empirical study on sale distribution between electronic and catalogue sales found that consumer maturity and ease of searchability of content translated into a more equal distribution of sales between both retail outlets.[36] Similarly, digital music retailer eMusic has also released some of its sales data, which according to them supports the long tail theory. In a music catalogue of five million songs, eMusic has

32. Anderson C, *The Long Tail: The Revolution Changing Small Markets into Big Business*, New York: Hyperion (2006), pp.38–40.
33. Ibid, pp.19–23.
34. Ibid.
35. Ibid, p.23.
36. Brynjolfsson E, Hu YJ and Simester D, *Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales*, SSRN Research Paper Series (2007), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=953587.

found that 75 percent of their catalogue has sold at least one copy,[37] a finding consistent with Anderson's own analysis of other electronic retailers.

Another study into Netflix, a popular online movie rental site, found that there is a strong long tail in this market as well. While some of the findings corroborate the existence of Pareto distributions of sales and the reliance of superstars, 15 percent of all Netflix rentals came from movies outside the top 3,000 titles, which are not stocked by traditional retailers.[38] Interestingly, the long tail is not only circumscribed to content markets; other electronic commerce industries seem to be displaying similar effects. For example, data shows that the popularity of online booking of air travel and the proliferation of small airlines produces what some are calling the "long tail of travel", where smaller players see an increase in their share of the overall market.[39]

We will see some doubts about the existence of the long tail next, but it seems clear that at least in digital markets, there is something essential taking place. A picture is starting to emerge – Pareto still reigns supreme, and superstars still have a big chunk of the market, but the long tail has opened opportunities for small players to take a larger share of the profits than they would under the Pareto distribution model.

As more corroborating data starts coming in, one must question how the long tail works. As with most markets, there are clearly two sides to the long tail, supply and demand.[40] The reason for the existence of the long tail in the supply side has already been discussed, and has more to do with straightforward economics than with complex and network theories. Once electronic retailers have no need to rely merely on superstar sellers to turn a profit, any sale counts and niche markets can emerge. Brynjolfsson, Hu and Smith explain this thus:

---

37. Nevins CH and Keeble A, *Emusic Sales Data Supports "Long Tail" Concept*, Press Release (15 January 2009), http://www.emusic.com/about/pr/PR2009115.html.
38. Tan TF and Netessine S*, Is Tom Cruise Threatened? Using Net IX Prize Data to Examine the Long Tail of Electronic Commerce*, Wharton Working Paper (2009), http://opim.wharton.upenn.edu/~netessin/TanNetessine.pdf.
39. Barnhardt S, "The Long Tail of Travel", *Travalution* (19 April, 2007), http://www.travolution.co.uk/articles/2007/04/19/834/the-long-tail-of-travel.html.
40. Brynjolfsson E, Hu YJ and Smith MD, "From Niches to Riches: The Anatomy of the Long Tail", 47:4 *Emerald Management Reviews* 67 (2006).

> On the benefit-side, brick-and-mortar retailers sell to consumers in their local geographic region. Consumers with mainstream tastes will be served before consumers with one-in-a-million tastes. Internet retailers, on the other hand, can aggregate demand on a national or even global scale. With the potential Internet market approaching a billion consumers, even if you have one-in-a million tastes, there are still over a thousand like-minded consumers who share your niche tastes.[41]

But supply alone does not serve to explain the long tail phenomenon. It does not matter how many more works are available online, people must be willing to purchase or rent works that are not usually available through traditional retail channels. This is where some network theory explanations may be useful. Specifically, as it has been explored in previous chapters, the Internet's architecture is a positively conducive to the distribution of information. Small world and scale-free networks could very well explain the emergence of long tail markets. For example, the Internet favours small world networks by allowing people with common interests to communicate and organise in clusters.[42] These networks rely on connectors who have disproportionate influence in the overall behaviour of the network. Therefore, one would expect that word-of-mouth and Internet influence (say, through social media or blogs) could have an effect on buying patterns.

There are several studies that support this hypothesis. Oestreicher-Singer and Sundararajan[43] conducted a survey in Amazon.com of recommendation networks. By looking at data from 200 distinct categories, they established that categories whose products are influenced more by recommendations have significantly higher demand distribution, which supports the existence of a network-driven long tail effect. Giles provides further evidence in a study that proposes that increased information in cultural works translates into considerable deviation from Pareto models, and accounts for increasing returns. He comments that:

---

41. Ibid.
42. Vázquez A, "Growing Network with Local Rules: Preferential Attachment, Clustering Hierarchy, and Degree Correlations", 67:5 *Physical Review E* 056104 (2003).
43. Oestreicher-Singer G and Sundararajan A*, Recommendation Networks and the Long Tail of Electronic Commerce*, Wharton Working Papers (2009), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1324064.

> There appear to be some similarities between the way in which particular music recordings gain popularity, and the ways in which this occurs for movies and theatrical performances. In each case, for example, word of mouth recommendations can play an important role. The more people who have listened to, and purchased, a musical recording, the more information there is available to other potential agents.[44]

Perhaps more relevant to the theory of the role of small worlds to the long tail is a study on the influence of blogs in music sales. Dewan and Ramaprasad[45] looked at influential music blogs, which in a scale-free network such as the Internet could be classed as connectors, and tried to see if mention in these sites could be correlated with an increase in sales. Their research produced some very interesting results that confirm the existence of small world influences to long tail sales. Firstly, they found that the music blogs explored were not mainstream, which immediately would seem to deviate from Pareto distribution markets; these blogs would as a result tend to attract niche audiences. Secondly, they found that blog readership and membership tends to translate in stronger tail sales for the musicians featured by that community. Although the study does not talk directly about small worlds, it would seem that the reason for such clustering can be explained in light of network theory. Small world networks are more easily influenced because of the short pathways between actors, hence the highly skewed influence towards the tail.

2.2 Long tail or tall tales?

While there is growing evidence of the presence of the long tail effect in digital markets, it must be said that not everyone agrees with either the existence of the long tail, or with how critical a shift it is. While some researchers concede that there are growing sales in the tail, they point out that the content industries still rely heavily on superstars.

---

44. Giles DE, *Increasing Returns to Information in the US Popular Music Industry*, Econometrics Working Paper EWP0510, (2007) http://web.uvic.ca/econ/ewp0510.pdf.
45. Dewan S and Ramaprasad J, *Impact of Blogging on Music Sales: The Long Tail Effect*, Paul Merage School of Business Working Paper (2007), http://www.citi.uconn.edu/cist07/1b.pdf.

Some of the evidence counter to the long tail has been observed in the DVD electronic markets. In a study of video sales (DVD and VHS) from 2000 to 2005, Elberse and Oberholzer-Gee found that there was indeed a growth in "tail" markets, but interestingly they also found that the market has been suffering considerable polarization: fewer titles account for an increasingly larger slice of the market, while smaller sellers trail off into the distance.[46] This is a remarkable find for two reasons: it seems to corroborate the long tail effect, but it also seems to hint at the presence of another network theory effect, that of "the winner takes all". Presented with more choice, consumers seem intent not only on buying superstar products, but the share of the market of the top sellers seems to be increasing considerably. While electronic commerce has opened new revenue doors, it also is polarizing the market. It is almost as if we are seeing a runaway Pareto principle.

Elberse[47] has found similar trends looking at other datasets. She inspected the figures for electronic music retailer Rhapsody, which featured prominently in Anderson's book, and are often cited as one of the best examples of the existence of the long tail. While she found that there is indeed a tail, she found a remarkable concentration at the head, where 10 percent of titles accounted for 78 percent of all clicks, and the top 1 percent of titles took a staggering 32 percent of all plays. Similarly, she looked at figures for video rentals from digital service Quickflix, and she also found that 10 percent of DVDs accounted for 48 percent of all rentals. Elberse comments that the polarisation has actually been detrimental to smaller creators:

> When I differentiate between artists on smaller, independent labels and those on major labels, I find that the former gain some market share at the tail end of the curve as a result of the shift to digital markets. However, that advantage quickly disappears as we move up the curve: A more significant trend is that independent artists have actually lost share among the more popular titles to superstar artists on the major labels. [...] The data shows how difficult it is to profit from the tail.

---

46. Elbersei A and Oberholzer-Gee F, *Superstars and Underdogs: An Examination of the Long-Tail Phenomenon in Video Sales*, Harvard Business School Working Paper (2006), http://www.people.hbs.edu/aelberse/papers/hbs_07-015.pdf.
47. Elberse A, "Should You Invest in the Long Tail?" 86:7/8 *Harvard Business Review* 88 (2008).

Another source of criticism for the long tail has come from the British Performing Right Society (PRS). Will Page and Andrew Bud looked at an unnamed dataset for music sales and presented their findings at an industry event. They found that:

> For example, we found that only 20% of tracks in our sample were 'active', that is to say they sold at least one copy, and hence, 80% of the tracks sold nothing at all. Moreover, approximately 80% of sales revenue came from around 3% of the active tracks. Factor in the dormant tail and you're looking at a 80/0.38% rule for all the inventory on the digital shelf. Finally, only 40 tracks sold more than 100,000 copies, accounting for 8% of the business.[48]

Unfortunately, the authors have not published their results, and there is no indication as to what dataset has been used. There has been speculation that the data may come from mobile downloads, which may account for the very wide divergence from some of the other electronic commerce services.[49] Mobile content is a unique market because it consists mostly of ringtone downloads. Users would probably want to download a very distinctive tune to have as their ringtone, which could explain why there is such a sharp skew in this dataset.

Page has also looked at data from the popular streaming service Spotify. Here the data is even more contrary to the long tail theory. Page found that by 2009 there were 4.5 million songs available in the service, but of those only 3 million had been played by the almost 2.6 million users.[50] This is a long tail of tracks with no plays. Moreover, listening figures clearly favoured popular artists, which seem to be consistent with the existence of a "winner takes all" scenario. Rich acts get richer, smaller acts languish at the tail.

Why is this concentration happening? As it has been explained, one of the reasons why there is a long tail effect in online environments is the ease of connecting

---

48. Telco 2.0, *The "Long Tail" Interrogated*, (12 November, 2008), http://www.telco2.net/blog/2008/11/exclusive_interview_will_page.html.
49. Anderson C, "More Long Tail Debate: Mobile Music No, Search Yes", *The Long Tail Blog* (November 8 2008), http://www.longtail.com/the_long_tail/2008/11/more-long-tail.html.
50. "Spotify: The UK Stats", *Music Ally* (15 October, 2009), http://bit.ly/ao9U66.

consumers and the existence of recommendation systems. Dellarocas and Narayan[51] conducted a survey of online recommendations for films in Yahoo Movies for 2002, and correlated that information with box office receipts. They found that online consumers were more likely to review popular products, and so, contrary to the long tail effect, online reviews may exhibit "tall heads" instead of "long tails". The "winner takes all" scenario that we have been witnessing could very well be explained by a situation where users are simply more likely to review popular titles.

Having said this, even the conflicting evidence still points towards a change in consumer patterns. Whatever importance one may give long tail economics, it is clear that the old Pareto distribution model is undergoing major shifts.

3. PEER-TO-PEER

3.1 Brief introduction to the technology

Having explored the legal side of the equation, it is time to turn to illegal file-sharing and copyright infringement. While this is a topic often covered in the literature, the actual technologies involved in wide-scale copyright infringement online are often misunderstood and even misrepresented.

Peer-to-peer (P2P) is a term that is most usually used to describe illegal file-sharing. However, at its most basic level, the term simply is used to refer to decentralised technical and/or organisational architectures. The term is used to describe decentralised banking,[52] lending,[53] social networks,[54] and many other non-technical arrangements.

---

51. Dellarocas C and Narayan R, *Tall Heads vs. Long Tails: Do Consumer Reviews Increase the Informational Inequality Between Hit and Niche Products?* Robert H. Smith School of Business Research Paper No. 06-056 (2007), http://papers.ssrn.com/sol3/ papers.cfm?abstract_id=1105956.
52. http://www.wiseclerk.com/group-news/.
53. Svolkia J, "Forget Citibank – Borrow from Bob" *Harvard Business Review* (2009), http://hbr.org/web/2009/hbr-list/forget-citibank-borrow-from-bob.
54. http://www.slsknet.org/.

In strict information technology terms, P2P is usually used to describe a generic way to distribute transport loads in a telecommunications network. It mostly "refers to the concept that in a network of equals (peers) using appropriate information and communication systems, two or more individuals are able to spontaneously collaborate without necessarily needing central coordination".[55] In other words, participants in the network share the resources necessary to make the system work, be it storage, bandwidth, energy, data, etc.

While the technology itself is self-evidently neutral, the term has become almost synonymous with illegal file-sharing because of the prevalent use of P2P networks to share copyright infringing copies. In the last decade we have seen three main types of P2P technologies used for sharing files: semi-central server systems, decentralised client-based networks, and BitTorrent.

The mediated[56] server-based P2P network is a model that relies on some form of central server to operate; the most famous example of which is Napster. In the Napster network, users downloaded the Napster client, and connected to a central server that held information on which files people were sharing. A user would then connect to other user's computer and download the file. This type of model is technically a P2P network because it connects two users, even though it relies on the central server in order to keep track of users and files.[57] Some literature refers to this type of architecture as a mediated system.

The decentralised client-based P2P networks operate entirely without mediation from a central server; examples of these are networks such as Fasttrack, eDonkey2000 and Gnutella; and software clients such as Aimster, Grokster, Limewire, eMule, eDonkey and Kazaa. In this model, the user would download a client which would connect to one

55. Schoder D and Fischbach K, "Core Concepts in Peer-to-Peer (P2P) Networking", in Subramanian R and Goodman B (eds), *P2P Computing: The Evolution of a Disruptive Technology*, Hershey PA: Idea Group Inc. (2005), p.21.
56. Backx P et al, "A Comparison of Peer-To-Peer Architectures", *EURESCOM Summit* (2002), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.151.9453&rep=rep1&type=pdf.
57. Saroiu S, Gummadi KP and Gribble SD, "Measuring and Analyzing the Characteristics of Napster And Gnutella Hosts", 9:2 *Multimedia Systems* 170 (2003).

or several P2P networks. Once connected, the user would be able to search files shared by other clients connected to the network, and then would download the content from one or many computers hosting the same file. The client developers do not run the actual networks; they just make a client that can connect to the network.[58] The main difference between the centralised and the decentralised models is precisely the lack of a central server that stores information about the files.

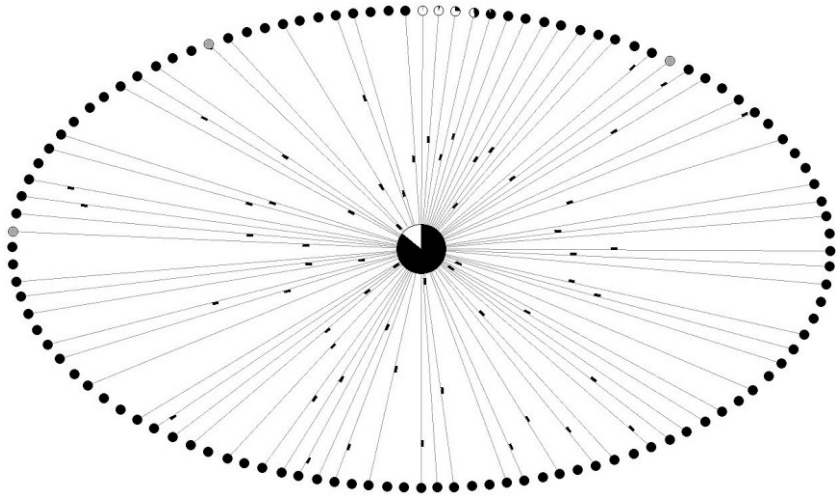


Figure 5.3 A typical BitTorrent swarm[59]

The BitTorrent network is a qualitative jump from the other two models described because it does not require a client, although users may still need a program that can process .torrent files. BitTorrent is a communications protocol that distributes file-sharing amongst users with an entire copy of the file (seeds), and/or amongst users with incomplete versions of the whole (peers). The information of who is sharing the files at any given time is distributed through a tracker file which allocates resources accordingly

58. Sen S and Wang J, "Analyzing Peer-To-Peer Traffic across Large Networks", *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement* (2002).
59. The circle in the middle is the local file. The full circles are seeds, and the incomplete circles are peers in the swarm. All of the peers are exchanging small parts of the file amongst each other. The screenshot was taken from Vuze from a legitimate copy of Open Office 2.2.

at the local level; all of the seeds and peers sharing a file form a network (swarm).[60] If a user wants to find a file, all he needs to do is to go to a search engine and type the name of, say, a movie. If the film is being shared, there may be a torrent file that contains information of those who are sharing the file at the moment. All that is needed is for one person to initially have the file and upload the tracker to a torrent tracking site; for that reason this person will be seeding the copy. Once other users find it, they will start downloading the file, but at the same time they will be sharing it with others in the swarm. Eventually, peers that complete the download and keep their BitTorrent client open will become seeds; the more seeds, the "healthier" a torrent is (Figure 5.3).

Of all the three models, the one that is prevalent at the time of writing is unsurprisingly the BitTorrent protocol. This is a very efficient manner of sharing large files, as it distributes the load amongst participating users. Because it is wholly decentralised, the decisions about the amount of participation and the time of connection are all left to the user. As long as the BitTorrent application is running and instructed to share files, it will do so.[61] It is relevant to stress the technical importance of the tracker: it helps peers connect to each other, tell each other which port they are listening into and the contact information on which seeds and peers are sharing the same file. This is perhaps the only centralised feature of BitTorrent, as it relies on the existence of tracker servers.

It must be noted that the BitTorrent protocol has been adopted by mainstream content owners in order to share files as well; for example, it is used by Microsoft in its consumer synchronisation service.[62] It is also used by game developers to distribute upgrades, such as World of Warcraft, and it is also a very popular manner of distributing open source software, such as Linux distributions and Open Office.

---

60. BitTorrent.org, *Protocol Specifications*, (2006), http://www.bittorrent.org/protocol.html.
61. Cohen B, "Incentives build robustness in BitTorrent", *Proceedings of Workshop on Economics of Peer-to-Peer systems* (2003), http://bit.ly/9SeJIW.
62. Windows Live Mesh.

3.2 P2P and network theory

It is easy to see why P2P networks are of interest to complexity theory. Here we have real-life examples of large-scale networks designed specifically to exchange information. Because most of the networks are non-proprietary, researchers often have access to almost entirely unprecedented vast datasets. P2P networks also seem to present us with corroboration of many of the principles of complex systems that have been described in earlier chapters, namely whether or not they present scale-free distributions, whether or not they are small world networks, their resilience and, perhaps more importantly for Internet regulation, whether or not they self-organise.

When researchers have looked at P2P networks using the analytical tools of network theory, they have found that they do indeed display power law characteristics, which may explain many of the features of scale-free networks, particularly stability and robustness. Ripenau, Foster and Iamnitchi[63] conducted a survey between 2000 and 2001 of the Gnutella P2P network to assess its structure. They found some interesting power law characteristics in the network. First, they discovered that P2P networks were scalable; in other words, while the network kept growing consistently, the overall features remained the same[64] – if you recall the discussion of power law distributions, this is a common tell-tale sign of the existence of power laws. Second, when looking at the distribution of links within the network, they found typical scale-free distribution of links – namely, most nodes had fewer links, while few hubs had a disproportionate amount of edges. Third, when looking at the connectivity to the network, that is, the amount of time a client stayed connected, they found a strong power law as well.[65] More

63. Ripenau M, Foster I and Iamnitchi A, "Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design", 6:1 *IEEE Internet Computing Journal* (2002).
64. Ibid, p.6.
65. Ibid, p.7.

research into the Gnutella network has been producing similar results,[66] which is strong evidence to assume that P2P networks are indeed power law networks.

If power laws are present in P2P networks, and there is no reason at the moment to assume otherwise, other characterises of complex systems should also be found. In another study, researchers charted path lengths in P2P networks in order to find out if they represented small worlds.[67] How would small worlds operate in P2P networks? It would of course depend on the type of network studied, but in a small world P2P network one would expect to find small interconnecting paths from any given node. This can be measured by looking at the average number of nodes and hubs that information has to travel to get to a random recipient. The aforementioned study discovered some power law behaviour, but researchers were surprised that links tended to cluster more than would otherwise be expected in a scale-free topology. The researchers then created their own P2P network, as they guessed that their results were being skewed by the efficiency of web site search engines. The resulting link distribution between nodes in the network corresponded to power laws. Further research into the topic tends to corroborate these findings, and serves as good indication that P2P are not only scale-free networks, but that they are also small world networks.[68]

P2P, and particularly BitTorrent, appear to be almost perfect examples of self-organisation in action. While the networks arise from architectural decisions at the start, the fact that they are almost completely decentralised means that there is no organising force, and consequently their growth and evolution is autonomous and organic. A key to the self-organising principles of complex adaptive systems is that seemingly chaotic conditions become ordered thanks to systemic and/or architectural conditions. Brahm

---

66. Jovanović M, Annexstein F and Berman K, "Modeling Peer-To-Peer Network Topologies Through 'Small-World' Models And Power Laws", *IX Telecommunications Forum* (2001).
67. Khambatti M, Ryu K and Dasgupta P, "Structuring Peer-to-Peer Networks using Interest-Based Communities", *International Workshop on Databases, Information Systems and Peer-to-Peer Computing*, Humboldt University, Berlin, Germany (September 2003).
68. See: Saroiu S, Gummadi KP and Gribble SD, "A Measurement Study of Peer-to-Peer File Sharing Systems", *Proceedings of Multimedia Computing and Networking 2002* (2002); and Adamic L and Huberman B, "Zipf's law and the Internet", 3 *Glottometrics* 143 (2002).

Cohen admits that he designed the BitTorrent protocol with two key features in mind, robustness and efficiency.[69] The BitTorrent client will attempt to form Pareto efficiency between peers, this is to say, it will try to maximise up to the point where both peers will benefit from the exchange. These two architectural conditions explain exactly why BitTorrent is so good at organising peers and seeds to serve large amounts of data efficiently.

If P2P networks display power laws, then it is evident that they would also be robust by design. This is because scale-free networks are resilient as any random attack on a node will not hit an essential one, and the network will remain operational.[70] Most evidence points that nodes and hubs in P2P networks follow a power law, so any attack on the system will not result in wider failure. Studies seem to confirm this finding. A study into the Gnutella P2P network found inherent vulnerabilities, but concluded that:

> There are two mechanisms that cause the formation of scale-free topologies. First, networks expand continuously by the addition of new vertices, and second, new vertices attach preferentially to vertices that are already well connected. In Gnutella, the first mechanism can be seen by the fact that new nodes are continuously entering and leaving the system, meaning the topology is undergoing constant change and growth. The second mechanism can be seen by the fact that there are only a few hosts that clients initially connect to [...]. Hence, the topology of the Gnutella network is scale-free because of its adherence to these two mechanisms.[71]

It is remarkable that most of the literature which studies P2P networks remarks on their resilience and stability.[72] P2P networks not only have a power law distribution of links, but they are also very fluid. Any given file shared using a protocol such as BitTorrent will have a steady number of seeds, but it will also have peers coming in and out of the swarm. Some seeds will act as hubs in the network by staying connected for

69. Ibid.
70. Newman MEJ, Barabási A-L and Watts DJ, *The Structure and Dynamics of Networks*, Princeton, NJ: Princeton University Press (2006), p.425.
71. Keyani P, Larson B and Senthil M, "Peer Pressure: Distributed Recovery from Attacks in Peer-to-Peer Systems", *Proceedings of the International Workshop on Peer-to-Peer Computing* 306 (2002), p. 307.
72. Qiu D and Sang W, "Global Stability Of Peer-To-Peer File Sharing Systems", 31:2 *Computer Communications* 212 (2008).

longer periods and sharing larger portions of bandwidth to the swarm, but most peers connected to the network will have both smaller connection times and smaller bandwidth to share. The removal of any given seed, even if it is a central one, will not affect the swarm. And this does not even touch on the most interesting feature of P2P BitTorrent networks. Each tracker creates its own network. Even if it was possible to remove one tracker, there are hundreds of others waiting to carry the load. This is what resilience is all about.

However, while extremely resilient, P2P networks could also have inherent vulnerabilities. The first potential issue is one of computer virus propagation. Because these are highly-efficient networks, P2P systems seem to be remarkably prone to computer virus epidemics. According to Adamic and Huberman:

> Finally, it has been shown that scale-free networks are more susceptible to viruses than networks with a more even degree distribution. Namely, a virus spreading in a random network needs to surpass a threshold of infectiousness in order not to die out. However, if the network has a Zipf degree distribution, the virus can persist in the network indefinitely, no matter what level of its infectiousness.[73]

Another issue with the legendary resilience of P2P networks is that they really cannot be completely decentralised. At some stage, any individual who wants to share files using a P2P network will have to connect to another computer and/or server in order to obtain information about where the file is being shared. Decentralised P2P client-based networks rely on peer connections in order to find hosts, which create inefficient search architectures.[74] A user operating a client-based network like Gnutella will broadcast a file search to its networks, which in turn will broadcast the search to their networks; this can slow down the system and make finding peers an inefficient exercise.[75]

---

73. Huberman and Adamic, supra note 68.
74. Ibid.
75. Ibid.

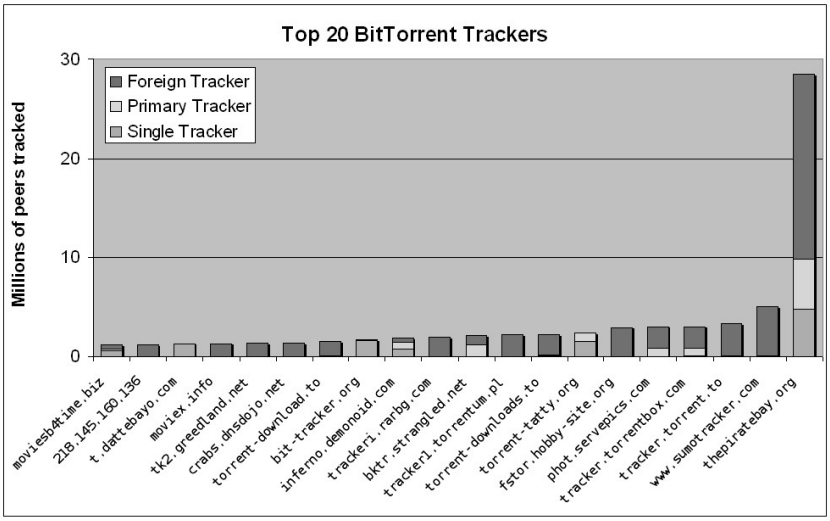


Figure 5.4 Long tail of tracker sites?[76]

BitTorrent is not immune from the problem of centrality either. While it is true that each file shared is its own network, BitTorrent is heavily reliant on tracker files. To illustrate this point, let us follow a typical infringing file-sharing download of one of the instalments of the popular teen vampire Twilight saga, *Eclipse*. The first stage is to find the tracker file. To do this one just needs to type "twilight eclipse torrent" into Google, which at the time of writing produced 29 million results. All of the links in the first four pages of the search directed to torrent tracker sites which are sites that either host the tracker or that link to places where the tracker is hosted. Let us remember that the tracker file is vital. The links in the search result would direct you to a site where you can download the .torrent file for that specific work. If you wanted to download the file, you would need a computer program that can handle torrent files. As it has been said before, these programs can be used for all sorts of legitimate uses. In this case we are using Vuze. Clicking on the link opens the program and then you have the option to download the file. This results in joining a swarm that is sharing the file. In the example

76. Vliegendhart R, *Top 20 BitTorrent Trackers*, (2009), http://www.tribler.org/trac/wiki/DistributedTracker.

we are using, the program is connected to over 2,000 seeds, and just over 100 peers. The tracker itself lists 3,100 seeds and 102 peers on average, and the tracker is hosted in different sites, each version carrying more than 2,000 seeds. It would be practically impossible to shut down all of these connections: even if one shuts down one seed, there are thousands of other users sharing the file.

Why is this vulnerable? The clue is in the tracker. Any user willing to download *Eclipse* will still have to connect to one single tracker file. However, an interesting and perhaps ironic feature of BitTorrent tracker files is the fact that most trackers are hosted in very few servers in a manner that resembles Pareto distributions. Research has been conducted into tracker sites, and it has become clear that few sites host most trackers (Figure 5.4).

It is obvious that by 2009 the PirateBay was by far the most popular tracker site. One can assume that if the PirateBay website were to disappear tomorrow others would take its place. Nonetheless, the chart above shows that BitTorrent is still highly centric, and centrality means potential vulnerability.

4. COPYRIGHT IMPLICATIONS OF NETWORK THEORY

It is hoped that the above sections have provided enough evidence that there are indeed practical applications of network theory to copyright subjects, both in the legal and illegal markets. Most of what has been explained so far is mostly useful for descriptive purposes. The creative industries operate under Pareto's Law, and digital content increasingly displays long tail distributions. In illegal file-sharing, P2P networks undoubtedly work as scale-free networks. Can network theory give us any prescriptive insights? Can network theory help us draft better copyright laws?

The first issue is a practical one. Historically, copyright law has been highly susceptive to lobbying by the content industries. However, there seems to be a growing

trend in intellectual property policy to draft future strategies based on evidence.[77] There are three relatively successful examples of evidence-based policymaking in Europe. The first was the considerable public consultation process and research going into the discussion of the European Directive on Computer Implemented Inventions, which resulted in the eventual demise of the proposal.[78] The second example has been the Gowers Review of Intellectual Property,[79] which has made a big point of putting evidence before the interests of powerful lobbying groups. The third example was the extensive consultation process that led to the drafting of the Digital Britain Report[80] in the UK, which would later inform the passing of the Digital Economy Act.[81]

While none of these examples has made use of the research highlighted in previous sections, the following section will attempt to pose examples of how network theory could inform legislators and policymakers in order to produce better-informed copyright policy. Just as in the rest of the chapter, both "legal" marketplaces and illegal file-sharing will be dealt with separately, although it is clear that there is room for cross-pollination between one and the other.

4.1 Towards a long tail copyright policy

The discussion about a possible long tail copyright policy must begin by making a clear distinction between the law and business models. It is perfectly possible to have in place copyright legislation that does not reflect existing business models, or that business models could adequately change without affecting copyright law and policy. If this is the case, then the emergence of the long tail would not necessitate changes in copyright law.

77. One of the most outspoken supporters of this approach is Professor James Boyle. See: Boyle J, "A natural experiment", *Ft.com* (November 2004), http://www.ft.com/cms/s/4cd4941e-3cab-11d9-bb7b-00000e2511c8.html.
78. *Proposal for a Directive of the European Parliament and the Council on the Patentability of Computer-implemented Inventions,* COM(2002) 92. For more about this, see: Guadamuz A, "The Software Patent Debate", 1(3) *Journal of Intellectual Property Law & Practice* 196 (2006).
79. Gowers A, *Gowers Review of Intellectual Property*, HM Treasury, (2006).
80. Department of Business, Innovation and Skills, *Digital Britain Report* (2009), http://interactive.bis.gov.uk/digitalbritain/report/.
81. Digital Economy Act 2010 (c. 24), http://www.opsi.gov.uk/acts/acts2010/ ukpga_20100024_en_1.

Is it possible for long tail models to exist under existing copyright regimes? Perhaps the mere existence of the long tail is an answer to this question. However, what if copyright policy is being drafted to maintain a decreasingly relevant business model? What if the newest and future legislation simply perpetuates defunct strategies? Would it not be vital to try to avoid passing legislation that is irrelevant the moment it is enacted?

Traditionally, copyright law has one main purpose, succinctly expressed in the US Constitution, which states that it copyright exists "to promote the Progress of Science and useful Arts".[82] As laudable as this goal is, were one to draft a justification for modern copyright law, it would probably read "to promote profits for copyright holders". Copyright law serves to sustain specific business models. Current business models are based – wittingly or not – on Pareto's Law, so current copyright law protects the status quo.

A look at some of the latest attempts to draft copyright legislation will serve to illustrate this point. As it has already been mentioned, the UK was recently involved in a crucial policymaking exercise in order to adapt copyright law to the challenges presented by the Internet. In one telling paragraph, the Digital Economy Report highlights why the evidence presented by network theory is more relevant today than ever before. The Report states:

> The popularity of X-Factor and Britain's Got Talent shows the enduring drawing power of content-creating talent that few people possess. The digital world allows more of that talent to find its way to more consumers and admirers than ever before. But it is not wholly democratic: some have the talent to create content; many others do not. As throughout history, there need to be workable mechanisms to ensure that content-creators are rewarded for their talent and endeavour. And the need for investor confidence is key. User generated videos can be hugely popular, but there remains a healthy appetite for big movies costing many millions to produce.[83]

---

82. United States Constitution, Art. I, Section 8, Clause 8. For more about the justifications to Intellectual Property in general, see: Hettinger EC, "Justifying Intellectual Property", 18:1 *Philosophy and Public Affairs* 31 (1989).
83. *Digital Britain Report*, supra note 80, p.109.

Unwittingly, the drafters of the Digital Britain Report have produced a paragraph that reeks of Pareto's Law. Hidden throughout the report is the assumption that only a few can create content, that only a few can profit from such content, and that these creators must be rewarded for their investment. Reading through the Report, anyone who knows about the Internet but, most importantly, who has seen the rise of the long tail cannot help but notice that here we are presented with policy solutions that are simply attempting to maintain Pareto distribution inequalities in place, regardless of the evidence. The above paragraph should have said that the popularity of shows like the *X-Factor* and *Britain's Got Talent* have a decreasing share in a growing market, and that it is now evident that more and more people can create content, regardless of talent. While this will be discussed in more detail in the next chapter, it should be stressed that such mentality cannot go unchallenged.

It is true that it is difficult to pinpoint specifically Pareto's influence in existing copyright law, but it is clear not only from the Digital Britain Report, but from almost any other legal document dealing with copyright and the Internet, that these assumptions are taken as a given.[84]

This brings us back to the principal question of the difference between business models and the law. It seems clear that copyright law is still today being drafted to accommodate Pareto distributions. Is that incompatible with new business models exemplified by the long tail? If the answer is no, then copyright policy can continue as it stands. But if there any specific area of existing copyright law and policy where the interests of the traditional copyright owner and the long tail business models diverge, then this should pose a significant conundrum for policymakers. There is indeed one area of copyright where there seems to be a conflict between the status quo and new business models brought about by digital marketplaces, and that is the chief role of the intermediary in online environments.

---

84. Take for example, the recitals in the *Directive 2001/29/EC of the European Parliament and the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society*, where the profits of copyright holders are given paramount importance.

One key feature of the Internet is the role played by intermediaries, be they Internet service providers, content aggregators and search engines. The amount of online information means that we rely more and more on these intermediaries for almost every aspect of our wired lives. ISPs allow us to connect, but in many instances they also provide hosting services. Search engines and content aggregators allow us to find content, some of which might be infringing.

The liability of these intermediaries for illegal actions taken by their users within the networks has been the subject of litigation, scholarly analysis and regulatory response since the early days of the Internet. Early on, content owners undertook legal action against ISPs and other intermediaries in order to attempt to obtain damages and/or injunctions for infringement taking place in their networks.[85] While some of these lawsuits were successful, the effect on early intermediaries was devastating, and it soon became clear that there needed to be some sort of rationalisation of the liability regime.[86] The rationale for this is that with a growing number of users and multiplying amount of content, it would be impossible for most intermediaries to police whatever took place in their networks unless they exercised strict editorial policies.

The solution was the creation of a limited indemnity for intermediary service providers online, exemplified by the EU Electronic Commerce Directive (ECD)[87] and the US Digital Millennium Copyright Act (DMCA).[88] The common denominator of both legislative solutions is to maintain liability for Internet intermediary services, but also to create a limited indemnity regime if ISPs have no previous knowledge of any illicit activity. This principle assumes that intermediaries have no editorial control over the large amount of information within their networks, and for that reason cannot have any

85. Amongst others, see: *Frank Music v. CompuServe Inc*., No. 93 Civ. 8153 (S.D.N.Y. 1993); *Playboy Enterprises, Inc. v. Frena*, 839 F. Supp. 1552 (M.D.Fla. 1993); *Sega Enterprises v. Sabella*, No. C93-04260 (N.D. Cal. 1996); and *AdobeSystems Inc. v. Tripod Inc*., No. 1:96CV157 (N.D. W.Va.).
86. For more on this, see: Edwards L and Waelde C, *Online Intermediaries and Liability for Copyright Infringement*, WIPO briefing paper WIPO/IIS/05/1, (2005), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1159640.
87. Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market.
88. Digital Millennium Copyright Act, 112 Stat. 2860 (1998).

knowledge of infringement being committed. As long as they act to remove the infringing content, this indemnity will remain in place. Art. 12 of the ECD works on the assumption that intermediaries act as "mere conduits", and the DMCA establishes safe harbours for intermediaries who provide a notice-and-take-down procedure for content owners.[89] While imperfect,[90] this system worked reasonably well for almost a decade; it gave intermediaries some respite from excessive litigation, and it gave content owners a workable system that still allowed them to take down infringing materials.

However, copyright infringement continued unabated, and as a result some copyright owners have been trying hard to bring back some form of intermediary liability into the statute books through lobbying and through case law. The opening salvo in the new intermediary wars was undoubtedly *Viacom v YouTube* in 2007.[91] In this case, media giant Viacom sued video-hosting site YouTube for $1 billion USD for direct infringement of the exclusive rights to public performance, public display and reproduction of owned content. In their complaint Viacom alleged:

> Defendants encourage individuals to upload videos to the YouTube site, where YouTube makes them available for immediate viewing by members of the public free of charge. Although YouTube touts itself as a service for sharing home videos, the well-known reality of YouTube's business is far different. YouTube has filled its library with entire episodes and movies and significant segments of popular copyrighted programming from Plaintiffs and other copyright owners, that neither YouTube nor the users who submit the works are licensed to use in this manner.[92]

This was not really a surprising development for those following the copyright wars. What seems surprising is that Viacom would go to the extent of suing a large service provider knowing that the law was not on their side. The judge agreed and granted

89. Bernstein A and Ramchandani R, "Don't Shoot the Messenger! A Discussion of ISP Liability", 1:2 *Canadian Journal of Law and Technology* 1 (2002).
90. For some criticism of the safe harbour provisions applying to search engines, see: Walker CW, "Application of the DMCA Safe Harbor Provisions to Search Engines", 9:2 *Virginia Journal of Law &Technology* 1 (2004).
91. *Viacom International Inc., et al. v. YouTube Inc., et al.*, Nos. 07-Civ-2103 (LLS), 07-Civ-3582 (LLS) (S.D.N.Y. 24 June, 2010).
92. The complaint can be found here: http://bit.ly/bPDzIY.

summary judgment in favour of YouTube by stating that the site is protected by the safe harbour provisions of the DMCA. Just before the complaint, YouTube had been purchased by Google, so this case became emblematic of the struggle between the content lobby and intermediaries.

The trend has been repeated in other jurisdictions. In Australia, *Roadshow Films v iiNet*[93] looked at similar questions. iiNet is an Australian internet provider, which was sued for secondary infringement by Australian film producer Roadshow Films, part of the Village Roadshow media conglomerate. The question at the heart of the proceedings was whether an ISP can be held liable for the copyright infringement committed by its customers. The judge in the case correctly identified that while there was ample evidence that there was infringement taking place in the defendant's network, but that they could not be held liable just by providing a connection to the Internet because iiNet could not be seen as "sanctioning, approving or countenancing copyright infringement".[94]

In Europe, a Belgian court came to a different decision in *Sabam v Tiscali*.[95] The case was brought by the Belgian Society of Authors, Composers and Publishers (Sabam) against ISP Tiscali (now called Scarlet). Sabam wanted Tiscali to install filtering software in its systems, which would allegedly curb illicit file-sharing in P2P networks. The first ruling in the District Court of Brussels agreed with the claimants based entirely on expert reports about the feasibility of deploying filtering systems. The case, however, has been appealed and is currently referred to the European Court of Justice.[96]

On the legislative front, content owners have been lobbying hard to reform or repeal the existing liability indemnity principles. One of the most publicised attempts has been the enactment of so-called three-strikes laws which shift the burden of enforcement from owners to ISPs. Under this regime, a content owner would issue an ISP with notification

93. *Roadshow Films Pty Ltd v iiNet Limited* [2010] FCA 24.
94. Ibid, para 14.
95. *Sabam V. S.A. Tiscali (Scarlet)*, District Court Of Brussels, No. 04/8975/A (29 June 2007).
96. Guadamuz A, "ISP Liability to Get ECJ Hearing", *TechnoLlama* (12 February, 2010), http://www.technollama.co.uk/isp-liability-to-get-ecj-hearing.

that a user is engaged in copyright infringement. The ISP would then issue a warning letter to the user and if he failed to comply a second letter would be sent, and further infringement would see the service being disconnected from the Internet altogether, hence the name. The first country to adopt such a law was South Korea, which in March 2009 passed reforms to its Copyright Act which gives authority to ISPs to send warning letters to infringing users asking them to stop transmission of illegal copies, and ultimately allows them to suspend or terminate the offending accounts.[97] The second country to enact similar legislation was France, which enacted the Loi favorisant la diffusion et la protection de la création sur Internet (HADOPI).[98] While controversial and hotly contested in the French Parliament, HADOPI has put in place a system of disconnection that has to be approved by civil courts, and thus it is not as burdensome to ISPs as some of the earlier proposals seemed to imply.[99]

Not to be outdone, the UK has included the possibility of disconnection in the aforementioned Digital Economy Act 2010, which might see users disconnected after repeated infringement notices have been sent.[100] At the time of writing the precise details of disconnection are under consultation, so they will not be discussed in detail at this moment. However, it can be remarked for the purpose of this work that during the debate leading to the enactment of the Digital Economy Act, the content industries were engaged in a monumental lobbying effort in order to see the notice and disconnection regime included in the final legislation.[101]

Finally, at the time of writing another possible large-scale shift is being discussed in the Anti-Counterfeiting Trade Agreement (ACTA). This is a multilateral trade

97. "South Korea's 'Three-Strikes' Law Takes Effect", *Zeropaid* (23 July, 2009), http://www.zeropaid.com/news/86703/south-koreas-three-strikes-law-takes-effect/.
98. HADOPI is not the name of the legislation, but the name of the authority which oversees the law, the "Haute Autorité pour la diffusion des oeuvres et la protection des droits sur Internet". However, the law has become synonymous with the institution.
99. Jondet N, "The French Copyright Authority (HADOPI), the Graduated Response And the Disconnection Of Illegal File-Sharers", *BILETA 2010*, University of Vienna (March 2010).
100. Ss 3-18.
101. See for example, Taylor G, "Never Mind The Billshock", *British Phonographic Industry Blog* (25 January, 2010), http://www.bpi.co.uk/blog/post/Never-Mind-The-Billshock.aspx.

agreement between the EU, the US, Mexico, Canada, Australia, South Korea, New Zealand and a few others, that is set to tackle trade mark and copyright infringement issues. While the negotiations are being kept secret, leaked versions of the text indicate that we could see an end to the liability indemnity regime. As it has been mentioned, the "mere conduit" and "safe harbour" provisions in the ECD and DMCA respectively operate on the basis that the intermediary has no actual knowledge of the infringement. Art. 2.1.2 of the leaked text of the agreement would eliminate the actual knowledge provision for injunctions, which may open up the liability floodgates once again.[102]

Why is all of this relevant for the long tail? After all, most of the attacks against intermediaries are against copyright infringers, not against tail-end content owners. Nonetheless, it seems clear that the long tail relies on search engines, content aggregators, and word of mouth for it to work. From all of the above evidence, it is clear that Internet intermediary services are under fire like never before. It seems obvious that a return to the days where intermediaries could be liable for content would serve as a chilling effect to the entire way in which the Internet operates. The attack against intermediaries is done precisely to keep the Pareto Law system going, a model in which content owners carefully choose the channels of distribution, but where they also control these channels. In the age of Google, content owners have lost this power and content is available through a growing number of legal and illicit sites. To make intermediaries liable for the content placed by users is to perpetuate a system that does not work well in cyberspace. It is no coincidence that some of the evidence debunking the long tail comes from the chief economist for the British Performing Right Society (PRS).[103] This is unsurprising, as the current framework of commercial content production has been built upon the assumptions of Pareto inequalities, and any change in the underlying business models could affect the existing regimes.

---

102. Guadamuz A, "How Will ACTA Affect UK Copyright Law?" *TechnoLlama* (July 15, 2010), http://www.technollama.co.uk/how-will-acta-affect-uk-copyrightlaw.
103. Supra note 48.

The copyright industry is trying to find its feet online. Its efforts should be directed at fostering the emergence of new business models, of which the long tail is just one. By looking at how users engage with content, copyright policy should keep this in mind, and policymakers should resist the siren calls of dying profit-seeking methods. A long tail copyright policy would keep intermediary liability to a minimum, so any change that leads us away from the current regime should be opposed.

4.2 Copyright, networks and P2P

4.2.1 Copyright and P2P

It may be needless to repeat that copyright infringement on the Internet is rife. If one was to take industry figures seriously (and that is a big if), by 2015 digital piracy will have cost the industry €32 billion EUR, and will have caused job losses of 611,300.[104] While scepticism about these figures is warranted,[105] there is also room for concern. The data about the effect of file-sharing on sales is still a hotly disputed economic argument. While some researchers place the effect at around 12 percent,[106] others have found practically zero impact.[107] Regardless of the actual losses that can be attributed to file-sharing, there can be little doubt that a situation where a generation of users is engaged in wilful copyright infringement is undesirable. If anything, unauthorised copying of other people's works is unethical, but also a situation where large parts of the population willingly flaunt the law because of personal choice creates a situation that is undesirable to say the least. Either we scrap copyright enforcement altogether, or we devise ways in

104. TERA Consultants, *Building a Digital Economy: The Importance of Saving Jobs in the EU's Creative Industries*, International Chamber of Commerce Report (2010), http://www.iccwbo.org/bascap/id35360/index.html.
105. Guadamuz A, "Critique of the ICC's Report on the Digital Economy in Europe", *TechnoLlama* (19 March, 2010), http://www.technollama.co.uk/critique-of-the-iccs-report-on-the-digital-economy-in-europe.
106. Zentner A, *Measuring the Effect of Online Music Piracy on Music Sales*, University of Chicago Working Paper (2004), http://economics.uchicago.edu/download/ musicindustryoct12.pdf.
107. Oberholzer-Gee F, Strumpf K, "The effect of file sharing on record sales: An empirical analysis", 115:1 *Journal of Political Economy* 1 (2007).

which most of these practices can be brought back to legality. The current situation does not seem to favour anyone.

Wherever one stands in the great copyright debate, it is clear that the existing landscape is unsustainable. Could network science tell us something about how to tackle file-sharing? Perhaps that would be too much to ask, but one thing for sure is that it can tell us where existing enforcement strategies have gone wrong. How this information is used will be up to the creative industries and policymakers.

From reading the evidence presented in section 3, it may be clear why an understanding of networks is vital when it comes to enforcement of illegal file-sharing online. P2P networks display strong power law behaviour, so it is baffling that while some of the evidence for this statement has been available for years, owners and policymakers seem to be completely unaware about the implications of such a fact, hence the failure to tackle widespread copyright infringement in digital systems. A cursory look at the case history of lawsuits against P2P networks showcases some astonishing ignorance both of the technology and of the implication of scale-free topologies.

Napster was the first P2P system to be subject to a lawsuit from content owners. In 2000, it was sued in the US by several music record companies for contributory and vicarious copyright infringement, and it eventually lost the case and subsequent appeal.[108] Because Napster was a mediated server-based network, it relied entirely on the existence of a centralised database in order to connect users to one another. This was the seed of its demise, as the network could not exist without the services provided by the company.[109] This is consistent with network theory, as the network relied entirely on one super-hub which connected all of the nodes in the system. By taking out the central hub, the network could not exist.

---

108. *A&M Records v. Napster*, 2002 U.S. App. LEXIS 4752.
109. Smith S, "From Napster to Kazaa: The Battle over Peer-to-Peer Filesharing Goes International", *Duke Law & Technology Review* 8 (2003).

The legal situation with decentralised client-based P2P networks was much more difficult. During a period between 2001 and 2008, several makers of P2P clients were also sued by content industry in various jurisdictions, but the most visible cases were in the US with Aimster,[110] Limewire[111] and Grokster,[112] and in Australia with Kazaa.[113] The common denominator of these cases was that initially, courts found it difficult to deal with the technology. These were services that featured a client connected to a wider network that ran independent from the client. It should be remembered that US courts had to take into consideration the Supreme Court decision of *Sony v Universal*[114] by applying the Sony Doctrine, which states that technologies that have substantial non-infringing uses cannot be held liable for secondary and vicarious liability. Both *Grokster* and *Aimster* were found to have substantial non-infringing uses, and the fact that the services were not centralised played in favour of the defendants.[115] *Grokster* eventually made it all the way to the Supreme Court, which came up with another doctrine, that of incitement to copyright infringement. Evidence was presented that Grokster was not only aware that the client was used to infringe copyright, but also that it encouraged such actions.[116] *Grokster* turned out to be a turning point for legal actions against client-based services, both Kazaa and Limewire were defeated in court, and all the early difficulties in enforcement against P2P networks seemed to be over.

However, all of these gains were not only short-lived, but proved to be the very definition of a pyrrhic victory. As it has been explained repeatedly, client-based P2P networks are not centralised, and consequently they could exist even if the company who made the client software disappeared. For example, Limewire allowed users to

110. *In re Aimster Copyright Litig*., 2004 U.S. App. LEXIS 1449.
111. *Arista Records LLC v. Lime Group LLC,* 2010 U.S. Dist. LEXIS 46638.
112. *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 125 S. Ct. 2764.
113. *Universal Music Australia v Sharman License Holdings* [2005] FCA 1242.
114. *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417.
115. Miles E, "In re Aimster & MGM, Inc. v. Grokster, Ltd.: Peer-to-Peer and the Sony Doctrine", 19 *Berkeley Technology Law Journal* 21 (2004).
116. Wu T, "The Copyright Paradox – Understanding Grokster", *Supreme Court Review* 229 (2005).

connect to the Gnutella network, which still exists even at the time of writing.[117] The key is that the lack of central servers means that these networks are exceptionally resilient. Attacking the client manufacturers did not really dent P2P usage. It is true that these networks are nowhere near as popular as they were when the lawsuits were filed, but the reason for this is not legal action, but the fact that users have migrated to BitTorrent.

The legal status of BitTorrent has been considerably more difficult to pin down. As it has been stipulated already, the technology itself is nothing more than an efficient way of distributing traffic loads between users, so it has enormous non-infringement potential. This would at lease serve to cover BitTorrent under the Sony Doctrine in the US and other jurisdictions where similar provisions exist. Nonetheless, it is clear that the BitTorrent protocol is also used for widespread online infringement. Who is liable when such actions occur?

To analyse the legal challenges of BitTorrent, it is essential to remember the nature of the technology. In a typical BitTorrent transaction, a person holding a digital copy of the work makes it available to the public by creating a tracker file and uploading it to a tracker site, thus advertising to the world that the work is available for download. The file then becomes a swarm, and potentially thousands of Internet users may be involved in exchanging parts of the file until they obtain a complete copy. In an excellent analysis of the legalities of BitTorrent, Rietjens[118] usefully identifies three key exclusive rights that come into play on each transaction:

1. *Reproduction*: Let us ignore first the origin of the copy; this might be a legitimate copy purchased through an online retailer, or it might be an unauthorised copy, in which case the person making the copy would be directly infringing copyright. Does a person making a full copy infringe the exclusive right to reproduce the work? The answer seems to be yes in most jurisdictions. For example, in Europe

117. For example, the Grokster 2 protocol boasts almost 200,000 users with at least 10 clients: http://crawler.trillinux.org/history.html.

118. Rietjens B, "Give and Ye Shall Receive! The Copyright Implications of BitTorrent", 2:3 *SCRIPTed* 364 (2005).

all that is required for infringement is that the reproduction takes place "in whole or in part",[119] and it would seem that BitTorrent copies would certainly infringe on this right.

2. *Distributing*: Art. 6(1) of the 1996 WIPO Copyright Treaty (WCT) establishes the exclusive right of the rights holder to distribute the original and copies of the work. This right, however, only exists for copies in tangible form, and for that reason does not cover digital copies.
3. *Making available to the public*: Art. 8 of the WCT determines the exclusive right of the copyright owner to make the work available to the public in a manner, time and place chosen by them. This is a much trickier question, is making the tracker file available to the public equivalent to making a full copy available to the public? Logic dictates that the answer is positive. A full copy is made available so to speak by the first seeder, subsequent members of the swarm exchange parts of the whole, and the end result is that a full copy is eventually present in the recipient's computer. So it seems like this is another exclusive right infringed by participants in a swarm, as they are making the work available to the public as well.

The main legal authority dealing with the legal nature of copyright infringement in BitTorrent sites is the aforementioned *Roadshow Films v iiNet.*[120] In this case, the judge presents some of the most clear-headed analysis of the legal issues surrounding the BitTorrent protocol. When looking at a typical BitTorrent transaction, it is hard not to think of the collective nature of the infringement. In previous P2P incarnations, enforcement was easier because individuals and organisations were more important to the end result of copying one work from one computer to another using the Internet. In a

119. Art. 2 of the Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society ("Infosoc Directive").
120. Supra note 91.

typical BitTorrent transaction, several things need to happen: there must be at least one seeder with the entire copy of the file, there needs to be a tracker site, there needs to be a client and there needs to be several users sharing the file. The judge in iiNet looked at these elements and rightly came to the conclusion that they constituted a plural entity, what he calls the "BitTorrent system" comprising the client, the tracker and the users.[121] Not only was that a plurality, but the judge asked whether each individual connection in a swarm should be considered an individual infringement instance, or if the entire swarm was the infringing action. He rightly answered that the individual connections were irrelevant, as the work was being shared and copied by multiple users, so each swarm would be an infringing instance for copyright purposes.[122]

From all of the above, it is easy to see why BitTorrent has been so difficult to enforce. At any given time there are millions of works being shared using the BitTorrent protocol. For example, at the time of writing, more than 31 million users were sharing 3 million torrent files in The Pirate Bay,[123] with more than 20 million seeders. Assuming that most of those files are infringing copies of copyright works, it would be impossible for copyright owners to try to sue so many users. P2P traffic online has only continued to grow. Depending on the methods used to measure Internet transfer, it has been said that P2P transactions can hit as high as 80 percent of all recordable online traffic.[124] How can they stop this P2P flood?

The obvious targets are the tracker websites. Taking a hint from previous victories against Napster and client-based services, representatives from the content industries attempted to attack The Pirate Bay which, as has been pointed out earlier, is the largest tracker site in the world.[125] In January 2008, Swedish prosecutors brought civil and criminal charges for the Swedish equivalent of secondary copyright infringement against

121. Ibid, paras 70–71.
122. Ibid, paras 310–311.
123. The statistics can be found at: http://thepiratebay.org/.
124. Parker A, *The True Picture of Peer-to-Peer Filesharing*, Report by Cache Logic (2005), http://www.cachelogic.com/research/p2p2004.php.
125. *Sony BMG Music Entertainment Sweden AB et al. v Niej et al*, Stockholm Tingsrätt, Case B 13301–06 (2009).

four individuals associated to The Pirate Bay.[126] The main question rested, unsurprisingly, on whether The Pirate Bay was guilty of making works available to the public without authorisation of its owners. The court found that the defendants were indeed guilty of providing a site with "sophisticated search functions, easy upload and storage, and a website linked to the tracker", sentencing each to a year in jail, and awarded 30 million SEK (approximately €2.7 million EUR) in damages and costs. In any normal situation, this would be a crippling indictment of the technology, but we are dealing with resilient networks here. The defendants had already fled Sweden, and by the time of the trial The Pirate Bay was hosted in several other countries.[127] Needless to say, the site is still operating at the time of writing.

If attacking the main services has proved futile, what else can copyright owners do? Their second strategy was to attack individual users. Starting in 2003, the Recording Industry Association of America (RIAA) began a controversial campaign of suing individual users for direct copyright infringement in P2P networks,[128] and later began a campaign of sending settlement letters to P2P users.[129] The International Federation of the Phonographic Industry (IFPI) followed suit by issuing thousands of complaints against individuals around the world.[130] While it is difficult to ascertain exactly how many thousand suits were filed and letters were sent, the campaign resulted in some high-profile PR disasters for the industry. Because they identified users via IP addresses, which is not an exact science by any stretch of the imagination, some of the suits were

126. Fredrik Neij, Gottfrid Svartholm, Peter Sunde and Carl Lundström.
127. Anderson N, "Pirate Party Hosting Pirate Bay in Pro-P2P Political Gesture", *ArsTechnica* (May 2010), http://arstechnica.com/tech-policy/news/2010/05/pirate-party-hosting-pirate-bay-in-pro-p2p-political-gesture.ars.
128. Ryan B, "Communication Breakdown: The Recording Industry's Pursuit of the Individual Music User, a Comparison of U.S. and E.U. Copyright Protections for Internet Music File Sharing", 25 *Northwestern Journal of International Law & Business* 229 (2004).
129. Marco M, "RIAA Bullies College Students with P2PLawsuits.com", *The Consumerist* (1 March, 2007), http://consumerist.com/2007/03/riaa-bullies-college-students-with-p2plawsuitscom.html.
130. Mook N, "IFPI Sues 8,000 P2P File Swappers", *BetaNews* (17 October, 1006), http://www.betanews.com/article/IFPI-Sues-8000-P2P-File-Swappers/1161101823.

issued against children, the elderly, and even the deceased.[131] There were also two very high-profile cases in the US against individual users which resulted in astoundingly high damages: *RIAA v. Tenenbaum*[132] and *Capitol v. Thomas*.[133]

All of these cases against individuals served to prove two points. Firstly, it painted a picture of greedy industry giants trying to squeeze as much money from users as possible, and created a perception of a David versus Goliath situation where the little man was being attacked by rich faceless corporations. Secondly, it served as further example that the fight against P2P file-sharing has no easy answers, as even despite all of these lawsuits, copyright infringement continues to exist.

4.2.2 How to stop worrying and learn to love P2P

Given all of the legal failures highlighted, can content owners learn something from network science?

There are two main areas where the scale-free nature of P2P networks would be relevant to legal efforts to curb infringement in those sites. The first is the built-in resilience of the network, and the second is that of cascading failures.

The fact that P2P networks are stable and resilient to random attacks has already been explained in detail, and their survival even after the most concerted legal attacks against the network seems to attest to the truth of that observation. This statement is consistent with what we know about the resilience of scale-free networks. Take, for example, the thousands of lawsuits and cease-and-desist letters sent by copyright owners against individual file-sharers. Even if each targeted lawsuit and letter managed to change the behaviour of those individuals, the chances are these were not central hubs in the network, and consequently their removal from the system has no overall effect on the whole. The removal of random items is not the only element that matters in legal attacks

131. See for example: "Grandmother piracy lawsuit dropped", *BBC News* (25 September, 2003), http://news.bbc.co.uk/1/hi/entertainment/music/3140160.stm.
132. *Sony BMG Music Entertainment v. Tenenbaum*, 672 F. Supp. 2d 217.
133. *Capitol Records, Inc. v. Thomas*, 579 F. Supp. 2d 1210.

to P2P networks, but even massive adversarial failure of nodes in a scale-free topology does not result in the destruction of the network.[134] P2P networks not only display this resilience, but they can even be designed to be more resilient to such attacks.[135]

During the heyday of the client-based network model, copyright owners undertook a different approach to that of the strict legal route, namely the "poisoning" of P2P systems by introducing decoys into the network. These are fake, faulty and otherwise unusable copies of copyright works.[136] The idea was that the presence of poisoned files would somehow affect the reliability of the network, and would deter future users. While there is research that confirms that such attacks could have had some limited effect,[137] it is clear that they did not affect the networks to the extent that they would be destroyed. The continuing existence of the networks seems to support that hypothesis.

Is there any sort of non-legal attack that could theoretically bring down a P2P network? The answer is yes, and here we move to the topic of cascading failures. In order to refresh this concept, a cascading failure can occur in a power law topology when key vertices in a network affect other vertices; in other words, if hubs and super-hubs that glue the network together are taken out, this could have a downstream effect on relying nodes within the system. Dumitriu et al[138] conducted a simulation on types of attacks that would result in a cascading failure of a P2P network. In their study, they accounted specifically for the presence of power-law graphs in P2P networks, and included graph-theory concepts into their attacks, such as protocol properties, graph properties, client-based counter-strategies, and even user behaviour. Their model is both

134. For examples of such networks, see: Saia J et al, "Dynamically Fault-Tolerant Content Addressable Networks," *IPTPS* (March 2002); and Fiat A and Saia J, "Censorship Resistant Peer-to-Peer Content Addressable Networks," *Symposium on Discrete Algorithms* (2002).

135. Loguinov D et al, "Graph-Theoretic Analysis of Structured Peer-to-Peer Systems: Routing Distances and Fault Resilience", *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (2003).

136. Christin N, Weigend AS and Chuang J, "Content availability, pollution and poisoning in file sharing peer-to-peer networks*", Proceedings of the 6th ACM conference on Electronic Commerce*, Vancouver, Canada (2005).

137. Ibid.

138. Dumitriu D et al, "Denial-Of-Service Resilience in Peer-To-Peer File Sharing Systems", 33:1 *ACM SIGMETRICS Performance Evaluation Review* 38 (2005).

elegant and simple, to introduce malicious nodes into the system that would overwhelm nodes and hubs. They explain:

> [W]e develop and study a new class of attacks designed to collapse a p2p network's goodput. In such an attack, a malicious peer modifies replies to queries for any file, before it forwards them to the client. In a "false reply attack", the malicious peer points the client to itself. When the client then requests a download from the malicious peer, it presents a corrupted copy of the file, forcing a repeated request and download in order for the client to obtain the true file. [...] Even a small percentage of nodes in a large-scale system can represent 100s or 1000s of hosts. We note two mechanisms by which attackers can control numerous hosts. First, the attacker can deploy all malicious nodes itself at a single or multiple Internet Data Centers. A second way to launch an attack is by subverting peers via a "trojan horse" program that serves corrupted content. Trojan horse programs are already common on both the Internet (e.g., those spread via email viruses, worms, and the web) as well in p2p systems.[139]

While devastating, this sort of attack would itself stretch the borders of legality. Moreover, this sort of attack seems to work much better in the second generation client-based P2P model, and would not work with BitTorrent.

In my first publication and subsequent presentations on this topic, I must admit that the cascading failure solution to P2P filesharing seemed to be the most viable option for content owners.[140] My own argument was that while scale-free networks are resilient, they also can be the subject of catastrophic failures when a vital super-hub in the system collapses. However, this initial assessment completely underestimated the level of resilience of P2P networks, particularly the resilience displayed by the BitTorrent protocol. Try to imagine an attack like the denial of service strategy described above on a BitTorrent file, and you will begin to see what the problem is. BitTorrent files do not rely on a semi-central client, so even if one were to poison some seeds, this would not affect a swarm as a whole. And even if a swarm was brought down, there would still be dozens of other swarms sharing the same work.

---

139. Ibid, p.39.
140. Guadamuz A, "Scale-Free Law: Network Science and Copyright", 70:4 *Albany Law Review* 1297 (2008), pp.1309–1311.

This does not mean that copyright owners have stopped trying. Poisoning has been also introduced to BitTorrent sites, where "fake" copies of a work are distributed all the time. However, sharing communities have found a way around this. Many tracker sites have the possibility of rating and commenting on specific files, so it is possible to tell the community if a file is fake, or if it has a virus, or if it does not work as intended. Peer-to-peer meets peer review.

There is only one answer to the BitTorrent conundrum. As it has been repeatedly stated, BitTorrent is not completely decentralised, it relies heavily on the existence of tracker sites such as The Pirate Bay. The obvious solution would be to take out or restrict access to these sites. However, this is not as simple as it sounds, as the example of legal action against The Pirate Bay exemplifies. It is perfectly possible that by the time you are reading this, The Pirate Bay no longer exists, but chances are there is another tracker site that has taken its place. Literary references to genies out of bottles, cats out of bags and opening forbidden boxes apply here. The technology exists, and it will be almost impossible to undo.

Based on all of what we know about network theory, the conclusion that should be reached by content owners is that – at least in the short-term – P2P networks cannot be brought down easily. Attacks against tracker sites and against individual file-sharers are not having any effect; their resilience is quite simply insurmountable, at least for the time being.

Nonetheless, there are possible scenarios that might work, but they might involve a change of how we see the Internet. Back in Chapter 3 we discussed how some countries have created their own separate Internet by means of regulation of the Internet entry-points into their countries, the chokepoint model of regulation. Under this scheme, it would be possible to attempt to restrict national access to certain sites. This strategy can be eventually scaled down to individual ISPs, whereby all of the addresses connected to a specific site would be blocked; this strategy has been used in several countries, from

Ireland to Denmark, with varying degrees of success.[141] National firewalls, however, can be easily circumvented by anonymisers, Virtual Private Networks, and other similar tools.

This game of cyber cat and mouse is wasteful once one understands the underlying topologies of the vast copyright network. Content owners are perennially engaged in the futile exercise of chasing users, service providers, P2P clients and tracker sites; all to no avail. The networks are too robust for traditional legal enforcement, and may even be too resilient for technical solutions unless the distributed nature of the Internet is somehow changed. The solution may be not to give in completely, but to try to use the existence of P2P networks to one's advantage. In a seminal work on copyright in the Digital Age, William Fisher proposes a compensation system for copyright owners. He says:

> A creator who wished to collect revenue when his or her song or film was heard or watched would register it with the Copyright Office. With registration would come a unique file name, which would be used to track transmissions of digital copies of the work. The government would raise, through taxes, sufficient money to compensate registrants for making their works available to the public. Using techniques pioneered by American and European performing rights organizations and television rating services, a government agency would estimate the frequency with which each song and film was heard or watched by consumers. Each registrant would then periodically be paid by the agency a share of the tax revenues proportional to the relative popularity of his or her creation. Once this system were in place, we would modify copyright law to eliminate most of the current prohibitions on unauthorized reproduction, distribution, adaptation, and performance of audio and video recordings. Music and films would thus be readily available, legally, for free.[142]

It may be strange to suggest a solution that ignores entirely the evidence from network theories so far, but it makes sense precisely because it would work regardless of the architecture and network behaviour described above. The only real legal solution at the

---

141. "Danish ISPs to Fight the Pirate Bay Block", *Torrent Freak*. (5 February, 2009), http://torrentfreak.com/danish-isps-to-fight-the-pirate-bay-block-090205.
142. Fisher WW, *Promises to Keep: Technology, Law, and the Future of Entertainment*, Stanford, CA: Stanford Law and Politics (2004), p. 205.

moment would be one that embraces the scale-free nature of P2P networks, and that works regardless of the underlying resilience. Fisher's proposal is compatible with what we know about networks. Not only that, it could work better once one understands concepts such as small worlds, hubs, and robustness.

I would also like to end this chapter on a positive note that unites the long tail and P2P. A study into P2P file-sharing has unearthed the fact that sharing does indeed seem to affect music sales from top earners.[143] Blackburn conducted research trying to ascertain what would be the effect for music sales of a reduction of file-sharing volumes by 30 percent. For top earners at the head, the result was a marked increase in sales. However, for those with minimum sales, decreasing file-sharing actually had a negative impact on sales.[144] If this data is accurate, then it could be said that P2P is good for the tail, but bad for the head.

P2P and the long tail may very well be the saviours of the creative industries.

143. Blackburn D, *On-line Piracy and Recorded Music Sales*, Working Paper, Department of Economics, Harvard University (2004), http://www.katallaxi.se/grejer/blackburn/ blackburn_fs.pdf.
144. Ibid, pp.45–46.

6. Peer-production Networks

> In many of the more relaxed civilizations on the Outer Eastern Rim of the Galaxy, the Hitchhiker's Guide has already supplanted the great Encyclopaedia Galactica as the standard repository of all knowledge and wisdom, for though it has many omissions and contains much that is apocryphal, or at least wildly inaccurate, it scores over the older, more pedestrian work in two important respects.
> First, it is slightly cheaper; and second, it has the words "DON'T PANIC" inscribed in large friendly letters on its cover.
> Douglas Adams, *The Hitchhiker's Guide to the Galaxy*[1]

In April 1999, author Douglas Adams founded a website called H2g2.com in honour of his successful series *The Hitchhiker's Guide to the Galaxy*. The stated purpose of the website was to create an informal guide to "life, the universe, and everything", a sort of online encyclopaedia edited and maintained by its users. The project was eventually taken over by the BBC in 2001, where it is still hosted. You may be forgiven for not having heard of this project, as its functions are replicated by a more recent online collaborative project called Wikipedia. Created in 2001, Wikipedia has become everything that Douglas Adams pretended for his online experiment. The online encyclopaedia boasts more than 12 million articles in 262 languages, of which 25 have more than 100,000 entries.[2] As a manner of comparison, the online edition of the *Encyclopaedia Britannica* contains 120,000 articles.[3]

Why has Wikipedia succeeded where Douglas Adams failed? Is it the technology? Is it the bottom-up approach of the wiki model? Is it serendipity? It could be argued that the reason for Wikipedia's success can be attributed to the rise of what is described as

1. Adams D, *The Hitchhiker's Guide to the Galaxy*, New York: Pocket Books (1981), p.2.
2. Wikipedia, *Wikipedia*, http://en.wikipedia.org/wiki/Wikipedia.
3. http://en.wikipedia.org/wiki/Wikipedia:Size_comparisons#Comparison_of_encyclopedias.

the participatory Web. For large periods of human cultural history, the process of creative creation has been subject to strict control of distribution channels. The status quo has rested on the assumption that only a few people can produce worthwhile works, and it has been up to publishers to serve as the judges of what should be communicated to mass audiences. This top-down approach does not only exist in cultural works, but can be seen in almost all facets of intellectual creation, from academic output to the production of software.

One of the many aspects where the Internet has prompted societal change has been that it has challenged the existence of this top-down approach. Distribution channels have been democratised, and it is now easier for hobbyists and amateurs to bypass the gatekeepers and make their works available to wider audiences. It is not the remit of the present work to judge the wisdom of the new bottom-up approach, but the existence of wider participation mechanisms cannot be denied. Wikipedia, Flickr, YouTube, blogging, Facebook, Twitter and many other social media have developed an environment of unparalleled creative momentum. This chapter will explore the phenomenon using complexity theory, trying to analyse whether the existence of peer-production and social media can be explained through some of the theories studied previously.

From the legal perspective, the subject of peer-production may seem of somewhat less importance than the commercial copyright subject dealt with in the last chapter. However, there are several aspects of the emergence of user-generated content where the tools of network science can be useful from a policy and user perspective in the copyright arena. The first is the subject of licensing of these works in the shape of open licences. The second is a more important policy question about the role of peer-production in the Pareto-driven copyright policies that we have in place at the moment. This chapter will concentrate on these two legal issues.

1. THE RISE OF PEER-PRODUCTION AND THE USER-GENERATED WORLD

1.2 Defining peer-production

Peer-production is a term that defines a particular form of developing content. In its most basic form, it simply describes a way in which individuals, firms and organisations come together to produce intellectual creations. As it was discussed in the previous chapter, the traditional view of content creation rests heavily on the idea of the creator as a struggling individual who requires protection in order to incentivise creativity.[4] The words "creation", "owner" and "author" have therefore been co-opted by the copyright industries in order to justify a particular commercial model. The reality, also discussed in the previous chapter, is that content creation is a collective effort involving a plurality of individuals, publishers and distributors. This responds to very specific business model, but also it exists as an institutional organisational process that distributes profits and allocates resources. Peer-production is a term that defines an alternative model to the existing system.

The term peer-production was first used by Yochai Benkler,[5] who was interested in looking in more detail at the paradoxical success of non-proprietary methods of developing software that seemingly ignored profit and Pareto's Law from the equation. In order to create a theoretical framework that could explain the reasons why individuals co-operate, write and distribute software seemingly without profit in mind, he looked at some of the ideas that explained the way in which individuals come together to form a collective organisational entity. He relied heavily on the work of economist Ronald Coase, who had established one of the principles of organisational economics – namely that if the cost of coming together as a collective was lower than the alternative, then individuals would create a group to achieve their goals, to reduce costs and enhance

4. Chapter 5, s. 1.
5. Benkler Y, "Coase's Penguin, or, Linux and the Nature of the Firm", 112:3 *The Yale Law Journal* 78 (2002).

productivity.[6] Under Coase's original model, the copyright industries could be described as entities that came together because the cost of doing business together was lower than the cost of individually negotiating and producing content. Benkler's great insight came in stating that there are different ways in which the production of works can be achieved. While traditionally this organisation has been reached by hierarchical and directed approaches, the cost of producing content through the undirected and non-hierarchic models is lower than the alternative. He comments:

> A distributed peer production model allows individuals to self-identify for tasks for which they are likely to be the best contributor. This makes peer production particularly well suited to organize activities in which human capital is the dominant input, as long as the coordination problems can be solved and that some mechanism to adjust for individuals' errors regarding their own suitability for a job is implemented.[7]

Without really thinking about it in that context, what Benkler really is talking about is self-organisation.

A related, yet more radical approach has been commented upon by Eben Moglen,[8] one of the main theorists of the free and open source movement. Moglen accurately pinpoints one of the most glaring problems faced by the commercial content paradigm, that which is succinctly encapsulated in the dominance of Pareto distributions. This is the fact that currently, commercial exploitation of content relies heavily on a system that brings together individuals that produce works, but the bulk of the profits go to a few. Moglen states:

> To the owners of culture, we say: You are horrified at our intending to do away with private property in ideas. But in your existing society, private property is already done away with for nine-tenths of the population. What they create is immediately appropriated by their employers, who claim the fruit of their intellect through the law of patent, copyright, trade secret and other forms of "intellectual property". Their

6. Coase R, "The Nature of the Firm", 4:16 *Economica* 386 (1937), p.403.
7. Benkler, supra note 5, p.6.
8. Moglen E, *The dotCommunist Manifesto*, (2003), http://emoglen.law.columbia.edu/publications/dcm.html.

> birthright in the electromagnetic spectrum, which can allow all people to communicate with and learn from one another, freely, at almost inexhaustible capacity for nominal cost, has been taken from them by the bourgeoisie, and is returned to them as articles of consumption–broadcast culture, and telecommunications services--for which they pay dearly.[9]

Forget the political undertones for a moment, and you will see that this is indeed a powerful description of the reason why the Internet has become such a disruptive force to the existing business models. When users become producers, we have reached a stage of fundamental change in the information economy. Whenever you read about one of the 20th century copyright industries complaining about the Internet, you need to keep in mind the fact that this change has been brought about by a shift in the economy of information.

Peer-production is a term that is often used interchangeably with other concepts that describe the creation of intellectual works by members of the public instead of the professional industries that have dominated for more than a century. You will hear the phenomenon described interchangeably as user-generated content (UGC), and even the so-called Web 2.0. UGC is usually used to refer to the actual content created by users,[10] as the name clearly indicates; while Web 2.0 is often defined as the set of tools used to create such content.[11] In the end, what we are seeing is that each term describes three aspects of the Internet content revolution, peer-production describes the process, UGC describes the content and Web 2.0 defines the tools. While these terms have nuanced meanings, it will be assumed that in the end they are all describing the same phenomenon, namely, a change in the organisation of the creation of content that includes larger numbers of users. So, when we talk about peer-production we are fundamentally talking about all three elements.

---

9. Ibid.
10. Lee E, "Warming up to User-Generated Content", 5 *University of Illinois Law Review* 1459 (2008).
11. O'Reilly T, "What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software", 1 *Communications & Strategies* 17 (2007).

Benkler brings together all of these in his definition of peer-production. In his influential work *The Wealth of Networks*, he sets an impressive framework for cultural means of production and the participatory Web. He states:

> This cluster of phenomena, from free and open-source software to Wikipedia and SETI@Home, presents a stark challenge to conventional thinking about the economics of information production. Indeed, it challenges the economic understanding of the relative roles of market based and nonmarket production more generally. It is important to see these phenomena not as exceptions, quirks, or ephemeral fads, but as indications of a fundamental fact about transactional forms and their relationship to the technological conditions of production. It is a mistake to think that we have only two basic free transactional forms – property-based markets and hierarchically organized firms. We have three, and the third is social sharing and exchange. It is a widespread phenomenon – we live and practice it every day with our household members, coworkers, and neighbors.[12]

The social element of sharing is precisely what had been missing in the previous assumptions about cultural production. Profitability in Coase's organisational economics cannot explain the wealth of production that we are currently witnessing. Social transfer of information plays an important role in our everyday lives; the Internet has allowed us to extend the social networks and allows users to share their creations with others.

Another theorist of peer-production phenomenon is Cass Sunstein. In his book *Infotopia*[13] he sees several problems with what he perceives as the growing balkanisation and atomisation of information, but still he opines that its potential is great. He argues that:

> [T]here are remarkable exercises in the development of cumulative knowledge, producing an astonishing range of new goods and activities. We shall see that some of the underlying methods are novel and exceedingly dramatic. They will be used far more ambitiously than they now are. With respect to the aggregation of information, we are in the first stages of a revolution.[14]

---

12. Benkler Y, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven, CT; London: Yale University Press (2006), pp.462–463.
13. Sunstein CR, *Infotopia: How Many Minds Produce Knowledge*, Oxford: Oxford University Press (2008), p.7.
14. Ibid, pp.8–9.

While there is much room for agreeing with the concept of peer-production described above, it is only fair that one should sound a word of caution. It would seem that in response to the dominance of Pareto, we could be responding with the myth of social production. While peer-production is an important development, one should not forget that the marketplace of ideas is still very much reliant on production inequalities. While none of the authors cited above has hinted that peer-production will replace commercial approaches, the outlook and impact of social media seems overstated. A common feature of those who tend to analyse new technological trends (the author included) is that we operate in a scholarly environment that is already familiar with the technologies involved. Could we be guilty of becoming too enamoured with peer-production because we are part of a group that constitutes its core user base?

This is what one could design as "If you build it, they will come"[15] argument. There is a growing trend to assume that by simply building a blog, a wiki or a Twitter stream, users will flock to a site and immediately participate and engage with the content. While this is mostly a personal anecdote based on a long list of failed UGC experiments, it is important to try to address this often overlooked fact about Web 2.0 applications. Do the successful showcases such as Wikipedia, YouTube and Twitter distract us from the long digital graveyard of content nobody has ever seen or cared for? Further sections will try to measure the true relevance of the peer-production economy to try to answer this valid question.

1.2 Clash of cultures

In a conference I attended in 2007, which dealt with licensing issues of peer-production, one of the legal keynote addresses was given by a representative from the International Confederation of Authors and Composers Societies (CISAC). Presenting to a potentially

15. Named so by the famous tagline used in the Kevin Costner movie *Field of Dreams*, http://www.imdb.com/title/tt0097351/.

hostile audience, the CISAC representative gave a balanced view of the interaction between collecting societies and open licences, and also commented that the history of collecting societies proves why collective management of intellectual property benefits creators. The reason for showcasing this seemingly innocent comment is that it exemplifies a view of the process of creation, and of intellectual property in general, as something that is only done with commercial interest in mind; and the fact that it was given to an audience convened specifically to talk about creation outside of the existing frames of reference displayed precisely why some in the copyright industries simply do not seem to get that the rules have changed. To think of creators as only those who make a profit or make a living may seem like common sense, but it is a view that ignores the fact that the Internet has changed the creative process. The advent of easy-to-use tools that allow the publishing of text, video, photographs, music and all other forms of digitised media have brought about a new generation of creators who are less interested with traditional distribution channels and are willing to explore other methods. We have all become potential publishers.

As it was discussed in the previous chapter, existing copyright law and policy has been drafted with the assumption that Pareto's Law is the only game in town. Copyright maximalism is a legal approach that makes sense when the vast majority of content is created with a specific purpose in mind, that of making profits based on a small number of superstar hits. The existing commercial model of creation was driven by the Pareto distribution model which relies heavily on thorough filtering by a minority. What is taking place nowadays is something which quite literally does not fit any of our previous ideas of what drives people to produce and maintain an intellectual work. There is a clash of cultures, of which the changes in commercial production of works is just the tip of the iceberg.

What may explain the clash is that there has been a fundamental change in how some people approach the production and management of resources in the information age. From a theoretical perspective, what we are talking about is in its very core about self-

organisation. If we take the way in which society generates information as self-organisation, then the way in which content is produced will be dependent on the prevalent conditions for it to occur. Up until recently, that predominant paradigm was through a system of a minority of creators who submitted works to intermediary distribution channels. Because of commercial constraints, only a few of those were distributed to the public on a large scale, be it through the printed press, music industry, audiovisual creations and art exhibitions, just to name a few. The majority of users were consumers.

With the Internet, such constraints do not exist. The price to enter the marketplace has been considerably diminished; all one needs is an Internet connection and the willingness to learn how to use basic tools in order to produce and publish content. Therefore, societal conditions and limitations have shifted enormously, which in theory should prompt the emergence of new models in true self-organising fashion. Nobody has dictated that people should produce more content, they just can. And they do.

This shift has resulted in a revolution in the process of content production the likes of which dwarf all other periods of human creativity. Google's CEO Eric Schmidt is given to some rather flamboyant and controversial statements, but recently he made a point that should place the current information revolution in perspective. He commented that since the dawn of civilisation and up to 2003, the total amount of information created was 5 exabytes.[16] However, by 2010 we have reached a point where collectively we produce 5 exabytes of information every two weeks.[17] By any measure, such a statement should prove to be a sobering thought.

Everywhere we look on the Internet, we can see examples of the vast amount of information that is being produced by individuals located outside of the traditional framework based on the creator/publisher economy. Lessig also has commented on the existing shift, and he places peer-production and user-generated content in terms of a

16. An Exabyte is 1018 bytes, one quintillion bytes, or a ten followed by 18 zeros.
17. Kirkpatrick M, "Google CEO Schmidt: 'People Aren't Ready for the Technology Revolution'", *ReadWriteWeb Blog* (4 August, 2010), http://is.gd/eNTTN.

clash between the read-only culture (RO) and the read/write culture (RW).[18] Lessig explains that for much of human history, the norm was to build upon cultural works, and therefore the act of creation was mostly a communal process. However, the Western idea of copyright that places the single creator at its centre prevailed, and therefore we were stuck with the read-only proprietary model. New technologies have made it possible to go back to a more organic way of cultural exchanges through the remix ethos that permeates much of the UGC universe. While Lessig claims that both cultures can coexist, an interesting feature of some of his work is to stress the relative importance of the RW culture by framing it in constant clash with its RO counterpart.

To stress this point, let us contrast some traditional content industries with their Internet-driven peer-produced counterparts. For example, in software a way to measure the amount of work that has gone into a program is to look at the single lines of code (SLOC), the individual lines of instructions that constitute it.[19] This is useful because it allows one to have an accurate estimate of the amount of programmer-hours that have gone into the development of a piece of software. Windows Vista is calculated to have approximately 50 million single lines of code, while Debian 4.0, an open source operating system released the same year, had 213 SLOC.[20] Peer-production can produce content in a manner that dwarfs commercial content.

Take also the newspaper industry; by 2010, there were 12,297 daily newspapers in circulation around the world.[21] Contrast that with the numbers of web blogs around the world (145 million at the time of writing),[22] and you may begin to see that there is a fundamental shift in how we view information delivery.

---

18. Lessig L, *Remix: Making Art and Commerce Thrive in the Hybrid Economy*, London: Bloomsbury Academic (2008), pp.28–29.
19. Wheeler DA, *More Than a Gigabuck: Estimating GNU/Linux's Size*, (2002), http://www.dwheeler.com/sloc/redhat71-v1/redhat71sloc.html.
20. Amor JJ et al, "Measuring Etch: The Size of Debian 4.0", *Debian Conference* (2007), https://penta.debconf.org/~joerg/attachments/33-measuring_etch_slides.pdf.
21. World Association of Newspapers, *World Press Trends 2010*, (2010), http://www.wan-press.org/worldpresstrends2010/home.php.
22. According to BlogPulse's daily blog survey: http://www.blogpulse.com/.

This clash of cultures does not mean that commercial creation and distribution of content is dying out; one can say that our culture is still dominated by Pareto's Law paradigm. One should also be careful when making quantitative and qualitative analysis of content figures; it may be true that there are considerably more blogs than newspapers, but nobody would contend that one can completely forego professional journalism and substitute it with blogs. There is still room for the prevalent model to exist, even if the rules have changed. What is being said here is that there has been a democratisation of content production, that has to be taken into account by policymakers and legislators.

Given this environment, it is no surprise that the enduring myth of the struggling creator has been translated into copyright policy. However, it is important to repeat a relevant example to stress the point that although there is a shift in content creation, policymakers do not seem to have noticed the changes. The UK government conducted a review of the content industries entitled Digital Britain,[23] which established the government's Internet regulatory strategy for the next decade. The report tackled several topics, including content. Unsurprisingly, the UK government's strategy with regards to content seems to be geared towards enforcement, piracy and unlawful use. What is surprising is that a report looking at the future of content in the digital environment practically ignored user-generated endeavours. The only mention of Web 2.0 and peer-production is in the glossary, which simply glosses over the rich opportunities brought by the participatory web by insisting on the outdated view of the top-down content provider. There is an obligatory mention to UGC and YouTube, but then the drafters have no idea what to do with it other than to mention that digital technologies lower barriers to new providers such as "the wide range of services now catering to ethnic minority communities and to specialist interest, the development of community services, of user-generated content whether on YouTube or on social networking sites".[24] In one

23. Department for Business, Enterprise and Regulatory Reform, *Digital Britain: The Interim Report* (2009), http://www.culture.gov.uk/images/publications/ digital_britain_interimreportjan09.pdf.
24. Ibid, p.45.

dismissing paragraph the UGC revolution is relegated to fringe status. It is disheartening that whenever it talks about content, it is talking about institutional content.

Interestingly, another result of the perpetuation of the myth of the commercial creator is that it has resulted in a clash between traditional media and Web 2.0 services. We are currently experiencing conflict between the top-down business models, and the organic business model championed by aggregated service providers and UGC application builders and managers, such as Google, YouTube and Facebook. Whichever way one would like to define peer-production, it is clear that its meteoric rise has been fuelled by the widespread availability of popular tools and applications that allows users to easily upload content online. It is no coincidence then that as traditional media sees their fortunes dwindle, content aggregator giants such as Google have benefitted from the popularisation of their services, at the same time as they promote its spread to ever growing sectors of the public. This conflict has culminated in legal challenges mostly against Google; such as the case of *Viacom v YouTube,*[25] and the recently settled cases against Google Books.[26] While the legal conflict between intermediaries and content owners has already been covered in the previous chapter, it must be stressed that the common denominator in these suits is that there is a palpable reaction against what is often described as parasitic practices by new media. Alongside the myth of the creator, a new one is arising, that of participatory technologies as leeches that thrive while the real content creators struggle. Perhaps the most prominent example of this was an article in the UK newspaper *The Observer* by Henry Porter. He commented:

25. *Viacom International, Inc. et al v. Youtube, Inc. et al* (United States District Court for the Southern District of New York, filed 13 March 2007, case no. 1:2007cv02103).
26. *Authors Guild v Google Inc* (United States District Court for the Southern District of New York, Docket No 2005 CV 8136, filed September 20, 2005); and *McGraw-Hill Cos, Inc v Google Inc* (United States District Court for the Southern District of New York, Docket No 2005 CV 08881, filed October 19, 2005).

> Despite its diversification, Google is in the final analysis a parasite that creates nothing, merely offering little aggregation, lists and the ordering of information generated by people who have invested their capital, skill and time.[27]

This seems to imply that anyone who does not create content and simply offers an aggregating service is by definition a "parasite" that has done no investment whatsoever. Google has invested large amounts of money in creating a vast and complex infrastructure that allows its users to access, create and aggregate content. These are tangible, useful and valuable services that make it easier for content creators to get their message across. Moreover, users are creating their own content.

What seems clear is that the RO and RW cultures, to use Lessig's terminology, will continue to be in conflict, as the RW culture gains more ground and the RO model continues to lose share in the market.

2. OPEN LICENSING

The changes discussed in the previous section are mostly important from a policy perspective, as the growth of peer-production has come to clash with policies informed by maximalist agendas proposed by the copyright industries.

There is, however, a legal issue that is crucial to the topic of peer-production, and this is the subject of open licensing. Traditionally, the commercial content industries have relied on strong copyright protection in order to distribute works. The protection is done through copyright law, but also through licensing. At the very basic level, a licence allows users to perform actions that would otherwise not be permitted under copyright law.[28] The most common form of copyright licence is one that gives users minimum

---

27. Porter H, "Google is just an amoral menace?" *The Observer* (5 April, 2009), http://www.guardian.co.uk/commentisfree/2009/apr/05/google-internet-piracy.
28. Guadamuz A, "The License/Contract Dichotomy in Open Licenses: A Comparative Analysis", 30:2 *University of La Verne Law Review* 101 (2009).

permissions, and where all other copyright is reserved, so they are "all rights reserved" licences.

Open licences are copyright licences that allow considerably more permissions to the user. If we have a spectrum of rights with "all rights reserved" licences at one end, and no protectionat the other (the work is in the public domain), open licences come somewhere in the middle. These licences still reserve some rights, but give users many others, what some call the "some rights reserved" licence.[29] There is ample literature describing this model, but this section will provide a basic introduction to the concepts, as their relevance will hopefully become evident when discussing the regulatory implications of peer-production and complexity later on.

2.1 Defining openness

Openness has almost become a buzzword attached to various movements and ideas. Open source, open content, open science, open standards, open databases; the very use of the work describes an opposition with networks that are closed by definition. But what does the concept mean in a peer-production context?

At its core, the definition of openness in peer-production is opposed to the concept of closed content. Under restrictive "all rights reserved" copyright regimes, the work that is being distributed comes attached with all of the exclusive rights allocated to its creator (or more accurately, its owner) through copyright law. These include, amongst others, the exclusive rights to copy, publish, distribute, execute, make derivatives and communicate the work to the public. If this is the definition of a closed work, then by definition an open work will, through the use of an open licence, allows users to perform some or all of the aforementioned exclusive rights.

29. Kansa EC, Schultz J and Bissell AN, "Protecting Traditional Knowledge and Expanding Access to Scientific Data: Juxtaposing Intellectual Property Agendas via a Model", 12:3 *International Journal of Cultural Property* 285 (2005).

The concept of openness in this context is historically related to the concept of freedom.[30] The father of the "free" movement was Richard Stallman, a software developer who became disillusioned with the collapse of what was perceived as an earlier golden age of programming where most code was shared between small numbers of creators. Stallman explains that software began to have restrictions imposed in the shape of proprietary licences that told users that they could not access the source code to modify the software, or share it with other people with the purpose of enhancing its functionality. If the user engaged in any tinkering with the code, then he/she stopped being a hobbyist and became a pirate.[31] Eventually, Stallman and other like-minded programmers created a powerful software development force under the general principles of non-proprietary software.

One of the main proponents of the idea of freedom as expressed by Stallman has been the Open Knowledge Foundation. In their definition of freedom, they have identified the main characteristics that a "free" (and consequently, an open work) should have. These freedoms are:

- the **freedom to use** the work and enjoy the benefits of using it
- the **freedom to study** the work and to apply knowledge acquired from it
- the **freedom to make and redistribute copies**, in whole or in part, of the information or expression
- the **freedom to make changes and improvements**, and to distribute derivative works.[32]

By definition, all licences that propose to be free contain all of these freedoms. While there is a philosophical argument between the users of the term "free" and "open", particularly in the software arena,[33] most of the times both camps are talking about the same core principles.

---

30. Moody G, *Rebel Code: Linux and the Open Source Revolution*, London: Penguin (2002).
31. *Revolution OS*, Directed by J.T.S. Moore, (2001).
32. Open Knowledge Foundation, *Freedom Defined*, (2008), http://freedomdefined.org/ Definition.
33. For a discussion of the differences, see: Guadamuz A, "Viral Contracts or Unenforceable Documents? Contractual Validity of Copyleft Licenses", 26:8 *European Intellectual Property Review* 331 (2004).

2.2 Free and open source software

There is a considerable amount of literature which deals with the subject of free and open source software, so this section will only serve as a small introduction for those unfamiliar with the concept.[34] There are two common names given to the application of peer-production models to software development, free software and open source software, which result in the compromise name free and open source software (FOSS).[35] Contrary to popular misconception about the movement, it is important to note that FOSS is not necessarily free of charge, and that it is not a movement opposed to traditional intellectual property protection.

In its more general form, FOSS is simply defined as software which is subject to later modifications by the user or other developers by allowing free access to its source code.[36] In this light, non-proprietary software is considered such if it is released under a permissive licence that allows such later modifications to the work, also known as "forks". These licences not only allow others to make their own modifications, but also permit users to distribute them accordingly. It is also understood that FOSS licences allow a wider range of rights to consumers that they would otherwise have, such as making copies of the work, or installing and distributing the software.[37]

FOSS licences cover a large spectrum of legal approaches and philosophies; and therefore a classification of licences is difficult. A survey conducted by the author found

34. For a more detailed introduction, see: Guadamuz A, "Free and Open Source Software", in Edwards L and Waelde C (eds), *Law and the Internet*, 3rd ed, Oxford: Hart (2009), pp.359–391; and Phillips DE, *The Software License Unveiled: How Legislation by License Controls Software Access*, Oxford: Oxford University Press (2009).
35. Also often referred to as FLOSS, Free Libre Open Source Software, adding the French and Spanish word for free as in freedom.
36. Source code is the programming statement expressed in a programming language that exists before the program is compiled into an executable application. The executable form of the software is generally known as the object code, and can only be read by the machine.
37. Guadamuz, supra note 34.

131 software licences that could be defined as either free or open.[38] However, all FOSS licences share some common elements. These are:

- *Attribution*. Copyright notices are to be kept intact, and the author(s) will be attributed in the code.
- *Access to the source code*. This is the most basic common element in all licences. The source code will be included either with the distribution, or is to be made available to the public in an open source repository.[39]
- *User rights*. Users are granted a non-exclusive right to use, copy and distribute the work.
- *Derivatives*. All FOSS licences allow developers to make modifications to the source code and make those modifications available to the public. This modification may come with restrictions.

There is, however, one element that is not shared by all licences, and that is copyleft. Copyleft is a legal mechanism contained in a licence which maintains the general freedoms awarded to FOSS users, but by acquiring a program released as copyleft, the user agrees to a licence that states that the software will not be used to develop closed source applications derived from it.[40] This is done by the inclusion of a clause that requires that all derivatives arising from the original code will be released under the same freedoms under which they were received. Arguably, the GNU General Public License (GPL)[41] is the most important copyleft licence.[42] One of the most essential

38. See: Guadamuz A, "Public Rights Licences", *WorldLII* (2008), http://www.worldlii.org/int/other/PubRL/.
39. Such as SourceForge, located at: http://sourceforge.net.
40. Rosen LE, *Open Source Licensing: Software Freedom and Intellectual Property Law*, Upper Saddle River, NJ: Prentice Hall PTR (2004), p.105.
41. http://www.gnu.org/copyleft/gpl.html.
42. At the time of writing, 62 percent of all projects charted by the open source project Black Duck are released under one of the different versions of the GPL. Of these, version 2 commands 47 percent, and the latest version (3.0) has 6.35 percent of the total share. See: http://www.blackducksoftware.com/oss/licenses#top20.

clauses included in the GPL is the copyleft clause, which sets restrictions against using the software in proprietary manners. The section reads:

> 2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions: […] b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties **under the terms of this License**.[Emphasis added][43]

What this means is that any software developed by using the source code of the copyleft program must ensure that the GPL is transferred to future users of the derivative software. This type of licence has been aptly named a "viral contract" as the contractual obligations contained are passed through a chain of distribution to other contracting parties.[44] The GPL therefore spreads in viral form, as the licensee must include the terms of the GPL in any subsequent derivative work they produce. Those subsequent licensees will be under the obligation to license their derivatives with the same obligations in place, and so forth.

This opens up a crucial legal question relevant to the present work. The presence of hundreds of FOSS licences creates sometimes operational problems for developers. Imagine that you are a developer who wants to include software under a FOSS licence to your project. From the start, you are presented with a choice of licence, assuming that you do not feel inclined to draft one from scratch. If this is a non-copyleft licence, then probably one of the main requirements is that you keep the source code open, and you are free to choose your own licence. But imagine that the code is released under a copyleft licence such as the GPL, then any modifications you do will have to be released under the same licence, or in some instances, a compatible one. This may seem straightforward, but with code being released under different licences you may end up

43. GPL, s.2(b).
44. Radin M, "Humans, Computers, and Binding Commitment", 75 *Indiana Law Journal* 1125 (2000).

with conflicting and incompatible code. This is the problem of licence incompatibility,[45] and it is of growing concern in FOSS licensing. Some possible solutions to this problem will be discussed later.

2.3 Open content

Open source licensing only applies to software, and it has undoubtedly been a very successful example of the peer-production sharing ethic.[46] A measure of the success of both the licensing model and the open source development philosophy is that the same ideas have been exported to other areas of intellectual creation, most significantly in the creative industries.

The most successful deployment of the FOSS "some rights reserved" licensing model to cultural works is to be found in Creative Commons (CC) licenses.[47] The Creative Commons project attempts to create so-called "intellectual property conservancies",[48] separating a block of human knowledge offered for the benefit of the public, but still protected by intellectual property.[49] This is analogous to nature conservation areas that exist for the wider social benefit, but have restrictions on certain uses. In the Creative Commons, the goal of intellectual property conservancies is achieved through the offering of a wide variety of licenses to protect creative works from misuse. This is done through the application of open source principles, where the work retains its copyright protection, but it is distributed freely[50] as long as the conditions contained in the license are met. An interesting part of the CC licensing environment is that users get to customise the rights given by picking from various licensing elements. Creators and authors need only to go to the CC website and select from different options offered in a

45. For more about this subject, see: Rosen supra note 40.
46. Weber S, *The Success of Open Source*, Cambridge, MA: Harvard University Press (2004).
47. See: http://creativecommons.org/.
48. Creative Commons, *Legal Concepts*, http://creativecommons.org/learn/legal/.
49. For more details about CC licences, see: Dusollier S, "The Master's Tools v the Master's House: Creative Commons V Copyright", 29:3 *Columbia Journal of Law & the Arts* 271 (2007); and Elkin-Koren N, "What Contracts Can't Do: The Limits of Private Ordering in Facilitating a Creative Commons", 74:2 *Fordham Law Review* 375 (2005).
50. In the Free Software sense, free here means free as in freedom, not free as in beer.

few drop-down menus; the system then chooses the license that fits the parameters entered. These licenses range from offering the work straight into the public domain, to more complex licenses with restrictions as to the commercial distribution of the work and the use of licenses in such distributions.[51]

Creative Commons licenses resemble their software counterparts in the fact that they maintain a minimum set of standards that are met by all of the licences. All CC licenses allow owners to retain copyright, but also allow users with the right to copy, distribute, display, digitally perform and make verbatim copies of the work into another format. This allows users to freely share, remix and redistribute content. For example, many of the images used in this book are released under a CC licence, and therefore can be included without having to ask permission from the owner.[52]

It is important to note that the baseline definition of CC licenses does not mention anything about modification or adaptation of a work; does not deal with copyleft-like clauses requiring the use of similar licenses to distribute the work; does not mention attribution; and does not deal with the distribution of copies for commercial purposes. Nevertheless, creators can choose a CC license that maintains all of the restrictions mentioned from all of the options offered. Authors then can choose from the following options to generate their license:[53]

- *Attribution:* The work is made available to the public with the baseline rights, but only if the author receives proper credit.[54]
- *Non-commercial:* The work can be copied, displayed and distributed by the public, but only if these actions are for non-commercial purposes.[55]

---

51. For more about Creative Commons see: Waelde C et al, *The Common Information Environment and Creative Commons,* Final Report to the Common Information Environment Members of a study on the applicability of Creative Commons Licences (2005).
52. See for example, Figure 2.4.
53. For more about the CC license elements, see: http://creativecommons.org/about/licenses.
54. Starting with Creative Commons version 2.5, the Attribution element is factually a baseline right and not an element that can be selected for.

- *No derivative works:* This license element grants baseline rights, but it does not allow derivative works to be created from the original.
- *Share-Alike:* This is based on copyleft principle. Derivative works can be created and distributed based on the original, but only if the same type of license is used, which generates a viral license.

It is possible to have licenses that combine several of these options.[56] The strongest (and most popular) CC license is the Attribution-NonCommercial-ShareAlike License,[57] which is the license that most resembles the strongest copyleft software ones (such as the GPL). All CC licenses are presented in three formats: the first is a short and easy to read "Commons Deed", which explains the terms and conditions of the license in a simple manner; the second format is the "Legal Code", which is the full license; the third is the "Digital Code", which is described in an HTML version of the license[58] that can be read by search engines and makes it easier to list the content in the Creative Commons directory.

Creative Commons presents a very positive step towards the wider distribution of creative works that is thoroughly compatible with the principle of peer-production. These licences are almost entirely tailored to respond to the needs of the user-generated content environment because they allow users to take control of their own works without having to consult a lawyer in order to draft a licence. It also allows the reuse of works, a goal that makes it particularly well-tailored for non-commercial uses widespread in the peer-produced Internet. Any time someone releases their work with a CC licence, it is an invitation to the world to share the work. Sharing information and the collaborative nature of content is precisely one of the most powerful characteristics of peer-

55. For more details about what constitutes a non-commercial work, see: Creative Commons, *Defining "Noncommercial": A Study of How the Online Population Understands "Noncommercial" Use*, (2009), http://wiki.creativecommons.org/Defining_Noncommercial.
56. However, the No Derivative and the Share-Alike elements are exclusive.
57. Version 2.5 can be found here: http://creativecommons.org/licenses/by-nc-sa/2.5/.
58. To be more specific, the code uses Resource Description Framework (RDF) metadata. For more about RDF, see: http://www.w3.org/RDF/.

production. Social engagement and distribution of content enriches the content environment. Piracy is not really a concern within the UGC universe.

The success of the Creative Commons licensing model cannot longer be in dispute. At the time of writing, Creative Commons calculated that there were at minimum 350 million CC licensed works available online. This figure is a minimum because it only offers a metric of works that can be searched on the Internet, and it is possible that there are many more.[59] Another measure of the licensing success is that Wikipedia has started sharing all of its content using CC licences;[60] which serves as further evidence of the peer-production credentials of the open content movement. Moreover, the image-sharing website Flickr had 13,290,440 individual images released under a CC licence.[61] But not only is Creative Commons used only for creative works, a considerable number of scientific and scholarly publications are now published using CC licenses.[62]

Open content is an excellent example of the relevance of peer-production; from photographs to blogs, from music to scientific works, the message is clear. Open content has become an important part of the Internet's content ecology, and the existence of easy-to-use licences that allow redistribution of content are a central part of their success.

3. COMPLEXITY IN OPEN LICENSING

3.1 Open self-organisation

One noteworthy aspect of the peer-production phenomenon implemented in open licensing environments is that it is clearly an exercise in self-organisation. It is not really

59. For a discussion on the difficulties of fining CC content, see: Bildstein B, "Finding and Quantifying Australia's Online Commons", 4:1 *SCRIPTed* 8 (2007).
60. Nicole C, "Wikipedia Now Uses Creative Commons", *Mashable Blog* (12 February, 2007), http://mashable.com/2007/12/02/wikipedia-creative-commons/.
61. http://www.flickr.com/creativecommons/.
62. For more about this, see: Guadamuz A, "Open Science: Open Source Licences for Scientific Research", 7:2 *North Carolina Journal of Law and Technology* 321 (2006).

necessary to repeat some of the concepts of self-organisation present in complex-adaptive systems, but one factor that can be refreshed is that these are dynamic networks where the constituents (be they individuals or organisations) respond to one another to form some coherent and stable environment. In true self-organising fashion, open licensed content (particularly open source software) operates in a manner that brings together actors from around the world in order to work on a project that brings about a desired result, be it a collected book, a remixed song, or a piece of software.

One of the most powerful metaphors that explain the self-organising nature of open content was expressed by programmer Eric Raymond in his influential work *The Cathedral and the Bazaar*.[63] There are two main ways to organise large numbers of people to produce a result, by hierarchical top-down project management, or by organic and chaotic means. To build something large and magnificent as a cathedral, one needs a funding entity, architects, masons, sculptors, painters, carpenters and a large number of workers. These types of projects can only come about by the shared effort of hundreds of individuals, but more importantly, they require some form of centralised organisational structure in order to bring about the desired result. Bazaars are entirely different organisational entities, they are non-hierarchical, people come together seemingly without pattern, and either by custom or local norms produce a stable yet chaotic environment. Despite lacking structure, somehow bazaars exist. Raymond commented that open source software development was akin to a bazaar, it should not work, but somehow it does, even by being open to "the point of promiscuity". While Raymond seems to be unaware of words like autopoiesis, emergence, and self-organisation, what he is describing is precisely a network that comes together and creates order from chaos. Open source is a complex adaptive system at work.

The end result is that open licensed content presents an excellent case study for self-organisation. Moreover, the networks of developers, creators, distributors and re-users

63. Raymond ES, *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, Sebastopol, CA: O'Reilly Media (2001).

also provide an excellent testing ground for the legal application of complexity theory, as these are environments where the copyright licence reigns supreme.

A cursory look at the area of FOSS development features some apparently impressive examples of self-organisation in action. Take, for example, the Linux operating system. Linux is a UNIX-based kernel[64] which began as a hacker project by Finnish programmer Linus Torvalds in 1990. He had been waiting for the developments of a UNIX-like kernel from free software developers, but when one was not forthcoming he developed his own, named it Linux and then placed it on the Internet for free, asking programmers to improve and build on it.[65] In true open source fashion, Torvalds managed to get lots of feedback from other programmers by making the source code for Linux available to everybody. By the end of 1991, a stable version of the Linux kernel was available to the public, and development has continued ever since.[66] This has led to different versions (known as distributions) of Linux being developed, giving this operating system a lot of stability and security, as the community was in charge of its support.

The remarkable fact from a self-organisation perspective is that while Linus Torvalds remained as the main maintainer of the Linux project through the years, the development has continued with the contribution of thousands of programmers around the world, and by 2001 the Linux kernel contained 2.4 million lines of code;[67] and it has continued to grow, with the latest version boasting 4.2 million lines of code.[68] Needless to say, such an amount of work would have been impossible if Torvalds had decided to work on his own. Most of the contribution to the project has been by hobbyists, people who give up

64. The kernel is the central component in an operating system, which manages system resources and allocates memory and processor usage.
65. Moody, supra note 30; pp.31–35.
66. Ibid.
67. Wheeler, supra note 19.
68. Wheeler DA, *Linux Kernel 2.6: It's Worth More!* (2004), http://www.dwheeler.com/essays/ linux-kernel-cost.html.

their free time to write code, test the program and debug its interoperability. This constitutes an impressive case of self-organisation in the truest sense of the term.[69]

Moreover, it is possible to calculate what the cost of any open source project would have been if it had been produced via commercial centralised methods, in other words, if it had been built like a cathedral. Because the source code for FOSS projects is available to everyone for examination, it is possible to calculate the number of single lines of code that each program has, and therefore it is also possible to calculate the number of programmer hours that have gone into producing the software. A report funded by the European Commission[70] looked at the five largest FOSS software projects, and calculated the cost that would have gone into producing the programs using the cathedral model. The results are:

Table 6.1 Production cost estimate for the five largest FOSS software products[71]

Software package	Lines of code	Months	Person-months	Cost (million EUR)
Openoffice.org	5,181,285	130	79,237	482
Linux Kernel-source-2.6.8	4,033,843	160	145,036	882
Firefox	2,437,724	87	25,339	154
GCC-3.4	2,422,056	113	54,048	329
Xfree86	2,316,842	90	27,860	169

69. Hars A and Ou S, "Working For Free? Motivations of Participating in Open Source Projects", *Proceedings of the 34th Annual Hawaii International Conference on System Sciences* (2001).
70. Ghosh RA, *Study on the Economic Impact of Open Source Software on Innovation and the Competitiveness of the Information and Communication Technologies (ICT) Sector in the EU*, European Commission Report ENTR/04/112 (2006).
71. Ibid, p.50.

These figures hint at the existence of a commercially viable self-organising model; somehow the bazaar works. But how exactly do these communities come together to produce millions of lines of code? This is where the study of complex adaptive systems comes into play. The FOSS developer communities are in the end networks of interacting agents that come together with an ultimate objective. These networks can be local, for example, a group of programmers that know each other and come together to produce software. However, for larger projects, the network spans across countries, which requires more complex organisational structures.

An interesting study into these networks has produced some intriguing results.[72] By looking at the social network structure of the agents in the network, and their constituent nodes and links, researchers found that open source software communities resemble other organisational types found in both natural and artificial networks. In particular, they studied the organisational structure of wasp colonies, and found a striking similarity with the social network structure of FOSS communities. Specifically, the wasp network relied heavily on successful nodes to maintain the colony together. Similarly, a degree and link analysis of email communication of 120 separate software teams produced a similar reliance on heavily centralised nodes.[73] Not only that, when plotting the number of links in both a small network of wasps and in small software communities into a logarithmic graph, both networks had similar communication distribution patterns.[74] The implications of the study are clear, and it is something that we can see all across the board when dealing with complex adaptive systems – self-organising networks rely on the hubs.

The apparent reliance of FOSS teams on hubs within the network may seem inconsistent with the bazaar model, but it makes sense. Distributed networks of programmers who may have never met may self-organise, but self-organisation does not

72. Valverde S et al, "Self-Managing Systems – Self-Organization Patterns in Wasp and Open Source Communities", 21:2 *IEEE Intelligent Systems* 5 (2006).
73. Ibid, p.39.
74. Ibid, p.38.

preclude some form of hierarchical structure, with the appointment of leaders and organisers that help the group to reach an objective. Crowston et al[75] conducted a survey of several high-profile FOSS projects, and found that while there is undeniably self-organisation at work in these networks, there is also a leadership structure that allows successful communities to produce software programs. They were particularly interested in teams where there was no assigned leadership; in other words, where leaders emerged from the community, also in self-organising fashion. The need for some leadership was required precisely because the group needed direction in ways to incorporate new members to the team, and also because there was need to develop norms and rules to maintain cohesion and efficiency.[76] Other research seems to corroborate this finding, as it seems clear that projects that are held together by an important hub in the network have more chance to succeed than others where no such role is present.[77]

All of these investigations hint at the existence of strong self-organising forces within FOSS development groups, but the presence of such emergent features in the network could also be found in other organisational models, such as Raymond's cathedral example. Another study[78] has tried to look precisely into this question, trying to determine if the self-organisation found in open source communities is similar to that found in proprietary software development groups. The first point to address is that FOSS self-organisation tackles specific areas of project management. An essential element of self-organisation of large distributed networks is the allocation of resources and the assignment of tasks. A group that has come together often without monetary goals is difficult to organise in the best of situations, so how these networks are able to allocate specific tasks within the group can be a structural nightmare, yet it happens in

75. Crowston K et al, "A Structurational Perspective on Leadership in Free/Libre Open Source Software Teams", *Proceedings of the First International Conference of Open Source Systems* (2005), http://oss2005.case.unibz.it/Papers/68.pdf.
76. Ibid.
77. Grewal R, Lilien GL and Mallapragada G, "Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems", 52:7 *Management Science* 1043 (2006).
78. Crowston K et al, "Self-Organization of Teams for Free/Libre Open Source Software Development", 49:6 *Information and Software Technology* 564 (2007).

thousands of FOSS projects. The study found that there were two main ways in which the communities self-organised: by self-assignment of tasks, and by some form of leadership assignment where programmer reputation served to allocate some level of hierarchical structure. This is an entirely different manner to that which proprietary software is produced, where assignment of tasks is the norm. The researchers found that there was a clear distinction in how proprietary and non-proprietary firms are organised, and therefore FOSS projects are indeed self-organising in manners that are dissimilar to how the cathedral model operates.

It seems clear that self-organisation exists in the free and open source software arena. However, does this translate into the open content field? This is a more difficult question to answer, particularly because the amount of research into open content does not match that found in software. The existence of large peer-produced open content projects, such as Wikipedia, hint at the existence of self-organising mechanisms here as well. However, Wikipedia is a difficult example to offer because while it is undeniably a peer-produced repository of knowledge, it has evolved over time. In the beginning, everyone could add and edit content to the online encyclopaedia, but this led to low quality of content in many entries, and also led to editing wars in controversial subjects. The community has organised itself in strict hierarchical manner, with heavy editors becoming de facto leaders of the community, and where an even smaller group of users with administrative powers can exercise heavy control.[79] While this still shows the predominance of emergent order, there seems to be some difference between this type of self-organisation and that found in software development. Duguid[80] has commented that this may be caused by the existence of strong quality controls in the software arena, as

79. Viegas FB et al, "Talk Before You Type: Coordination in Wikipedia", *Proceedings of the 40th Annual Hawaii International Conference on System Sciences* 78 (2007).
80. Duguid F, "Limits of self-organization: Peer production and 'laws of quality'", 11:10 *First Monday* (2006), http://is.gd/ePW1U.

opposed to the open content creation. Nonetheless, he still found strong self-organising features in several UGC sites, such as Gracenote, Project Gutenberg and Wikipedia.[81]

While it is true that many of the studies highlighted provide some descriptive insight into self-organisation in general, there are also some prescriptive lessons to be learned. Particularly, it has been one of the goals of this work that a theory self-organisation in Internet regulation is possible. The above examples into the emergent order of FOSS communities offer actual examples that some level of regulatory control and norm-setting will arise within online networks regardless of the existence of top-down guidance. Social networks with a common objective will produce some operational structure; this simple yet powerful finding is in itself a vital piece of the regulatory puzzle. Moreover, these networks still rely on hubs to coalesce and produce works.

3.2 Open scale-free networks

The self-organising nature of peer-production opens up another possible use of network theory in the analysis of the phenomenon. The evidence into the organisational structure of open communities may lead to the hypothesis that the ordering described could be caused by the presence of scale-free networks. This seems to be a common feature of complex adaptive systems, so whenever we see such order, it is possible that this might be caused by the presence of power law distributions within the networks. There are two main elements of peer-production where the presence of power laws could provide some insight into self-organisation. Firstly, there are the networks of developers and creators themselves. Secondly, there are the works created under peer-production models. It is important to keep this distinction in mind, as the presence of a power law in one may not be translated into a power law in the other. It would be interesting to see if self-organisation occurs only from the existence of a scale-free network of developers, even if the actual content created does not display power laws.

81. Ibid.

There is growing evidence that peer-production networks are indeed scale-free. Maillart et al[82] conducted a network analysis of the content in Linux distributions, namely the growth in size of the number of packages contained within the Debian Linux distribution, which were created under the FOSS development model. They measured the growth of the distribution over time, and found that Debian had been growing steadily following Zipf's Law. If you recall, Zipf's Law takes place where the frequency of an occurrence is inversely proportional to the one next in rank. When looking at the growth of Debian over time and plotting the number in a logarithmic scale, then the resulting graph is a straight line, where one version of Debian would be proportionally larger than the preceding version. The study found that this was probably caused because there was a power law in the links between the software packages that constituted the distribution. The reason for this is that not all of the software included in Debian was created at the same time, but was dependent on the regular flow of created objects. New software would be included only where there was an existing package that supported it, hence the steady growth. In other words, software grows steadily because new software relies on the existing one. While this may seem a completely intuitive result, the presence of power laws in the growth of Linux tells us that there is clearly a self-organising force at work.

Other researchers have found power laws in the developer network. Madey, Freeh and Tynan[83] collected data from almost 60,000 FOSS projects between 2001 and 2003, and conducted network analysis of the size of projects as a whole, the number of developers within each project, and also identified important actors within each network by trying to determine if they collaborated with other projects. The study found power law distributions in all areas, for example, the size of projects presented a power law, but also the clustering of developers within each network. However, when the study looked

82. Maillart T et al, "Empirical Tests of Zipf's Law Mechanism in Open Source Linux Distribution", 101:21 *Physical Review Letters* 218701 (2008).
83. Madey G, Freeh V and Tynan R, "Modeling the Free/Open Source Software Community: A Quantitative Investigation", in Koch S, *Free/Open Source Software Development*, Hershey, PA: Idea Group (2005), pp.203–218.

at clusters of developers, they found high clustering in some groups; while the majority of projects followed a power law, some projects were highly clustered around smaller number of developers. The researchers comment:

> The analysis provides support for the contention that the F/OSS community is a self-organizing system, and it also yielded an unexpected finding regarding the structure of the community. Several different types of analyses on the F/OSS data obey the power-law, which gives support to the hypothesis advanced by many qualitative researchers that the F/OSS community operates as a self-organizing system. When one examines the size of connected projects and developers, however, a second phenomenon emerges. It appears that there may be a dual nature to the structure of the F/OSS community, at least at this point in time. While the less well-connected clusters fit the power law, suggesting that part of the network is operating as a self-organizing system, there is a substantial percentage of the network (34.6% in March 2003) that is behaving differently, and that cluster does not fit the power-law pattern of the rest of the network data.[84]

This result may be caused by another self-organising characteristic of scale-free networks that the study seems to have ignored, and this is the small world phenomenon. It seems evident that some open source projects collapsed into tight clusters, which is also consistent with scale-free networks, as what may be happening is a "rich get richer" scenario. Another study[85] that looked at a similar dataset of developers came to this conclusion when it found that there was indeed some high clustering of some of the communities, and explained that this was caused by preferential attachment: nodes already attached to the network attract new nodes, but the more connected nodes attract more nodes than others with a lesser degree of centrality. Preferential attachment is precisely one of the reasons behind the "rich get richer" phenomenon.[86]

From the above, it seems like the self-organisation present in peer-production networks is not only the result of the scale-free network distribution of the networks, but may also be caused by the presence of preferential attachment. Success breeds success

---

84. Ibid, p.209.
85. Jin X et al, "A Topological Analysis of the Open Source Software Development Community", *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, (2005).
86. Watts DJ, *Six Degrees: The Science of a Connected Age*, London: Vintage (2004).

online, successful sites draw more links, and therefore the network organises itself around these hubs. This seems to be occurring as well in open content networks where power law distributions are present. A study into Wikipedia data produced hard evidence that the online encyclopaedia was organising itself around popular links.[87] Wilkinson and Huberman explain that while large sites such as Wikipedia are usually open to large-scale editing from almost anyone, the patterns suggest that some articles accrue most of the editing efforts:

> We have shown that although Wikipedia is a complex system in which of millions of diverse editors collaborate in an unscheduled and virtually uncontrolled fashion, editing follows a very simple overall pattern. This pattern implies that a small number of articles, corresponding to topics of high relevance or visibility, accrete a disproportionately large number of edits. And, while large collaborations have been shown to fail in many contexts, Wikipedia article quality continues to increase, on average, as the number of collaborators and the number of edits increases. Thus, topics of high interest or relevance are naturally brought to the forefront of visibility and quality.[88]

All of the above serves as further proof that a theory of self-organising Internet regulation is possible. It seems that not only is there self-organisation occurring in peer-production, but that this is caused by highly clustered popular nodes in the network. Wherever we find self-organisation, it is possible to postulate that we will also find networks that display power laws caused by preferential attachment of nodes. The prescriptive lesson is that it may be impossible to regulate these networks in traditional top-down approaches, but also, the presence of power laws may explain why it has been so difficult for regulators to stifle some practices such as illegal file-sharing. The emerging picture from both file-sharing and peer-production networks is that these are highly resilient, self-organising complex entities. The above evidence points towards the

87. Capocci A et al, "Preferential Attachment in the Growth of Social Networks: The Internet Encyclopedia Wikipedia", 74:3 *Physical Review E* 036116 (2006).
88. Wilkinson DA and Huberman BA, "Assessing the Value of Coooperation in Wikipedia", 12:4 *First Monday* (2007), http://ojphi.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/ viewArticle/1763.

existence of structural laws of the Internet that have to be understood if one hopes to regulate the online environment.

3.3 Open licences, social clusters and fitness

Open licences are an important constituent part of peer-production. While not all user-generated content is released under an open licence, there are very important sectors that are. By definition, FOSS has to be released under one such licence, and there is also content that is released under "some rights reserved" schemes. As it has been stated earlier, these licences allow users to republish, reuse, redistribute and modify the original work.

It is in the reuse aspect of the licences where network theory tools may provide some helping hand in trying to unravel the licensing maze of potentially incompatible terms and conditions. As it has been mentioned earlier, open source software is released under various different licences, some of which may have terms that are incompatible with one another. When we are dealing with only one project, this is often not a problem, as the developers may choose source code and software packages incompatible with their own licensing objectives. Now consider larger projects that collect code from various existing packages, and you may start to see the problem.

To illustrate this problem, German and Hassan[89] examined the licensing terms of 124 FOSS projects including some popular applications such as Apache web server, mysql, and GCC. These software packages have in common that they are not monolithic pieces of code; they are often complex software projects which include a large number of components. They found that the inspected software used more than 20 licenses, which could be roughly classed under 11 sub-groupings once different versions of the same licence were merged. Unsurprisingly, a large number of the inspected software was released under the GPL (45 in total), but the worrying result was the range of

89. German DM and Hassan AE, "License Integration Patterns: Addressing License Mismatches in Component-Based Development", *International Conference on Software Engineering* (2009).

incompatibilities found. While the study found ways to get around potential incompatibility problems by mixing and matching packages released with compatible licences, the challenge to developers is clear.

Research in the specific area of licensing is scarce, and most investigative efforts seem to be directed towards the analysis of the development communities as such.[90] Other social network tools have been deployed to analyse and visualise open source projects specifically from the codebase by looking at the interaction between developers. Take Gource, a powerful and dare we say beautiful visualisation tool that allows developers to look at all aspects of their project (Figure 6.1).

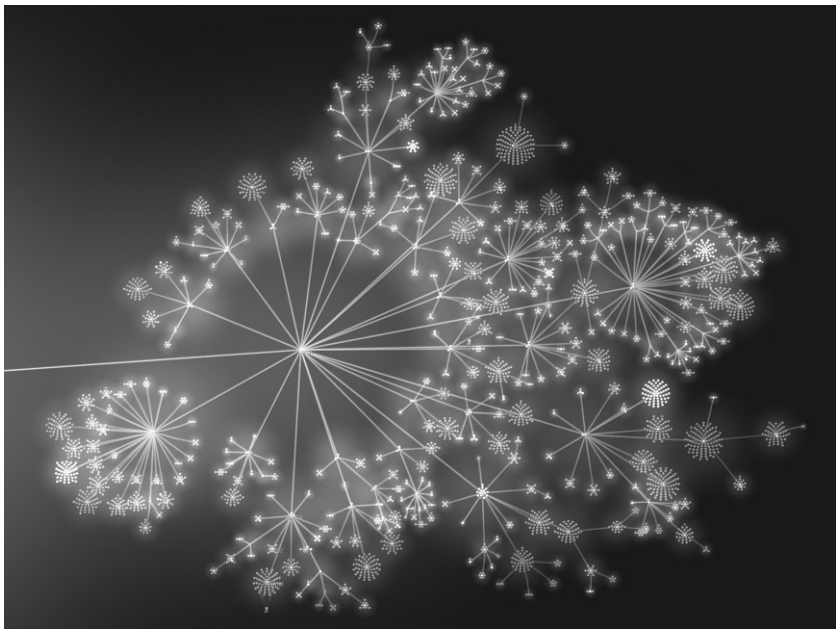


Figure 6.1 Social network representation of Linux[91]

90. See for example: Lanzara GF and Morner M, "The Knowledge Ecology of Open-Source Software Projects", *Proceedings of the 19th EGOS Colloquium*, Copenhagen (2003).
91. Found at: http://code.google.com/p/gource/wiki/Screenshots. Released under GNU General Public License v3, http://www.gnu.org/licenses/gpl.html.

The lack of research into source code distribution chains along licensing pathways may be the result of the various manners in which FOSS projects can be modified to produce new software. Source code can be simply adopted by another project, as has been explained, this would be a first generation adaptation from the original code, and it is known as branching.[92] It is also possible for a project to split into smaller components for development purposes, and then for the code to be brought together into the parent project, which is known as merging.[93] A third way of adapting source code is for a project to diverge into another version, something that is known as forking.[94] All of these derivative strategies hint at the prevalence of the functional nature of software. FOSS licensing is an efficient way to make sure that the source code remains open for modification. This functional nature may dissuade researchers from looking at the distribution chains, as what matters is the decision-making process that produced the modification. Cultural works are different because creative expressions are intrinsically subjective and non-functional; each modification is dependent heavily on the aesthetic and cultural value of the original.

Interestingly, this difference between the functional aspect of software and the value of cultural remixes has allowed more research into the open content licensing modification chains. ccMixter[95] is a relatively small online community of musicians and music enthusiasts where creators make their music available to the public under a CC licence, encouraging the reuse and remixing of these works. Cheliotis[96] conducted a network analysis of the reuse of materials within a community of 1,850 active users sharing 7,484 music works. He drew a directed graph of works, where each work was the node, and if there had been reuse of a work, then that would be established as a link.

---

92. Cheliotis G, "From Open Source to Open Content: Organization, Licensing and Decision Processes in Open Cultural Production", 47:3 *Decision Support Systems* 229 (2009).
93. Ibid.
94. Ibid.
95. http://ccmixter.org/.
96. Cheliotis G, *Remix Culture: Creative Reuse and the Licensing of Digital Media in Online Communities*, Participatory Media Lab Working Paper (2008), http://pml.wdfiles.com/local--files/working-papers/Remix_Culture_Web_Version.pdf.

By looking at a smaller dataset, the study was able to look at several aspects of how peer-production occurs in real life but, more importantly, it was possible to examine the regulatory role of the licences in the remixing community. Cheliotis comments that:

> Interestingly we have thus far observed that in ccMixter (a) authors of derivatives tend to respect the licenses of the works they reuse, and (b) in cases where they could legitimately license their derivatives under more restrictive terms, they generally do not. Upon closer inspection we found out that this is primarily the result of an ingenious licensing mechanism implemented by the site administrators. Every author of a remix must state the sources used in the derivative work. As the license of each source work is stored in a database, the website will automatically select an appropriate license for the remix. Thus license compliance is ensured.[97]

This is a valuable regulatory insight into the peer-production licensing ecology. It seems like architecture plays an important part in the organising decisions made by content creators. Similarly, community norm-setting is still as important as the licence itself, as the community will be more likely to establish licensing decisions that should be followed by those participating in the sharing environment. Another finding of the study is that it analysed the chain of reuse of participating musical works; in other words, it could follow the number of remixes that the original piece received. This provides a concise demonstration of distribution chains that are also present in other peer-production areas, such as software. Here Cheliotis found that most works were only remixed one, and that subsequent modified versions were less likely to occur, down to the fifth generation, where remixing was almost negligible. Finally, the study produced results that are consistent with the rest of the literature on peer-production networks, as it became clear that the network was organised around central hubs and popular mixes (Figure 6.2).

97. Ibid, p.8.

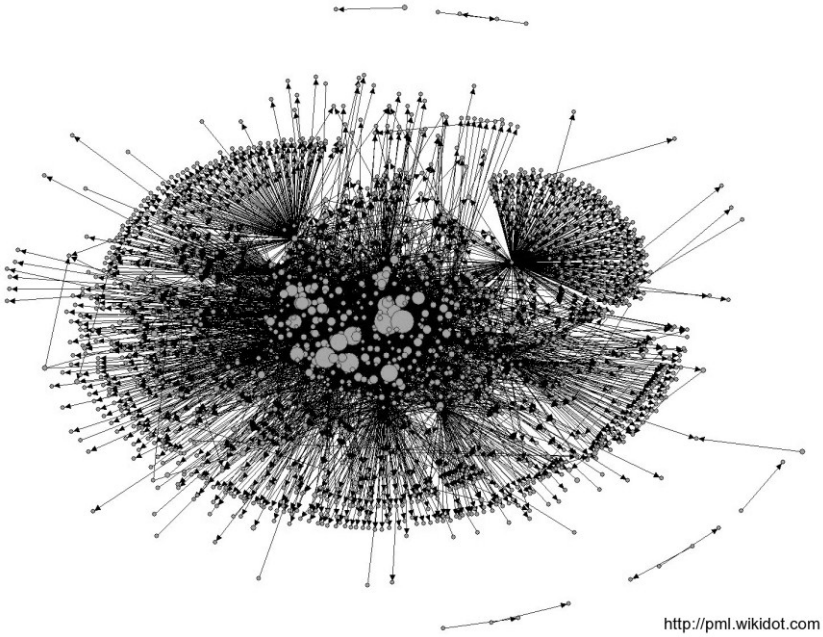


Figure 6.2 Network of works produced in the ccMixter community

Finally, the viral nature of peer-produced works released under copyleft licences could also be seen as well from a complexity perspective, and it could give us valid regulatory insights as well. One of the effects of copyleft licensing is that copyleft clauses require works to be released under the same terms licence. As has been mentioned before, this could prove to be a viral effect: the licence would replicate because it is explicitly required to do so. This serves as a very strong regulatory imperative to the community of creators and developers that are currently sharing and using the licensed work.

The question is whether copyleft licences are unintentionally creating a licensing environment that intrinsically favours copyleft content. Thinking back on some of the concepts of self-organisation, complex systems can sometimes reach order within a fitness landscape where one of the fitness states is more stable than the other options. Could it be possible that copyleft licences are a fitness peak? The popularity of the licences is undeniable. As it has been mentioned already, copyleft licences are dominant

in the FOSS ecology. Similarly, the ShareAlike licence element in Creative Commons licences is one of the most prevalent elements. More importantly, the share of CC-licensed works under a copyleft licence has been growing. In 2006, only 45 percent of CC-licensed works were copyleft,[98] while by May 2010, 57 percent of all content was copyleft.[99]

Here we have a powerful indication of the existence of fitness landscapes in the self-organisational nature of the Internet. Theories of emergence tell us that order can be achieved when a system tries to find the most effective solutions to a problem. The simple addition of a clause in a copyright licence may serve as the point in which regulatory coalescence occurs, acting as a lightning rod for preferential attachment to occur. Something as seemingly innocent as a paragraph in a licence is an important self-organising force. Self-organisation can be directed, which has important implications for the proposed nascent theory of Internet regulation.

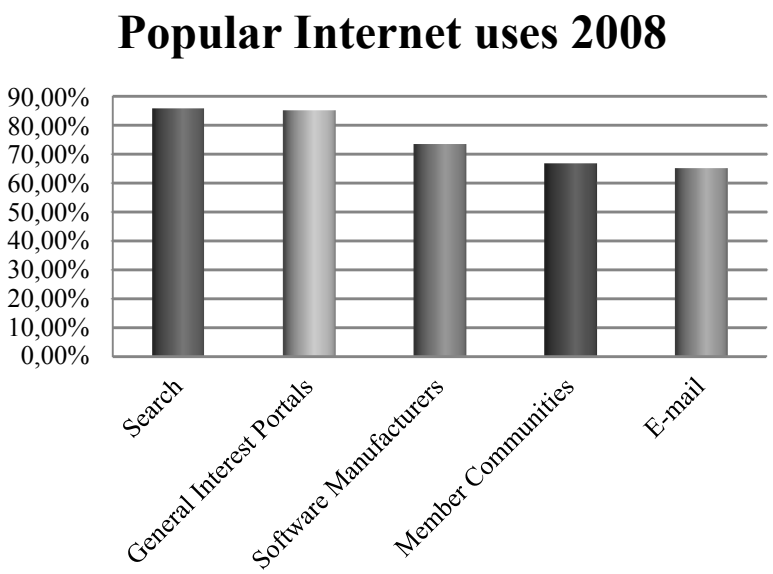
4. PARETO REVISITED

In the last chapter we examined the dominance of Pareto distributions in commercial creative works, but also the emergence of new business models exemplified by the long tail. Given the fact that there is currently a clash of cultures between peer-production and commercial content, it is logical to ask whether peer-production displays different consumption patterns to those found in their "all rights reserved" counterparts. Most, if not all, peer-produced content is offered free of charge online, so one would expect that it displays very different usage characteristics to those present in the established creative industries. Is this the case?

98. Creative Commons, *License statistics*, (2006), http://wiki.creativecommons.org/License_statistics#Raw_search_engine_query_data.
99. Creative Commons, *ccMonitor World Statistics*, (2010), http://monitor. creativecommons.org/World.

4.1 Measuring the UGC ecology

A mere look at the size of the UGC universe produces an impressive picture of the numbers involved. Wikipedia contains more than 12 million articles in 262 languages, of which 25 have more than 100,000 entries.[100] In November 2008 Flickr reached the 3 billion picture mark.[101] By March 2008, YouTube had 78 million videos on display, with 200,000 new videos uploaded every day.[102] Exact numbers on blogging are difficult to come by; blog aggregator Technorati has indexed a total of 133 million blogs since 2002, 900,000 of which had postings within 24 hours.[103] According to Nielsen, social communities (blogs and social networking sites) are now the fourth top online activity of internet users, beating email (Figure 6.3).[104] By February 2009, Twitter had received 7,038,000 visitors, a 1,382 percent change from previous year.[105]



100. Wikipedia, *Wikipedia*, http://en.wikipedia.org/wiki/Wikipedia.
101. Flickr, *3 Billion*, http://blog.flickr.net/en/2008/11/03/3-billion/.
102. Digital Ethnography, *YouTube Statistics*, (2008), http://ksudigg.wetpaint.com/page/YouTube+Statistics?t=anon.
103. Technorati, *State of the Blogosphere 2008*, http://technorati.com/blogging/state-of-the-blogosphere/.
104. Nielsen, *Global Faces and Networked Places*, Nielsen report (March 2009), http://is.gd/eTSFc.
105. Nielsen, "Twitter's Tweet Smell Of Success", *Nielsen Wire* (18 March, 2009), http://blog.nielsen.com/nielsenwire/online_mobile/twitters-tweet-smell-of-success.

Figure 6.3 Internet use according to Nielsen

Whichever way one looks at these figures about the biggest players in the Web 2.0 sphere, one cannot help but be impressed. However, as has been remarked before, there are commentators that see such numbers as little more than widespread organised copyright infringement. Are we witnessing a true upswell in creativity, or simply rehashed postings and mash-ups from "real" creators? This is not a baseless question; one of the most publicised copyright infringement case hinged precisely on this issue. In *Rowling v. RDR Books*,[106] JK Rowling and Warner Bros sued Steven Vander Ark, the author of the Harry Potter Lexicon, and his publishers alleging copyright infringement over the publication of the Lexicon in printed form. The Lexicon was a reference work which contained numerous entries detailing the Harry Potter world. The case highlighted the existing clash between competing media, as the Lexicon was very much the embodiment of user-generated content, where a fan of the Potter books had taken considerable time and effort to document and reference the tomes for an online audience. However, upon closer inspection, the court found that while the Lexicon conveyed "new information, new aesthetics, new insights and understandings" to the original work, it had copied and pasted entire passages unattributed, in what amounted to little more than plagiarism.[107]

This case serves to illustrate a serious problem encountered by user-generated content in particular and peer-production in general, and it is the perception amongst some sectors that most of the content is either pirated or copied from established sources. It is difficult to dispel this myth given examples such as the Harry Potter Lexicon, but surely there is a wealth of true creativity and inventiveness involved in peer production. How to measure this then?

106. *Warner Bros. Entertainment Inc. et al v. RDR Books et al* (575 F.Supp.2d 513).
107. Ibid.

Unlike commercial distribution of content, impact and success are not measured in sales, so the first task is to find an adequate way to assess its relevance. Mere hits could give us an indication of popularity, but website metrics could be clouded with a question of granularity of content. A more indicative measure could be done through link backs to content, as this could give a better idea about the way in which information is being shared online. Technorati does precisely this with blogs by measuring impact through peer linking within the blogosphere, what they call a blog's authority. This type of measure provides a better indication of what users find relevant as they link back to that content. Interestingly, a look at Technorati's top 100 blogs reinforces the idea that we are faced with some scale-free topologies in UGC content, just as was found in the FOSS environment. For example, the top authority blog is The Huffington Post with more than 25,000 links to it, while there is a sharp fall in authority, producing a long tail chart (Figure 6.4).

The inequality is even sharper when one looks further down the chart. As of the time of writing, an authority of 375 ranks at around 7,000; 200 authority ranks 17,000; 95 authority ranks 48,000; and 41 authority ranks 133,000. The blogosphere is a long tail of millions of blogs, but those at the top have a disproportionate amount of followers when compared to less read ones.

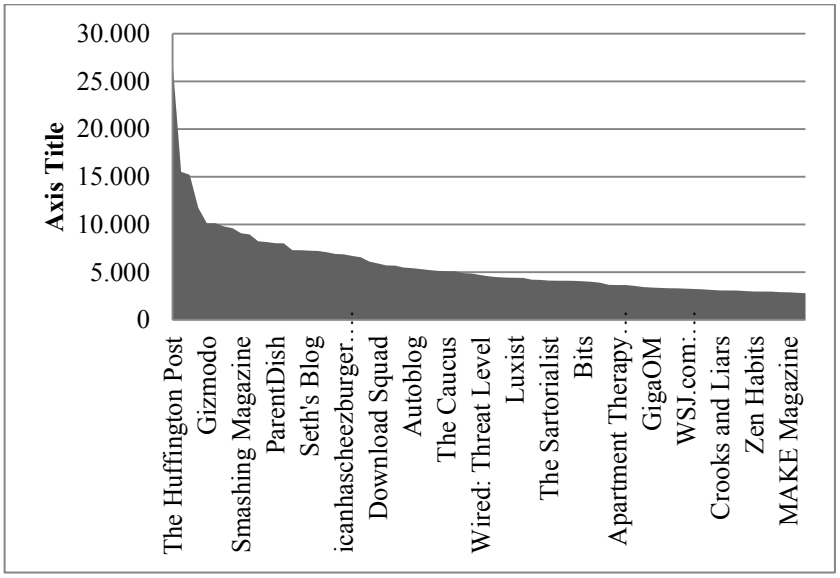


Figure 6.4 Technorati's top 75 blogs[108]

This phenomenon is replicated in other user-generated sites with uncanny regularity. Incoming links to Wikipedia articles provide yet another long tail,[109] as does the measure of article views.[110] A study of 1.8 million random pictures on Flickr during a 10-day period produced an almost perfect power law, where 7 percent of images accounted for almost 50 percent of the views.[111] Another study on YouTube video popularity also produced a heavy skew towards the top viewed content, while it also displayed a heavy tail.[112]

108. Chart as of 11/04/2009.
109. Wikipedia, *Histogram of Incoming Links for English Wikipedia Articles*, (2009), http://commons.wikimedia.org/wiki/File:Histogram_of_incoming_links_for_English_Wikipedia_articles,_January_2009.jpg.
110. See: Wikirank: http://wikirank.com/en.
111. Van Zwol R, "Flickr: Who is Looking?" *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence* (2007), http://www.semedia.org/PubFolder/ vanZwol-FlickrWhoLooking.pdf.
112. Cheng X, Dale C and Liu J, "Statistics and Social Network of YouTube Videos", *16th International Workshop on Quality of Service* 229 (2008), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4539688.

It seems clear that there is considerable content inequality in the top UGC websites. Is this replicated elsewhere? This is more difficult to find out, as the research into UGC impact tends to concentrate on the bigger websites, probably because of the availability and openness of datasets and the ease of crawling through content. Studies into the usage of open source software have produced similar results, where it is clear that software popularity also follows Pareto distributions.[113] Research into the use of content released under Creative Commons licenses seems to indicate similar long tail behaviour.[114] Although there are inherent limitations with producing accurate data in consumption of CC content,[115] one cannot ignore that one result seems to keep coming up time and time again. Pareto is alive and well in peer-production.

These findings seem completely counter-intuitive when one considers the many differences between peer-production and the copyright industries, a difference that is not only about price; the entire licensing model is different, as well as the motivation of players. In fact, whether a work is offered for free online or paid seems irrelevant when one looks at sheer download figures, as evidenced by the fact that even P2P content follows the commercial popularity of the work: more popular films will be more downloaded for free.[116] Therefore, one would have to ask the reason for the similarities in usage statistics displayed by both commercial content and user-generated content. It is possible that in both environments, information replicates in similar manners. Both UGC and commercial content appear to respond similarly to the self-organising laws that operate in cyberspace.

As seen in the previous chapter, commercial content is remarkably susceptible to information: the more buzz there is for a work, the more likely it is that it will display

---

113. Hunt F and Johnson P, "On the Pareto distribution of Sourceforge projects", *Proceedings of the Open Source Software Development Workshop* (2002).
114. See: Cheliotis et al, "Taking Stock of the Creative Commons Experiment Monitoring the Use of Creative Commons Licenses and Evaluating Its Implications for the Future of Creative Commons and for Copyright Law", *35th Research Conference on Communication, Information and Internet Policy* (2007), http://web.si.umich.edu/tprc/papers/2007/805/ CreateCommExp.pdf.
115. Bildstein, supra note 59.
116. Pouwelse JA et al, "Pirates and Samaritans: A Decade of Measurements on Peer Production and Their Implications for Net Neutrality and Copyright", 32:11 *Telecommunications Policy* 701 (2008).

successful sales figures. This feature seems to be shared by successful user-generated content, where hits are often referred to as "going viral". Be it a blog post, a video on YouTube, or a picture on Flickr, there are some instances when the content accumulates incoming links causing a tipping point, and the work is replicated throughout the internet. What makes content reach this point is still a mystery, for example, there seems to be little aesthetic logic or uniting theme in top video content on YouTube.[117] As anyone who follows popular culture closely, the same seems to apply to commercial content. However, one feature is shared in both RO and RW cultures, and this is the fact that preferential attachment and the "rich get richer" phenomenon are strong features of both production philosophies.

Nonetheless, UGC is much richer than the few instances where a work has gone viral, yet if it can be described using the same distribution curves that apply to traditional creation models, then perhaps both systems are not as different as one would expect. While it is clear that UGC represents a departure from commercial methods of exploitation, the apparent similarity in distribution could mean that both types of works could be covered under the same set of copyright policies, much as it is done now.

117. Burgess J, "'All your chocolate rain are belong to us?' Viral Video, YouTube and the dynamics of participatory culture", in Geert Lovink and Sabine Niederer (eds), *Video Vortex Reader: Responses to YouTube,* Institute of Network Cultures: Amsterdam, (2008), pp. 101–109.

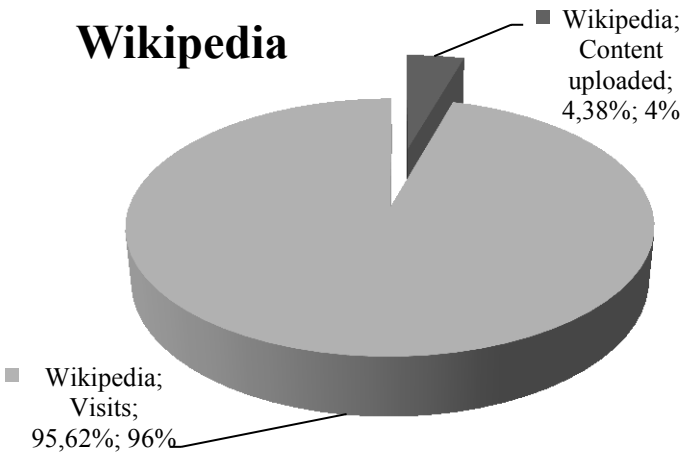


Figure 6.5 Visits versus user participation in Wikipedia

There could be a deep structural similarity between the RO and RW cultures, and it is that both production models share one striking characteristic, and this is the inequality in the numbers of creators versus consumers. Under Pareto's Law, the content industries operate with small number of high earners at the head of the charts, followed by the long tail of lesser acts. More importantly, this inequality can also be found in the number of creators versus the number of users. While peer-production offers us a potential democratisation of the creative process, it is important to note that with few exceptions, the number of users that create content is still a minority when compared to the number of consumers of that content. Research firm Hitwise conducted a study into the amount of visitors top UGC websites received against the participation from users, and it found that there was considerable disparity in this regard. YouTube and Flickr had user participation of just 0.18 and 0.12 percent respectively, while Wikipedia had 4.38 percent of user participation against visits (Figure 6.5).[118]

118. Tancer B, *Measuring Web 2.0 Consumer Participation*, Hitwise US Research Note (2007), http://hitwise.com/downloads/reports/Hitwise_US_Measuring_Web_2.0_Consumer_Participation_June_2007.pdf.

The reason for this is that not everyone is interested in the process of creation. Having the technical capacity to create a blog does not immediately turn one into a writer. Similarly, there is no reason to believe that just because one can upload content online, one should do so. Depressing as it may sound for the proponents of peer-production (of which the writer is one), the similarities between UGC and traditional methods still remain. This is a subject where the deterministic nature of scale-free networks seems to dictate the terms of usage.

4.2 Copyright policy implications

The relevance of these statistics for copyright policy is considerable. The popularisation of user-generated content, the growing number of Web 2.0 applications, and the widespread replication of content online (both legal and infringing), pose several challenges for copyright law and policy. So far, policymakers have been single-mindedly intent in curbing copyright infringement and boosting the enforcement of intellectual property. These are solutions to problems that affect a small number of earners. Such interest in enforcement is understandable as the copyright industries generate large amount of income for the economy. However, policy should also consider the potential relevance of peer-production content as a source of wealth, research, creativity and innovation.

One corollary from the evidence presented so far is that policymakers cannot ignore peer-production. New business models are emerging and, like it or not, Web 2.0 tools and UGC have become an important part of the content ecology. One need only look at a comparison between blogs and the mainstream media to realise that peer-production models are often reaching as many users as their commercial counterparts.[119]

It also seems clear that the Internet favours the long tail model for both UGC and mainstream content. Particularly when talking about Internet content, the rising

119. Sifry D, "State of the Blogosphere August 2005 Part 5: The A-List and the Long Tail", *Sifry's Alerts Blo*g (10 August, 2005), http://www.sifry.com/alerts/archives/2005_08.html.

popularity of content-creation by users may be occurring because Web 2.0 tools encourage the publication of creative works, and their ulterior dissemination to the entire Web. Policymakers should take this into consideration when looking at ways to regulate copyright in the digital domain. Evidence points towards the fact that an important number of copyright owners are located in the peer-production sector. It would be useful if policy was no longer designed with the idea of profit in mind. More often than not, the creator will be a hobbyist, never expecting a monetary return for her troubles. Nevertheless, it is easy to make this point while forgetting that while peer production is on the increase, readership of such content may not be. Just because something is online does not mean that it has an audience. Similarly, "traditional" offline world ideas of quality and peer-review still apply to the online environment.

Nonetheless, the pervasiveness of Pareto inequalities in both RO and RW cultures is only important if we see copyright as a mere economic right. If that is the case, then peer-production will be only relevant if it creates popular super-hits that could be transformed into commercial works. The continuing emphasis on the economic value of creative works would serve to dispel the utopian ideal of the Internet as a democracy of information where data is the currency.[120] There would be an inherent inequality in the nature of information, everybody is free to participate in the online environment, but only the works at the head of the long tail will be of importance.

However, a fact that is often forgotten is that copyright is not only about economics. A copyright policy based on Pareto distribution models ignores completely the fact people are still willing to create without hope of remuneration. There is indeed an economic incentive provided by copyright, but this incentive is nebulous; many Internet users are happy to produce works subject to copyright protection for all sorts of reasons. There is something to be said about those who willingly inhabit the long tail.

---

120. Some of these ideas can be seen expressed here: Lessig L, *The Future of Ideas: The Fate of the Commons in a Connected World,* New York: Random House (2001), pp.240–244.

So, while the data shows that the clash of cultures is not as deep as we may otherwise think, there cannot be any doubt that we are witnessing a monumental change in the manner in which information is produced and disseminated. If this is the case, and to be honest there is little reason to believe otherwise, then copyright has to respond. It is true that usage figures in peer-production are still governed by the same rules that apply to commercial works, but this does not change the fact that those who were only consumers of content are now also producing it. Copyright policy must recognise this in some form or another, perhaps by the inclusion of peer-production friendly rights.

There does not need to be a monumental shift in copyright policy, but perhaps a reminder that copyright is not only about economic rights could be a start. One of the most important licence elements in Creative Commons licences is the Attribution. This has been a moral right present in international copyright law for more than a hundred years.[121] Licence and copyright inevitably meet. So, peer-production friendly policies could include more exceptions for non-commercial uses, just like those present in CC licences, but also a renewal of the importance of moral rights.

There is a final point about the relevance of user-generated content to copyright policy. As more and more consumers become creators, they are also becoming copyright owners. Copyright used to be something that happened to wealthy and famous people, but now it increasingly happens to many of us. If you upload a picture on Flickr, post a video on YouTube, create a blog, program source code, release a song, or even post in a social media site, then you are also a copyright owner. This very fact should be advertised everywhere. Copyright has to be respected not because it might affect a millionaire author in a country far away, but it might also affect you. This is where both cultures meet, in the blurring between creator and consumer.

121. *Berne Convention for the Protection of Literary and Artistic Works* (1886, as amended), Art. 6bis.

7. Cybercrime and Networks

> ***John McClane:*** Hey, what's a fire sale?
> ***Matt Farrell:*** It's a three-step systematic attack on the entire national infrastructure. Okay, step one: take out all the transportation. Step two: the financial base and telecoms. Step three: You get rid of all the utilities. Gas, water, electric, nuclear. Pretty much anything that's run by computers which... which today is almost everything. So that's why they call it a fire sale, because everything must go.
> *Die Hard 4.0*

In the early hours of April 9 2009, an unprecedented attack took place against the cyber-infrastructure of a relatively small American town. Ten fibre-optic cables were cut by unidentified intruders in four separate locations in the town of Morgan Hill, near San Jose California.[1] This seemingly small incident had tremendous implications in the surrounding area, affecting telecommunication services for Morgan Hill and three other counties, completely knocking down "911 service, cellular mobile telephone communications, land-line telephone, DSL internet and private networks, central station fire and burglar alarms, ATMs, credit card terminals, and monitoring of critical utilities".[2] The disruption to the digital telecommunications network was such that it stretched emergency services to such an extent that local hospitals had to revert to analogue technologies such as ham radio, as the affected area was completely cut-off from the rest of the country.

As of today, nobody has been charged with the attack, but what remains clear is that the perpetrators uncovered a serious vulnerability in digital telecommunications

1. Asimov N, Kim R and Fagan K, "Sabotage attacks knock out phone service", *San Francisco Chronicle* (10 April, 2009), http://bit.ly/q5iOd.
2. Perens B, "A Cyber-Attack on an American City" *Silicon Alley Insider* (25 April, 2009), http://www.businessinsider.com/a-cyber-attack-on-an-american-city-2009-4.

infrastructure, its centralisation, a pattern that may remind readers of the vulnerability of centralised P2P systems discussed in Chapter 5. The Internet was precisely created to avoid this kind of targeted attack. In theory, damage to part of the network should be redirected and most of the communication systems should have been distributed across other hubs. The problem in Morgan Hill was something that is increasingly common in the Internet's architecture, and that is the fact that while the logical infrastructure is distributed, the actual physical cabling serving large geographical areas rely on too few chokepoints. Taking down only four such hubs blacked out a disproportionately large area considering the small nature of the attack.

While this incident was isolated, it can be used as an illustration of the growing problem presented by recent trends in network architecture. Instead of having a resilient decentralised telecommunications network, the actual implementation of the Internet has proved to be subject to cascading failures. The modern Internet relies on a centralised system that is increasingly vulnerable to coordinated attacks such as the one in Morgan Hill.

The present chapter will look at the issue of network centrality from the perspective of complexity theory, looking particularly at the implications of such centrality for cybercrime, both from a preventive perspective and also from an enforcement angle. The growing centrality of the Internet's architecture can be seen as a problem if we want a robust network capable of withstanding cyber-attacks. However, cyber-criminals also operate as networks, and as such it might be possible to design enforcement strategies that rely on network theory in order to uncover patterns, helping to design more resilient networks, but also may help us in dismantling criminal gangs thanks to small-world analysis.

1. CYBERCRIME

One of the most important aspects of Internet regulation is that of online criminality, a phenomenon often described in the literature as cybercrime. While the Internet offers

untold positive opportunities, there can be little doubt that it is also a breeding ground for illicit activity. This is only to be expected, since every big step in human invention provides opportunities for both use and misuse, and cyberspace is no different in that regard.

One of the challenges of those interested in studying the topic of cybercrime is how to define and delimitate the subject. For example, the invention of the telephone created new opportunities for criminal activity, so should there be such a thing as "telephone crime"? The answer is obviously no, the medium is not vital to the commission of the crime, it may simply facilitate it. So, why should we have a separate area of legal scholarship that studies crimes committed through a computer? It would consequently be essential to try to demarcate the subject of study. Nowadays, most human activities have an online element, and an increasing number of "offline" criminal actions also have an online element. For example, a fraudster that uses email to contact a target should not really be considered cybercrime in the strict sense, just as whether the person telephoned the target would not be important to the final result.

The existing definitions of cybercrime are not useful in this regard. The Computer Crime Research Center defines it as "crimes committed on the internet using the computer as either a tool or a targeted victim".[3] This is of course too broad and perhaps not very useful definition because it would cover practically anything where a networked computer has been used in one way or another. The Council of Europe's Convention on Cybercrime seems to take the narrower approach when it defines cybercrime as any "action directed against the confidentiality, integrity and availability of computer systems, networks and computer data as well as the misuse of such systems, networks and data by providing for the criminalisation of such conduct".[4] This narrows the field, as it adds the element that the criminal action must be against the network or computer itself.

---

3. Joseph A, *Cybercrime Definition*, Computer Crime Research Center Papers, http://www.crime-research.org/articles/joseph06/.
4. 2001 Council of Europe Convention on Cybercrime, Preamble, para 9.

A more useful approach is to think of cybercrime as those illicit activities that would not otherwise exist without the Internet. This filters out everyday crimes that may have an online element at some stage of commission. This narrow definition, however, would filter out some criminal types that have been already enacted into legislation. The aforementioned Convention on Cybercrime includes some offline offences that have an online element. The convention recommends the inclusion of the following categories of cybercrime:

1. *Offences against the confidentiality, integrity and availability of computer data and systems*. These include illegal access, illegal interception, data interference and misuse of devices. The types of crimes covered here would be hacking offences where groups or individuals access a system without authorisation in order to attack the system or remove information.
2. *Computer-related offences.* These include computer-related forgery and computer-related fraud, and the objective is mostly to alter or delete information with fraudulent means for economic gain.
3. *Content-related offences.* These deal mostly with child pornography.
4. *Offences related to infringements of copyright and related rights.* These are self-explanatory.
5. *Ancillary liability and sanctions.* These include corporate liability, and aiding and abetting in the commission of one of the aforementioned offences.

National laws tend to treat the subject of cybercrime by a combination of the application of old norms and the enactment of new legislation. For example, in the UK, the main cybercrime law is the Computer Misuse Act 1990,[5] which contains only three wide-ranging offences: unauthorised access to computer material; unauthorised access with intent to commit or facilitate commission of further offences; and unauthorised

5. Computer Misuse Act 1990 (c.18).

modification of computer material. Other legislation has been changed specifically to cover new variants of old crimes, such as the Fraud Act 2006.[6]

However, while the law in this subject remains static, the complexity of criminal acts committed online continues to grow. One could see the offences described as traditional cybercrime. There are various new variations of criminal offences, what Edwards calls the next generation of cybercrime.[7] These new offences include variations of existing themes, but that because of their technical complexity and international nature require new means of legal classification. These include:

1. *Phishing*: This is a combination of hacking, fraud and account hijacking. A phishing attack usually takes the shape of an email that seemingly comes from a financial institution, payment system, or e-commerce site, which asks the user to connect to a website and enter login and password. The site is a fake portal that takes these details and uses them to enter the user's account in the real service, and then removes funds, goods, or uses the facility to make purchases.[8]
2. *Botnets:* These are networks of hijacked computers that have been infected by viruses or trojan programs. The program may lie dormant in the system for later use, or it may be used for sending out unsolicited email or phishing attacks.[9]
3. *Denial of Service (DoS)*: This is related to botnets. Infected computers that form part of a botnet may be used to send information requests to a specific website; the idea is that the target server will be overwhelmed and it will eventually be brought down.[10] Botnets can be very effective, and have been used successfully by hackers to affect even large service providers such as Yahoo and eBay.[11]
4. *Cyber-terrorism:* This is not so much an offence as such, but can be defined as the combined use of other existing forms of cybercrime to commit terrorist attacks. Cyber-terrorism therefore could use denial of service attacks against important infrastructure in order to disrupt services and create chaos, but in the wider sense

6. Fraud Act 2006 (c.35).
7. Edwards L, *Cybercrime 2009: The Legal Perspective*, RUSI CyberSecurity Conference, London (October 2009).
8. Drake C, Oliver J and Koontz E, "Anatomy of a Phishing Email", *Conference on Email and Anti-Spam* (2004), http://bit.ly/coZCQa.
9. Maurushat A, "Zombie Botnets", 7:2 *SCRIPTed* 370 (2010).
10. Edwards L, "Dawn of the Death of Distributed Denial of Service: How to Kill Zombies", 24:1 *Cardozo Arts & Entertainment Law Journal* 23 (2006).
11. Brown I, Edwards L and Marsden C, "Information Security and Cybercrime", in Edwards L and Waelde C, *Law and the Internet*, 3rd ed, Oxford ; Portland, OR: Hart (2009).

it could also be used to describe websites that contain bomb-making instructions, or forums used to incite violence.[12]

5. *Cyber-warfare:* This may seem not to be so much a new type of cybercrime, as the subject of international public law. However, modern cyber-warfare could be defined as a type of cybercrime because it uses some traditional and new offences in order to conduct some international policy strategy. Cyber-warfare could be a systematic attack via botnets against a country's infrastructure, but it can also take the shape of a hacking attack against a corporate entity in order to obtain sensitive data. This would include forms of cyber-espionage.[13]

Regardless of the categorisation used to define cybercrime, there can be little doubt that it is a phenomenon that has been growing in importance in recent years, as more of our daily lives contain some online element. For example, phishing is of increasing concern to the financial services. The Anti-Phishing Working Group is a global industry and law enforcement association dedicated to removing phishing websites and conducting research into phishing attacks. They report that phishing peaked in August 2009 with a record 40,621 unique phishing reports, and a staggering 56,362 unique phishing websites.[14] Of these sites, 39 percent targeted the financial sector, 33 percent payment systems (such as credit cards and online account-based systems like PayPal) and 13 percent targeted online auction sites.[15] While the actual damage is hard to calculate due to inaccuracy in reporting, card-not-present online fraud in total accounted for £134 million GBP in the first half of 2009.[16]

Brown, Edwards and Marsden[17] have compiled a number of cybercrime statistics that give good idea of just how big the problem is. Amongst these are the following:

---

12. Wilson C, *Botnets, Cybercrime, and Cyberterrorism: Vulnerabilities and Policy Issues for Congress*, CRS Report for Congress (2008), http://bit.ly/aIzr1k.
13. Janczewski L and Colarik AM, *Cyber Warfare and Cyber Terrorism,* Hershey: Information Science Reference (2008).
14. Anti-Phishing Working Group*, Phishing Activity Trends Report* (2009), http://www.antiphishing.org/reports/apwg_report_Q4_2009.pdf.
15. Ibid, p.7.
16. Fraud Action UK, *Financial Fraud Action UK Announces Latest Fraud Figures*, Press Release (7 October, 2009), http://www.banksafeonline.org.uk/documents/ 2009H1FraudPressRelease.pdf.
17. Brown, Edwards and Marsden, supra note 11.

- In 2008 Internet security firm Symantec identified 1,656,227 distinct new malware and spyware software programs.
- In that same year, Symantec identified 9,437,536 infected machines part of botnets and other zombie networks.
- The FBI/Computer Security Institute reported that by 2007, there were 10,000 denial of service attacks daily, with costs for each target ranging between $90,000 USD to $6.45 million USD.
- In 2007, the US Federal Trade Commission received 221,226 Internet-related fraud complaints, totalling $525,743,643 USD.[18]

By any reckoning, all of these figures serve to stress the seriousness of the problem presented by cybercrime to businesses, governments and individuals. The level of threat is compounded by the fact that very often, law enforcement is not particularly prepared to deal with high-technology threats, and most of the attacks tend to go unpunished. Given the level of damage done by hackers and other malicious cybercriminals to the economy, the level of enforcement is remarkably low. Take phishing for example. Given the large number of offences every month, these rarely result in indictments, and while there are exceptions[19] it seems like cybercriminals are usually operating practically unopposed.

This situation has led to a shift from strict law enforcement strategies and has resulted in technical solutions being favoured. It is now practically unheard of to have a computer connected to the Internet that has no firewall and anti-virus software, which gives an indication of where the fight against cybercrime truly lies.

Take botnets as an example. Given the fact that there are millions of infected computers, it has become clear that it would be very difficult to try to go directly after the perpetrators. The best strategy seems to try to create filtering systems that might

18. Ibid.
19. "Phishing Indictment Includes More Than 100 Defendants", *Computing Now* (7 October, 2009), http://www.computer.org/portal/web/news/home/-/blogs/1866045.

diminish the negative effects of botnet action, particularly in denial of service attacks.[20] While this approach may seem defeatist, it is the most logical and efficient course of action.

Cybercrime relies heavily on existing infrastructure and architectures to thrive. Scholars dealing with Internet regulation have commented on the fact that architectural decisions made early about the way in which the Web operates have made cybercrime easier. Zittrain[21] has been at the forefront of warning about the choices made in the Internet's early days, which have now been translated into a more vulnerable system. He posits that the Web was created as a generative space more concerned with stability, scalability, resilience and the ease of spreading information than with security. Once it became profitable for unscrupulous individuals to try to disrupt the network, the existing architecture was ill-prepared to meet the challenge. He comments:

> [S]urfing the World Wide Web often entails accepting and running new code. The Web was designed to seamlessly integrate material from disparate sources: a single Web page can draw from hundreds of different sources on the fly, not only through hyperlinks that direct users to other locations on the Web, but through placeholders that incorporate data and code from elsewhere into the original page. [...] To visit a Web site is not only to be asked to trust the Web site operator. It is also to trust every third party – such as an ad syndicator – whose content is automatically incorporated into the Web site owner's pages, and every fourth party – such as an advertiser – who in turn provides content to that third party.[22]

It is slightly ironic that what makes the Internet such a vibrant medium also makes it vulnerable to cybercrime offences. While he does not mention it specifically, Zittrain's model of the generative web is yet another example of self-organisation in action, and one that in this case can have negative effects for users. If the underlying structure of the Internet is vulnerable, then we will continue to play catch-up with cybercriminals because the system allows the misuse of the technology in order to commit offences.

---

20. Maurushat, supra note 9.
21. Zittrain J, *The Future of the Internet: And How to Stop It*, London: Allen Lane (2008).
22. Ibid, p.56.

This is why a better understanding of the network can help design more effective strategies to tackle cybercrime. It is hereby proposed that network theory can be used to tackle some of the illicit actions listed above. To do that, we must understand a bit better what network science has to tell us about the Internet's structure. The next sections will highlight two areas where network theory may do just that.

2. NETWORK CENTRALITY

As stated repeatedly in Chapter 4, the Internet is supposed to be a distributed network where data is sent through various intervening points within the network in packets, and therefore it is not a centralised system. Moreover, the rapid growth of the Internet has produced a network that displays several power law topologies, and it also operates in a highly autonomous fashion. However, in that same chapter it was demonstrated that the Internet is also increasingly centralised at selected choke points, particularly at the national level, as various governments around the world attempt to regulate cyberspace by generating a national infrastructure that has fewer points of entry, and therefore it is easier to control and filter content from the wider network. This situation has created a much less distributed network than originally envisaged. The problem with a more centralised Internet is precisely that it is much more vulnerable to attacks than the distributed network that we were supposed to have.

It is essential to study what we mean exactly by network centrality to understand its relevance to the present work. One of the most vital elements shared by both small world networks and scale-free networks is the significance of individual hubs within the network. For example, it has already been explained that in any given social network there are central hubs that serve as connectors, improving the inter-connectedness of the entire network. Think back to your own social network, and certainly you can think of one or several individuals that stand out as knowing lots of people. These can be said to be more central vertices in the network. In graph theory, centrality means a way to

measure the importance of any given vertex within the network.[23] However, there are many measurements that one could try to use in order to ascertain the importance of any given node. Is the node very informed, but isolated? Is the node social, but bad at communicating information? Graph theory looks at three main measures of centrality to study the importance of a node:

> Three measures were formalised: degree, closeness, and betweenness. Degree was the number of ties or neighbours of a node; closeness was the inverse of the sum of all shortest paths to others or the smallest number of ties to go through to reach all others individually; and betweenness was the number of shortest paths on which a node was on.[24]

While these three main points have been added to and modified throughout the years,[25] what is important to note is that there are analytical tools available if one is to attempt to measure network centrality. What is vital for such analyses is that accurate data is gathered with regards to the three main elements of network centrality, namely, that one knows the number of links a node has to neighbouring nodes, the shortest number of paths to other nodes in the network, and the average shortest path (Figure 7.1). If you recall what has been discussed about small world networks, it is evident that the concept of centrality is vital to occurrences such as the six degrees of separation because it helps us to determine the importance of a node within a network, but also allows us to measure the shortest paths between nodes.[26]

---

23. Freeman LC, "Centrality in Social Networks: Conceptual Clarification", 1:3 *Social Networks* 215 (1979).
24. Opsahl T, Agneessens F and Skvoretz J, "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths", 32:3 *Social Networks* 245 (2010).
25. Ibid.
26. Newman MEJ, Barabási A-L and Watts DJ, *The Structure and Dynamics of Networks*, Princeton, NJ; Oxford: Princeton University Press (2006), p.342.

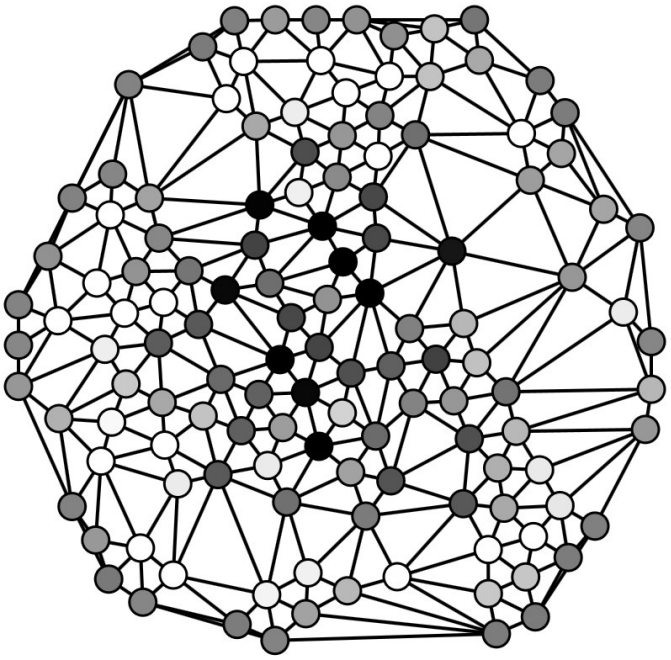


Figure 7.1 Graph betweenness[27]

So, what does the Internet look like taking centrality into account? This is actually a problematic question, as there are various methods of trying to measure the Internet topology, from fine-grained system-by-system IP measurement, to a broader autonomous system (AS) analysis.[28] The diversity of measurements can result in highly biased results, as the observed topology will be painted by the vantage point used to observe the network.[29] The relevance of this important fact for the current work is that if one is expecting to measure things like the potential centrality of a node in the network, the result may be biased by the vantage point used to measure such centrality. Imagine that one wants to try to ascertain the relative centrality of a node. The results would vary depending on whether one was behind a national firewall, or if the measurement was

27. Darker dots represent more central nodes as measured by betweenness, outside dots are peripheral. Image under an Attribution-Share-alike CC licence: http://en.wikipedia.org/wiki/File:Graph_betweenness.svg.
28. Shavitt Y and Weinsberg U, "Quantifying the Importance of Vantage Points Distribution in Internet Topology Measurements", *Proceedings of INFOCOM 2009 IEEE* 792 (2009), p.793.
29. Ibid.

conducted in a more "central" hub, such as a survey conducted in Google servers. While there are proposed solutions[30] to solve the problem of bias in Internet surveys, it must be stressed that any analysis about the Internet topology may still be biased.

Keeping that in mind, what does the Internet topology look like from the perspective of network centrality? One must add two more concepts about Internet architecture that are relevant to the question of centrality: (1) the physical elements of the Internet, namely routers, name servers, fibre optic cables, satellites and wireless hubs; and (2) the logical elements, such as the intervening logical paths information must travel from one point to another, incoming and outgoing links, search engine relevance and hosting service providers. Let us illustrate the difference by using a blog as an example. Imagine a website that is hosted in a server located in the UK, and you want to access it from the United Arab Emirates. The physical elements connecting you to that site would be the intervening servers, some of which may be more vital, for example the national firewall, the physical network that connects a server from the UK to your laptop in a hotel in the UAE. If one were to measure the centrality in such a network, one would have to look at the Internet backbone, the infrastructure connecting sites. Now, let us imagine the logical centrality, you would be looking at its importance within the overall Internet. Does the site show up in search engines? Are there other sites linking to the page you want to access? Were one to measure the centrality in the physical network, the more central elements would probably be the critical intermediary choke points, such as name servers and routers. Were one to measure the centrality in the logical network, the principal elements would probably be search engines, or influential linking sites. In other words, the physical and logical central nodes need not be the same.

Therefore, Internet topology depends on whether one is looking at the infrastructure, or the flow of information. So, we ask again, what does the Internet look like?

30. For example, a solution is to deploy distributed measuring agents, See: Shavitt Y and Shir E, "DIMES: Let the Internet Measure Itself", 35: 5 *Computer Communication Review* 71 (2005).

From the physical perspective, it is becoming clear that the Internet is nowadays more centralised than originally designed. In an interesting study on centrality,[31] researchers found that when one looked at the Internet backbone connections between large cities, and compared it to pre-Internet networks such as the airline transportation system, there was a striking similarity in which hubs were central to the network. In this study, researchers took data from a survey of international ISPs headquarters from 59 countries and 180 cities, and paired cities in order to measure how data flowed from one city to another, assuming that there were no direct backbone connections to each other. This is consistent with the graph theory concept of centrality explored above. Then they looked at the connecting passenger data between cities by looking at airline traffic between city pairs. What this approach provides is a picture of the centrality of nodes in both networks. In other words, if you wanted to get information or passengers from one city to another, and they were not connected directly, then a central hub would be a city where either would have to go through in order to reach its destination. The results were striking, as both the Internet backbone and the airline system had the same top five cities as the more central hubs: London, Paris, New York, Amsterdam and Frankfurt, with London being the most central city by far in both networks. This tells us that Internet centrality matches geographical centrality. It is easy to see why the results are as they are, as the historical importance of London as a transportation hub has survived the Internet revolution. Be it data or passengers, London remains central.

While London is a key hub in the physical infrastructure of the Internet, for historical reasons the United States as a whole still remains as the most central country in cyberspace terms. After all, the Internet arose from US military and academic networks, so it is only reasonable that it retains a high level of centrality. For example, the description of an early average data transfer from the UK to Australia serves to illustrate this point:

---

31. Choi J, Barnett GA and Chon B-S, "Comparing World City Networks: A Network Analysis of Internet Backbone and Air Transport Intercity Linkages", 6:1 *Global Networks* 81 (2006).

> The first example is a trace from University College London to the website of an Australian Internet Service Provider. Unlike a telephone transmission, which sets up a dedicated circuit that remains open between caller and receiver, Internet data travels in discrete, destination-marked packets more similar to the way letters are transmitted through a postal system. After leaving the university, data packets cross the Atlantic on a dedicated link to New York leased by JANET, the UK's scientific research network, and transit the United States on the UUNet network. Arriving in Los Angeles, they leave for Sydney where they will be offloaded onto the Australian Internet service provider's network.[32]

This level of centrality is a hang-up from the way in which the global backbone arose, so the reliance in hubs located in the US and Europe have been there from the start. The result is that most of the Internet traffic passes at some stage through the United States, even if the exchanging countries are close to each other. Cukier cites the example of Singapore and Malaysia, two neighbouring countries that used to send more than ten times the amount of traffic to the United States than to each other.[33] Another example of the inefficient infrastructure could be found in Africa, where almost every country needed to connect to the Internet using an industrialised nation.[34]

While the centralised nature of the global Internet backbone has been improving, it still shows large levels of centrality around a few countries. For example, the DIMES Project distributes data-gathering autonomous agents around the world to produce pings and traceroutes of the global Internet in an attempt to paint a more accurate topology of its centrality.[35] The emerging topology is one where servers and routers located in the United States still reign supreme as the most central constituents in the worldwide network, and results in some staggering images of just how centralised the Internet still is (Figure 7.2).

---

32. Townsend AM, "Network Cities and the Global Structure of the Internet", 44:10 *American Behavioral Scientist* 1697 (2001).
33. Cukier KN, "Bandwidth Colonialism? The Implications of Internet Infrastructure on International E-Commerce", *INET99 Conference*, San Jose CA (June 1999), http://www.isoc.org/inet99/proceedings/1e/1e_2.htm.
34. Ibid.
35. Shavitt and Shir, supra note 10.

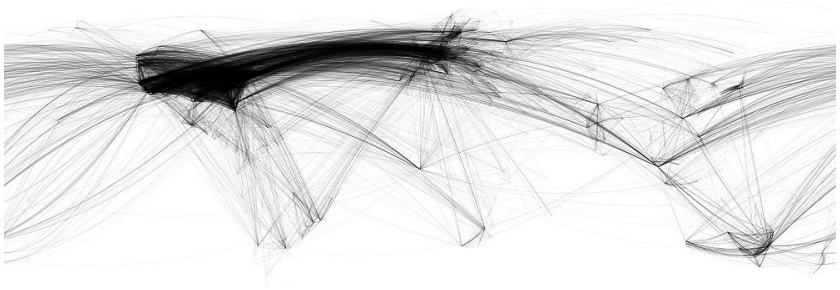


Figure 7.2 Internet city-to-city backbone connections[36]

What about logical centrality? This could potentially be different to the underlying architectural centrality of nodes, as what we are dealing with are people, websites and more dynamic systems. After all, the physical architecture of the network tends to be more static, and as it has been shown above, the pre-existing importance of nodes is carried out through time. Logical elements of the Internet need not respond to these constraints, websites accumulate links, become more or less popular, hubs come and go, companies fall from favour and new ones pick up in importance. Three years ago MySpace seemed destined to dominate the social network environment online, and not many people would have thought that Facebook would rise to the levels of popularity that the service has at the time of writing. There is even a chance that by the time you are reading this, these networks have been surpassed by a newcomer. The logical importance of hubs is ever-changing.

Nevertheless, evidence here also seems to indicate a high level of centrality in the information aspects of the network, with the United States again in front. For example, a study in 2001 looked at the number of national domain names as a measure of content-creation for each country, and found that 54 percent of all Internet content at that time

36. Image created by Chris Harrison (Carnegie Mellon University) using DIMES 2007 data: http://chrisharrison.net/projects/InternetMap/index.html. Each line represents a direct data connection between cities.

was either hosted in the US, or had an American-assigned domain name.[37] Recent data seems to corroborate this trend. ICANN maintains a list of all Internet domain name registrars around the world; these are the entities in charge of assigning and selling domain names. The US has almost 60 percent of all registrars, and has almost four times more than the second country in the list, Canada.[38] Interestingly, this distribution would also hint at the underlying existence of power laws.

But the location of domain names and registrars is only one piece of the information layer of the Internet – there appears to be a great disparity about the number of Internet users, and the location of content online. For example, by July 2010, North America had only 13.5 percent of the entire Internet population, with Asia and Europe commanding 42 percent and 24.2 percent respectively.[39] However, a survey of the localisation of IP addresses, that is, the actual computers connected to the Internet, produced some contradictory results. The United States commanded a staggering 1.5 billion IP addresses, while China had 257 million in second place, and the UK had just over 200 million. In fact, the US had more IP addresses than the other top ten countries combined.[40] This dominance translates into page views. Most of the 100 most popular websites in the world by January 2010 are American companies or organisations, with only 12 non-US-hosted websites.[41] All of this indicates that while other countries are making inroads in Internet penetration and the number of users online, most content and computers in the network are still located in the United States. This would certainly display a high level of logical centrality in the global system.

37. Zook MA, "Old Hierarchies or New Networks of Centrality? The Global Geography of the Internet Content Market", 44:10 *American Behavioral Scientist* 1679 (2001).
38. ICANN, *Geographic Distribution Of Registrars* (2010), http://forms.icann.org//idashboard/ public/.
39. Internet World Stats, *World Internet Usage and Population Statistics* (2010), http://www.internetworldstats.com/stats.htm.
40. Domain Tools, *IP Counts by Country* (July 2010), http://www.domaintools.com/internet-statistics/country-ip-counts.html.
41. "Superpower: Visualising the Internet", *BBC News* (January 2010), http://news.bbc.co.uk/1/ hi/technology/8562801.stm.

Another study into the centrality of hyperlinks, an essential part of the information layer of the World Wide Web, seems to produce similarly high levels of US centrality.[42] The study looked at 356 million hyperlinks and analysed incoming and outgoing links in each page, noting where the site was located and to which country it linked, and vice versa. Overall, the US came up as the most central country, followed by Australia, the UK, China, Japan, Canada and Germany. The US had the most incoming links, while Germany had the most outgoing links to other countries. The interesting aspect of this study is that it looked at reasons for the level of centrality, trying to correlate the results with economic and cultural preferences. Unsurprisingly, economic aspects accounted for high centrality in the hyperlink network. Perhaps more surprising is that there was a cultural correlation as well, with countries with highly individualistic cultures dominating also.[43] This tells us that centrality has not only architectural elements, but also a cultural level.

These findings appear to be consistent with what we know of how large networks, and particularly scale-free networks, operate. Besides the architectural and historical importance of the US to the emergence of the Internet, it is only natural that because of these initial conditions, the global network would display high distribution of nodes according to the initial conditions in the network. The network simply self-organises around those conditions. Nobody tells people to visit US sites more, it just happens because those nodes tend to be older, and older nodes accumulate links faster. Moreover, large networks tend to display centrality levels distributed according to power laws;[44] in other words, scale-free networks will always have fewer important nodes, the very definition of centrality in graph theory.

---

42. Barnett GA and Sung E, "Culture and the Structure of the International Hyperlink Network", 11:1 *Journal of Computer-Mediated Communication* 217 (2005).
43. Ibid.
44. Latora V and Marchiori M, "A Measure of Centrality Based on Network Efficiency", 9 *New Journal of Physics* 188 (2007), p.196.

3. SOCIAL NETWORK ANALYSIS

One of the most exciting areas of network study is the subject of social network analysis (SNA). As has been discussed extensively in previous chapters, social networks are groups of individuals that are connected to and interdependent on one another. These individuals have interactions that can range from family ties to friendship, employment information and development. A social network can be understood "as any bounded set of connected social units".[45] Social networks then rely on three key building blocks: the boundary of social elements studied, be it a family, a tribe, or a country; then the connected element between the social units, which are the links that tie the units together; and the definition of social unit itself, these are usually individuals, but also can be groups of groups, so we could have social units consisting of organisations and institutions.[46] SNA is therefore a systematic way of looking at these networks in an analytical fashion by using graph theory in order to provide useful information about the group.

Social network analysis is not a new subject, it started as the study of social groupings through psychological analysis of things like group cohesion and friendship, in what became known as sociographs.[47] The research and development into SNA continued in areas such as sociology and anthropology, but it was not until later in the 20th century that graph theory was used to analyse social interactions, such as looking at influence networks, and with the development of mathematical tools capable of analysing dynamic networks.[48]

Graph theory is therefore just a method of making sense of the various interpersonal interactions in a social system. Scott explains:

---

45. Streeter CL and Gillespie DF, "Social Network Analysis", 16:1 *Journal of Social Service Research* 201 (1993), p.202.
46. Ibid.
47. Freeman L, *The Development of Social Network Analysis*, Vancouver: Empirical Press (2006), p.14.
48. See Chapter 2.2.

> A common framework for social network analysis programs is the mathematical approach of graph theory, which provides a formal language for describing networks and their features. Graph theory offers a translation of matrix data into formal concepts and theorems which can be directly related to the substantive features of social networks. If the sociogram is one way of representing relational matrix data, the language of graph theory is another, and more general, way of doing this. While it is not the only mathematical theory which has been used for modelling social networks, it is a starting point for many of the most fundamental ideas of social network analysis. [...] The concepts of graph theory, then, are used to describe the pattern of connections among points. The simplest of graph theoretical concepts refer to the properties of the individual points and lines from which a graph is constructed, and these are the building blocks for more complex structural ideas.[49]

So, what is the usefulness of SNA? As one can expect from a field of study that has been around since the 1930s, social network analysis has provided a diverse and rich level of scholarship that looks at social interactions from a formal perspective. Typical questions asked in SNA range from comparing the level of connectedness achieved in a specific social setting, such as the difference between a family group and another social unit, to studies looking at whether one medium may influence the way in which people interact.[50] However, it is useful to look at examples of research into SNA that are relevant to the Internet, that being the main topic of this work.

A good example of practical applications of social network analysis in an online environment is a comparative study conducted in a group of distance learners, trying to determine how often they made contact with each other depending on the medium of communication used.[51] An important hypothesis presented by SNA studies is that the medium determines communication in some social contexts. In other words, while the content of the communication will be the same regardless of the medium, the number of people one can communicate with will depend on the medium used to establish a link; contact with non-friends will tend to be non-emotional regardless of the number of

---

49. Scott JP, "Social Network Analysis", 22:1 *Sociology* 109 (1988).
50. Several examples can be found here: Wasserman S and Faust K, *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press (1994).
51. Haythornthwaite C, "Social Networks and Internet Connectivity Effects", 8:2 *Information, Communication & Society* 125 (2005).

people involved, but with Facebook one can reach more friends and family. In this study, it was established empirically that this seemingly obvious statement held true by looking at the way non-emotional ties were acquired in the aforementioned group of distance learners. For example, when comparing the number of connections of the group between those using asynchronous communication tools such as email, and those using synchronous communication via Internet Relay Chat (IRC), it became clear that those who were chatting were more likely to interact with more peers than those simply emailing.[52] Moreover, the study also looked at communication over time, and found that wilful top-down direction from course organisers would have considerable effects on the shape of the social network graph – by changing groups around and requiring specific communication media, the level of interaction would increase considerably.[53] While many seasoned teachers would probably be able to give similar advice, the value of social network analysis in an online setting is that it gives us clear evidence about pedagogic practices that encourage interaction between learners. This is by all means a valuable tool that tells us a lot not only about teaching, but about how online social networks operate.

SNA is not only giving us valuable insights into media, but also about a question that has been at the forefront of network theory since the days of Solomonoff and Rapoport:[54] how do people influence each other in a social context? Christakis and Fowler have conducted a much publicised and enlightening study into the small American community of Farningham across several decades using medical records and online social networks.[55] They looked at the way people accumulated "friends" through Facebook, and found that something as simple as smiling in your profile picture would be a strong determinant in predicting the number of friends you would have. The authors

52. Ibid, p.131.
53. Ibid, p.134.
54. Solomonoff R and Rapoport A, "Connectivity of Random Nets", 13 *Bulletin of Mathematical Biophysics* 107 (1951).
55. Christakis NA and Fowler JH, *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives,* New York: Little, Brown and Co. (2009).

looked at two types of datasets, the physical network of Farmingham, and the Facebook network of an unnamed American university to compare levels of friendship and "happiness" between the online and offline world. The results were astoundingly similar. Firstly, by looking at happiness indicators in the Farmingham network, they found that people who seemed happy tended to cluster around one another in strong hubs, while unhappy people would linger at the periphery. Similarly, people who smile on Facebook tend to cluster around other smiling people. The data showed some interesting facts related to small world clustering:

- Each happy friend increased an individual's probability of being happy by 9 percent.
- Unhappy people connected to more happy people were more likely to become happy in the future, while those separated by more degrees would tend to remain unhappy. This fact carried through up to three degrees of separation.
- There is no noticeable difference between the number of close friends online and offline (average 6.6 on both). However, Facebook users have more casual "friends" (average 106 per user).[56]

The third relevant example of SNA in the online environment demonstrates that social networks display certain levels of predictability that can be useful in identifying nodes and links within the system. Krotoski[57] conducted an impressive in-depth social network analysis of the virtual world Second Life, trying to find whether one could predict attitudes and behaviours from social units by studying their interaction within the system. Krotoski adds to the growing evidence that social networks have extensive influence on individuals by stating that:

---

56. Lehrer J, "The Buddy System: How Medical Data Revealed Secret to Health and Happiness", 17.10 *Wired* (September 2009).
57. Krotoski A, *Social Influence in Second Life: Social Network and Social Psychological Processes in the Diffusion of Belief and Behaviour on the Web*, Ph.D. thesis, University of Surrey (2009), http://alekskrotoski.com/media_files/SocialInfluenceInSecondLife.pdf.

> Analysts have proposed that the structure of a network has implications for how much potential influence the social system may have on the individual. This is based on the principle of network exposure, which anticipates that the more people who have an attitude or perform a behaviour who are directly connected with an individual, the more likely the individual will adopt that behaviour or attitude. Exposure is progressive and maximal.[58]

One of the most interesting findings in this study is that social network influence seems to be immune to online anonymity; people will be influenced regardless of them being connected to a social setting or to a virtual persona in the shape of an avatar.[59] Here we revisit the concept of centrality (or network position to use the SNA terminology). Within SNA, influential nodes in the network are logically more influential. However, does this translate to online worlds? The answer seems to be a resounding yes; even when one removes a personal element, the influence of disembodied avatars matches that of personal social units.[60]

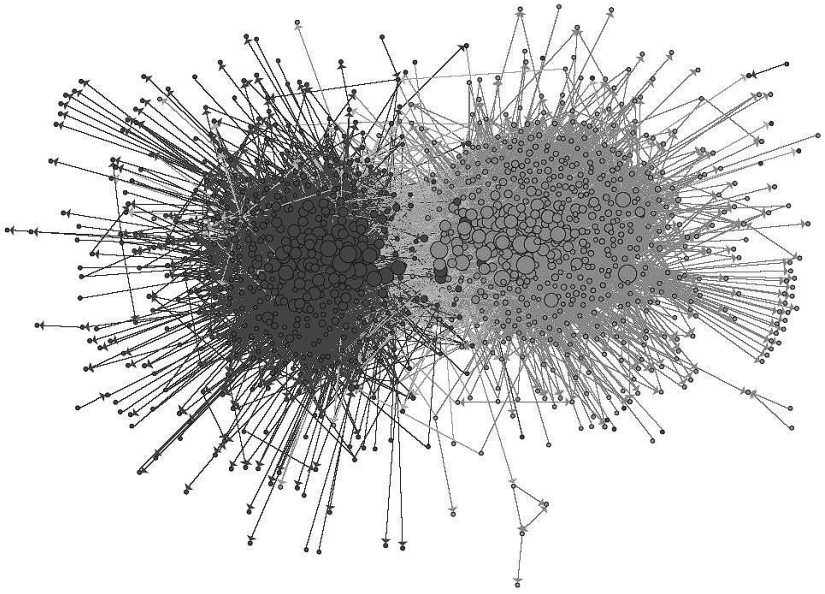
The final useful characteristic of social network analysis relevant to this work is the presence of a phenomenon that seems intuitive, and that is the fact that social networks tend to produce clusters of units and close-knit communities. According to Girvan and Newman:

> A third property that many networks have in common is clustering, or network transitivity, which is the property that two vertices that are both neighbors of the same third vertex have a heightened probability of also being neighbors of one another. In the language of social networks, two of your friends will have a greater probability of knowing one another than will two people chosen at random from the population, on account of their common acquaintance with you.[61]

---

58. Ibid, p.52.
59. Avatar is an Internet jargon term that describes an online persona which may or may not represent the physical human being operating it.
60. Schroeder R, *The Social Life of Avatars: Presence And Interaction in Shared Virtual Environments*, London: Springer-Verlag (2001).
61. Girvan M and Newman MEJ, "Community Structure in Social and Biological Networks", 99:12 *Proceedings of the National Academy of Sciences* 7821 (2002).

Why do we need methods of finding communities within larger scale networks such as the Internet? With a global network that has too much data about individuals operating in a social setting, it might be useful to delimit a community for various reasons. One may want to market only to targeted communities instead of mass spamming the entire network. SNA allows us to pinpoint accurately a cluster of individuals by analysing the ties between each one. The opposite might be true as well. For example, if you wanted to find an individual, it would be easier to do this by analysing his/her network of known associates. Granted, this type of analysis may tell us something we already know about networks and clusters, but nonetheless, the results can sometimes produce informative outcomes. For example, an often cited study of political blogs[62] in the US analysed blog communities by looking at which blog was linking to other blogs. The result is a striking display of political blogs distributed between conservative and liberal clusters that present one of the most beautiful and scary visual representations of the political divide (Figure 7.3).



62. Adamic LA and Glance N, "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog", *Proceedings of the 3rd International Workshop on Link Discovery* 36 (2005).

Figure 7.3 The US political blogosphere[63]

The relevance of this area of research will become evident later.

4. NETWORK THEORY AND CYBERCRIME

So, can network science tell us anything about cybercrime? This question can be split into two sub-components. Firstly, can we design better networks based on what we know about how they operate? Secondly, can we use some elements of graph theory to design enforcement mechanisms that will lead towards a much better record in finding and apprehending cyber-criminals? This section will try to answer both elements of the question by concentrating on the issues of centrality and social network analysis explored above.

4.1 Centrality and vulnerability

The first aspect to analyse when looking at the interaction between complexity and cybercrime is to look at whether the centrality of the networked systems may facilitate the commission of cybercrime offences.

The most explored area of research on this topic is that of computer virus propagation through the network. There is ample evidence that points towards the presence of high infection rates in networks that present some form of power law distribution of nodes. One study, for example, looked at the rate of spread of a virus through email, and found that it would spread faster in a power law network, than it would in another type of distribution, such as a small world network or a random graph.[64] The reason for this is that in a power law network the spread of a virus would be highly dependent on the

63. Those on the right, are conservative blogs, those on the left are liberal blogs.
64. Zou CC, Towsley D and Gong W, "Email Virus Propagation Modeling and Analysis", *CiteSeer* (2003), http://bit.ly/difAHB.

number of recipients an email had, but also would depend on the frequency with which a node would check his/her email. In a power law network, such as the email network studied by the paper, both these variables responded to power laws. In a random graph, where users are distributed randomly and where there are no influential hubs, this did not have much of an effect.

Moreover, virus spread seems to be highly dependent on the malicious software reaching a central node in the network, namely a node with a high degree of connectedness, then it is highly likely that the virus will reach epidemic proportions.[65] While this finding seems to be intuitive, it serves as further proof of the potential importance of network centrality for cybercrime purposes. Hubs are critical parts in the Internet architecture, and protecting them would be vital to avoid wider spread of infections.[66]

Computer virus infections can serve another purpose, and that is to look at potential social network elements of how malicious software spreads throughout the network. From the above, it seems that physical centrality within a scale-free network encourages viral infections. What about the logical centrality? For example, will a virus affecting the computer of an influential node in a social network have similar effects in the rate of spread of a virus? In a study set out to look precisely into this question,[67] Guo and Cheng gathered social network data for 14,933 students at an undisclosed university using MySpace, and produced a directed graph charting relational data between nodes, calculating centrality and clustering in the sample. The researchers simulated computer virus infections selecting random nodes in the network and then changed the simulation based on the centrality of nodes; in other words, they chose those nodes that seemed to have higher hierarchical value in the social network. The results were consistent with

---

65. Berger N et al, "On the Spread of Viruses on the Internet", *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithm*s 301 (2005).
66. Further evidence can be found in: Gang Y et al, "Epidemic Spread in Weighted Scale-Free Networks", 22:2 *Chinese Physics Letters* 510 (2005).
67. Guo H and Cheng H, "Computer Virus Propagation in Social Networks", *ICIS 2007 Proceedings* (2007), http://aisel.aisnet.org/icis2007/124.

what has been highlighted in other studies, and that is that centrality seems to be a strong determining value in the level of virus spread, the more focal the node to the network, the higher the infection rates.

While all of this may seem intuitive, it is baffling that an understanding on networks is not part of law enforcement strategies, and while the topic of cybercrime gets an increasingly significant treatment in legal scholarship, the basics of how viruses spread online is still largely ignored outside of computing systems research.

Study into other network-dependent cybercriminal offences offer similar findings about the importance of network topology in the detection, and potential filtering, of attacks on a network. An obvious example of this would be in denial of service attacks, where by definition, one node, or perhaps even a central cluster of nodes, is being subjected to an external attack. The traffic is incoming, so it does not really matter if the system is a central part of the network or not, as it is a target. However, centrality may have an essential bearing on the situation if the target is an important hub in a network, as the intention may be to knock down computers connected to the hub. In a typical web-based DoS attack, a botnet is used to send an overwhelming number of service requests against a server. The most effective way in which this can be stopped is by technical means, mostly through the deployment of some form of filtering that will keep out suspected attackers from the system.[68] The relevance of understanding network centrality in this issue is precisely to know where to deploy DoS defences. Key central hubs will have to be protected through technological means, as not doing it might compromise networks downstream.

It is precisely the issue of the Internet's architecture what makes centrality such an essential subject for the legal study of cybercrime. The Internet is resilient, but an attack on a central part of the network will have the potential to cause cascading failures in the network. This is a fact that cannot be stressed enough. As has been highlighted already,

68. Park K and Lee H, "On the Effectiveness of Route-Based Packet Filtering for Distributed DoS Attack Prevention in Power-Law Internets", 31:4 *SIGCOMM Computer Communications Review* 15 (2001).

the Internet is increasingly centralised, so the scope for potential large-scale attacks increases exponentially. Taking both the evidence from virus spread and DoS vulnerability explored already, an obvious concern emerges. If the Internet is more centralised than we had previously believed, and if malicious information spreads easily through networks that display power law distributions (and therefore rely on connecting hubs), it is possible to postulate the hypothesis that anything that affects the global hubs in the Internet backbone could easily be replicated throughout the Internet.

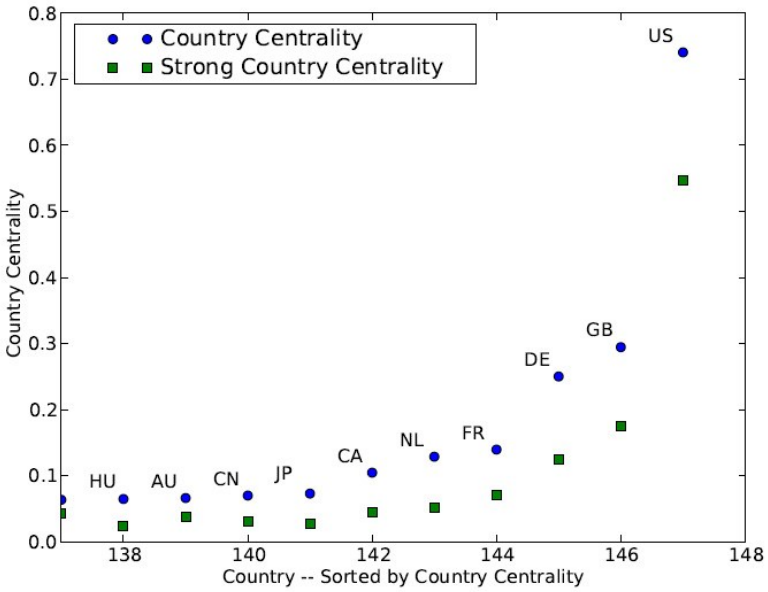


Figure 7.4 Strong country centrality (zoomed)

While not directly related to cybercrime, a study has looked precisely at this question. Karlin, Forrest and Rexford[69] conducted a survey of country centrality to try to determine the potential downstream negative effects of country-wide censorship of the Internet. The objective of the paper was to establish an analytical framework for determining the influence of each country within the flow of international traffic. The

69. Karlin J, Forrest S and Rexford J, *Nation-State Routing: Censorship, Wiretapping, and BGP*, arXiv Working Paper (2009), http://arxiv.org/abs/0903.3218v1.

researchers collected traceroute data between countries trying to determine the paths taken by information in the global network. This produced high levels of centrality consistent with the other studies highlighted in previous sections. What is novel about the approach of this study is that it also calculated what they call "strong country centrality" (SCC). They assumed that under some circumstances there may be other paths to information that do not go through one country. SCC would take place when all other viable paths led through that country as well; in other words, data had no other way of getting from A to B other than through that country. In findings consistent with other centrality studies, they found that the United States, the UK and Germany were the most central countries on the Internet, but also displayed high levels of SCC (Figure 7.4).

This is a study of great consequence for various reasons, particularly because it is one of the first to try to rank centrality of data at a country level, but more relevant to the subject of cybercrime, it offers strong evidence that points towards a worrying level of country centrality that still to this day favours Western countries. Most importantly, it shows that the Internet relies on those central hubs too much. Any attack on the global infrastructure will undoubtedly target the central points in the network.

The implications of such centrality are clear for another aspect of cybercrime, that of cyber-warfare. While this may not seem like a cybercrime subject, it seems that it has become so in recent years given the nature of attacks. One of the most publicised cases of cyber-warfare took place in 2007 against Estonia. The Baltic country found itself at the centre of a wide-scale cyber-attack from hackers against its information technology infrastructure due to a perceived slight against a Soviet-era statue.[70] The unprecedented attack targeted almost every aspect of the country's Internet presence, including media, government institutions and financial services. The reason why this is more of a cybercrime issue is that the attacks were conducted by Russian hackers, and while

70. Traynor I, "Russia Accused of Unleashing Cyberwar to Disable Estonia", *The Guardian* (17 May, 2007), http://www.guardian.co.uk/world/2007/may/17/topstories3.russia.

official Russian involvement has been denied, it is clear that large numbers of individual hackers were involved in some form or another.[71] The attacks managed to disrupt the country's digital infrastructure for days, although filtering international efforts managed to curtail the worst part of the attack. There are several interesting lessons about this incident, but one significant feature was that the attacks used dozens of botnets located around the world, encompassing almost a million separate computers.[72] The other issue is that while the attacks were conducted against individual websites in Estonia, the entire attack managed to knock down Internet connection throughout the country because the level of traffic overwhelmed the national infrastructure. It took high-level action from European root server authorities to try to minimise the damage, but many sites had to cut off their connection to the outside world.

A similar cyber-warfare attack took place against Georgia just before Russian troops invaded the country in August 2008. During the build-up to the Russian invasion, government, police and media websites were subjected to coordinated botnet attacks of such a scale that they brought down the networks, prompting several official services to relocate to servers outside of Georgia.[73] The pattern was similar to the Estonia incident, with the difference that these attacks were followed up with physical military intervention. The end result could not compete with the fact that there was an actual conflict taking place, but it did result in the virtual disappearance of Georgia from the global Internet due to the virulence of the cyber-attacks.[74]

Estonia and Georgia are just two examples of what the problems for country-level centrality exposed above could look like. While in these two cases the affected countries were not particularly central, imagine a similar co-ordinated attack on a more influential country, and you could begin to see the potential for wider disruption. In both attacks,

---

71. Davis J, "Hackers Take Down the Most Wired Country in Europe", 15.09 *Wired* (21 August, 2007), http://www.wired.com/politics/security/magazine/15-09/ff_estonia.
72. Ibid.
73. Danchev D, "Coordinated Russia vs Georgia cyber attack in progress", *ZDNet* (11 August, 2008), http://zd.net/aKJese.
74. Markoff J, "Georgia Takes a Beating in the Cyberwar with Russia", *The New York Times* (11 August, 2008), http://nyti.ms/HYPeX.

the Internet backbone in those countries was severely affected because of increased traffic, despite the fact that the assaults were directed not at the actual infrastructure, but at websites within the countries. A smarter and more targeted strike against national domain name servers could have had an even greater effect. Now imagine a similar scenario taking place against countries that are even more centralised, such as countries with national firewalls, and it is be easy to see how this could remove those countries from the Internet altogether.

Even more worrying, an attack against a country with strong country centrality could affect Internet traffic not only within the target, but also would affect international traffic that relies on data going through the network at a central hub.

The message is clear: the more centralised we make the Internet, the more vulnerable we become to co-ordinated attacks. While it may be easy to laugh at the preposterous scenarios painted by movies like *Die Hard 4.0*, the threat is real, and we need to heed the warnings from network science in this respect. Thankfully, some sectors of law enforcement seem to be taking the potential threat seriously. In 2007, the US Department of Homeland Security conducted a military exercise called Cyber Storm, where a simulated assault from domestic terrorists, German hackers and some insiders was able to crash the Federal Aviation Administration computer control system, post false data and shut down commuter services.[75] While this was a limited simulation, the message about the troubles posed by centrality is real.

4.2 Social network analysis and cybercrime

The second area of study of network theory that may have valuable input in the detection and prevention of cybercriminal offences is that of social network analysis.

The application of network theory to criminal social networks is one of the most exciting practical applications of the various theoretical characteristics of networks

75. Brenner S, "At Light Speed: Attribution and Response to Cybercrime/Terrorism/Warfare" 97 *Journal of Criminal Law and Criminology* 379 (2007), p.395.

described in previous chapters. Humans are social creatures; our interaction with one another is an important element of social structures and criminality is undoubtedly one situation where interaction occurs. Criminals have to operate as well in these social settings; they have friends, families and conspirators, so it is only logical that any study of human links will look into the seedier aspects of social life. For example, there appears to be a strong correlation between societies that have strong connections and low crime rates.[76] Social integration in the shape of strong social ties, or a sense of community demonstrated in the some forms of organised meetings between its members, seems to demonstrate social cohesion that translates into less criminal activity due to peer pressure and community control and surveillance.[77]

Not only is social interaction an essential determinant to criminality levels, but criminal organisations themselves seem to self-organise in ways where social connectedness and centrality operate in recognisable patterns present in other social settings. An interesting study[78] of African American and Hispanic street gangs in Newark, New Jersey identified 736 gang members distributed into four main gangs. While each individual gang maintained a small world structure, the interesting part is that there was interaction between gangs by the presence of connecting individuals, or cut-points. The removal of these connectors sent the organisation into disarray, and had a strong effect on the overall social organisation.[79] This is consistent with what we know of how social networks operate, but offers an interesting insight into criminal groups. Given enough data, it is possible to chart criminal groupings just as any other social network, and it is also possible to try to use this data in practical ways. McGloin suggests, for example, that by knowing the social structure of a gang, law enforcement

76. Bellair PE, "Social Interaction and Community Crime: Examining the Importance of Neighbor Networks", 35:4 *Criminology* 677 (1997).
77. Ibid, p.680.
78. McGloin JM, "Policy and Intervention Considerations of a Network Analysis of Street Gangs", 4:3 *Criminology & Public Policy* 607 (2005).
79. Ibid, p.620.

can direct its efforts into trying to gather useful data about associates and social patterns to better tackle intervention and allocate resources accordingly.[80]

The practical application of social network analysis to criminology rests on the assumption that criminal organisations can display similar characteristics to other social networks, and specifically small world networks. Coles[81] proposes that the same principles present in Milgram's small world networks are also to be found in larger organised crime groups; he states that acquaintance chains are also at work, and suggests that what keeps the network together are those specialised individuals in the network that act as connectors. What is innovative in Coles's analysis is that he postulates that if one wants to look at how the network is organised, one should look at the networks of acquaintances of those who have been confirmed to be part of the organisation, and by looking at the "friends of friends" of these individuals it is possible to get valuable insight into the composition of the network. Nonetheless, Coles has been criticised as presenting a rather non-nuanced approach to organised criminal networks. For example, Chattoe and Hamill[82] have commented that any social network analysis of criminal groups requires more than simple gossip about who is friends with whom, and that a quantitative study of the structure of the network is required. They use as an example terrorist network analysis to make their point:

> The disruption of terrorist networks has rapidly spawned a literature in the aftermath of 9/11. However, almost without exception, the work that is not merely anecdotal proves results using formal models that disregard both distinctive ethnographic knowledge of terrorist 'culture' and the working practices and insights of law enforcement agencies. For this reason, it is unlikely to have any lasting policy impact. Simulations constructed using ethnographic police data and access to the kind of

80. Ibid, p.623.
81. Coles N, "It's Not What You Know – It's Who You Know That Counts. Analysing Serious Crime Groups as Social Networks", 41:4 *The British Journal of Criminology* 580 (2001).
82. Chattoe E and Hamill H, "It's Not Who You Know – It's What You Know about People You Don't Know That Counts: Extending the Analysis of Crime Groups as Social Networks", 45:6 *The British Journal of Criminology* 860 (2005).

reasoning the intelligence services use about networks seem likely to be far more productive.[83]

This is an essential point to keep in mind. It is tempting to try to gather data and make assumptions about the shape of a social network based on vertices and edges in a graph. Quantitative and qualitative understanding of the network is still required at some level. Undoubtedly, having better pictures of the small world shape of any social group is a starting point into promoting better understanding of how criminal networks operate.

Can the work being conducted into criminal gangs be translated into useful analysis of cybercrime? Criminal groups online also operate in social networks, so there is no reason why this should not be the case as well. Chau and Xu[84] have conducted an interesting analysis of blogs loosely identified as "hate groups", which are sites that publish blatantly racist content. First they identified blogs that had already been highlighted by other research as containing extremist racists views, then they set out an autonomous agent that extracted useful link data from the blog (external links, comments, incoming links), and also copied content for textual analysis of the content. The researchers then conducted a topological structure of possible networks of racist blogs by using centrality analysis, such as looking at average shortest paths, clustering and degree distribution, which as has been mentioned are commonly found in SNA and centrality studies online. Perhaps unsurprisingly, this resulted in a tell-tale power law graph where a few blogs accumulated a larger number of incoming links, hinting at a pattern of link distribution that is to be found in other online communities.[85] The social analysis of the groups also produced expected results that are compatible with offline groups such as the ones described earlier. Even in an online environment, racist blogs exhibited high clustering, similar to the formation of gangs, and these networks relied on popular connectors but also exhibited influential individuals and blogs that crossed

83. Ibid, p.867.
84. Chau M and Xu J, "Mining Communities and Their Relationships in Blogs: A Study of Online Hate Groups", 65:1 *International Journal of Human-Computer Studies* 57 (2007).
85. Ibid, p.62.

across several communities, just as is the case in gang-related studies.[86] While the blogs may not be criminal per se, this study hints at the existence of centralised, clustered social structures in online groups.

However, the aforementioned method of using SNA to determine the composition and structure of the criminal network is limited by the fact that more serious cybercriminals do not advertise online, or have blogs with incoming and outgoing links. While this limitation should temper the enthusiasm for the deployment of social network analysis against cybercrime, this does not mean that there are not areas that could be subject to SNA study.

One controversial area of study where SNA has both been suggested and deployed is the subject of cyber-terrorism. This is unsurprising, as the so-called "War on Terror" set up by Western countries after the September 2001 attack against the United States has opened researchers to a welcome source of research funding. One of the first to suggest the use of network science in the fight against terrorism is Barabási, who makes an impassioned argument about the potential use of graph theory and SNA in the detection and destruction of terrorist cells.[87] He explains that the understanding of social networks could be used against terrorist networks by identifying members, and then he postulates that they could be vulnerable to targeted strikes against the hubs holding together the network. It is a tantalising promise, but one would need evidence that terrorist organisations behave according to power laws. While the evidence for this statement is still in the early stages, there is growing indication that at least security services are taking it seriously. Some basics of network theory are being taught in military schools in the United States.[88] Perhaps most intriguingly, several reports indicate that Saddam

86. Ibid, p.68.
87. Barabási A-L, *Linked: The New Science of Networks,* Cambridge, MA: Perseus Pub. (2002), pp.222–223.
88. *Connected: The Power of Six Degrees* (2008), Science Channel Documentary.

Hussein was found and apprehended by using social network analysis by looking at his network of associates and their movements around the loyalist area of Tikrit.[89]

While SNA could be useful in identifying terrorist networks, its application to so-called cyber-terrorism is more difficult, because up until now there has not been a documented cyber-terrorist attack.[90] However, it seems clear that the Internet has been used by terrorists to organise and communicate with one another, and this opens up the scope of the use of SNA to try to detect and possibly even prevent terrorist attacks. The challenge for this, and many other Internet-related criminal activities, is that while the data may be available, the analysis of the information may be lost in a sea of reports that drown out useful intelligence.[91] This could be a challenge met by careful and judicious use of social network analysis. By looking at websites which exchange terrorist and extremist materials, and correlating it with social network data and small world analysis, it could be possible to get a picture of potential future terrorists.

The current use and effectiveness of online social network analysis in the identification of online terrorist networks is difficult to ascertain at this point because intelligence services are understandably not forthcoming about their methods. However, there is a documented example of a missed opportunity with regards to telephone surveillance data which may serve to illustrate the potential validity of using SNA to fight terrorism. It is possible that the UK's intelligence community may have had in its grasp valuable data that could have been used to prevent the 7 July 2005 terrorist attacks in London. In March 2004, the Metropolitan Police conducted Operation Crevice, a raid launched against terrorist cells of Pakistani origin, which resulted in the indictment and conviction of seven individuals.[92] As a result of the arrests, police and security services conducted an analysis of phone calls made within the network, and identified 4,020 calls

89. Wilson C, "Searching for Saddam: The Social Network That Caught a Dictator", *Slate* (February 22, 2010), http://www.slate.com/id/2245228/.
90. Brenner, supra note 75, p.390.
91. Chen H and Xu J, "Intelligence and Security Informatics", in Cronin B (ed), *Annual Review of Information Science and Technology*, Volume 40, Medford NJ: Information Today (2006), p.237.
92. "UK seven 'were ready to start bombing'", *The Guardian* (21 March, 2006), http://www.guardian.co.uk/terrorism/story/0,,1736228,00.html.

related to the Crevice investigation.[93] No social network analysis was conducted of the participants of those calls. If there had been some form of systematic and informed analysis made, then authorities may have identified two members of the phone call network, Mohammed Siddique Khan and Shazad Tanweer, two of the 7/7 bombers (figure 7.5).

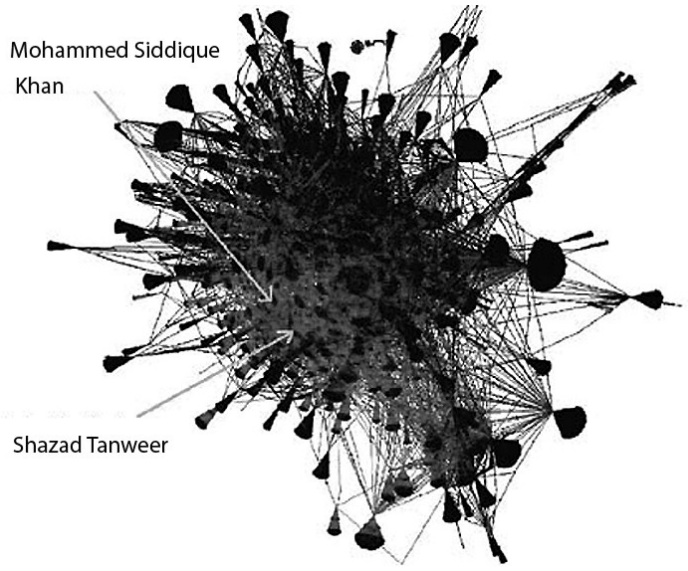


Figure 7.5 Phone call network of Operation Crevice surveillance

This seems like a clear instance where any sort of understanding of social networks might have prevented an atrocity. While it is unfair to deal in "what ifs", one only needs to look at the chart above to wonder if a qualified expert in SNA might have identified crucial central nodes in this terrorist network.

Besides the study of criminal rings as social networks, there is another problem that can be highlighted. What would happen if cybercriminals became aware of network theory, and started using social networks to commit crimes? If one accepts the theory

93. Intelligence and Security Committee, *Could 7/7 Have Been Prevented?* Cabinet Office Review of the Intelligence on the London Terrorist Attacks on 7 July 2005 (2008), http://www.cabinetoffice.gov.uk/media/210852/20090519_77review.pdf.

that cybercrime has become a cyber-arms race between law enforcement agencies, industry and criminals, then it would be possible to envisage a situation where individuals intent on committing offences may use the same theories explained previously in order to commit better crimes.

This is not such a far-fetched idea. There is an indication that some cybercriminals are already using the Internet to gather useful data about potential targets in what some researchers call "context aware phishing".[94] This is a more targeted phishing attack on specific targets, where freely-available data about friends, shopping preferences and browser history can be gathered online and can be used to tailor a very believable message that has more chance of prompting a response from the victim. Jagatic et al[95] conducted an experiment using Indiana University students as subjects. The researchers collected public data from blogging sites, social network sites and other machine-readable data in order to gather information about the target's contacts and preferences. They then harvested the data and produced a database with tens of thousands of relationships. They then divided the group into a control group that would receive anonymous messages, and one that seemingly came from someone within their social network. The study sent an actual yet harmless phishing attack on both groups, asking them to enter their secure University credentials. Only 16 percent of the control group provided their details, while an astounding 72 percent of those in the social network group responded positively to the attack.

If you add this study to our understanding of social networks, then the results are worrying to say the least. Take Facebook, for example, where at the time of writing, 500 million users shared their personal details with friends, family, acquaintances and co-workers. While a large number of Facebook users nowadays share personal data only with their "friends", that is, other people in their network, a less publicised feature of this vast network is that it allows search engines to crawl through the contact details.

---

94. Jakobsson M, "Modelling and Preventing Phishing Attacks", *Lecture Notes In Computer Science* (2005), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.64.1926.
95. Jagatic TN et al, "Social Phishing", 50:10 *Communications of the ACM* 94 (2007).

This is an invaluable tool for anyone who is looking to mine information. Bonneau et al[96] conducted a social network analysis on Facebook's public listings in order to gather information about users' networks and their centrality within the network. The stated purpose of the study was to create social graphs by using public friendship links. The results were worrying, as the paper found that it was possible to construct an accurate picture of a person's closest contacts by simply analysing data that can be obtained through web searches. When one connects this research to "context aware phishing", then we should really be concerned about the availability of information made available through social network sites.

The implication of all of the evidence presented is clear. Social network analysis offers a powerful tool against cybercrime, both as a means of trying to identify criminal organisations, but also as a warning about the amount of data available on the Internet that can be misused. It is hoped that by highlighting the rich analytical tools available to policymakers and law enforcement agencies, colleagues in the legal profession may start to look harder at the potential of network science to the fight against cybercrime. In danger of over-stating one of the objectives of this work, a better understanding of networks can only produce positive results.

5. A NEW INTERNET?

If complexity theory can teach us anything about how networks operate, it is that complex dynamic systems such as the Internet are self-organising networks. Enough evidence has been provided to support the statement that the Internet shows emergent characteristics resulting from the ad hoc engineering decisions that created it.[97] From a

96. Bonneau J et al, "Eight Friends Are Enough: Social Graph Approximation via Public Listings", *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, Nuremberg, Germany (2009).
97. Prehofer C and Bettstetter C, "Self-Organization in Communication Networks: Principles and Design Paradigms", 43:7 *Communications Magazine, IEEE* 78 (2005).

few decisions setting out the Internet's architecture, we currently have a system that has grown as a result of those initial decisions.[98]

This can result in the amazingly vast array of information described in the previous chapters but, as has been presented here, it can also lead to a deeply embedded structural fault line that allows the misuse of the Internet's architecture to allow the spread of viruses, the prevalence of zombie networks and the vulnerability of the entire system due to centrality issues. The nature of the system is one that favours openness, distribution and the free spread of information, using free here in both the freedom and economic meaning of the word. But that openness comes at the price of allowing spam, botnets, viruses, DoS attacks and a pervasive difficulty in regulating the system.

The structural problems with the Internet architecture have been known for quite a while. While the Web was built initially as a scalable and adaptable network, adding new protocols and more data to the existing architecture resulted in changes that did not fix the architecture, but simply latched on delivery systems on the existing protocols.[99] Not only that, the system was created originally with flexibility of content in mind; the design of the original protocols facilitated the development of a network that could fulfil military, academic and commercial objectives by making a minimum set of assumptions about the type of data that was shared within the network.[100] The network had one overarching design feature; it did not really care about the content of the data, as long as it got to the intended recipient. This lack of discriminatory filters in packet switching and information delivery are both a great advantage and a design disadvantage, as it makes it hard to remove undesirable data from the network. As Zittrain puts it:

> [I]f the Internet had been designed with security as its centerpiece, it would never have achieved the kind of success it was enjoying, even as early as 1988. The basic

98. Alderson D and Willinger W, "A Contrasting Look at Self-Organization in the Internet and Next-Generation Communication Networks", 43:7 *Communications Magazine, IEEE* 94 (2005).
99. Shenker S, "Fundamental Design Issues for the Future Internet", 13:7 *IEEE Journal on Selected Areas in Communications* 1176 (1995).
100. Clark D, "The Design Philosophy of the DARPA Internet Protocols", *Symposium Proceedings on Communications Architectures and Protocols*, Stanford, CA (1988).

> assumption of Internet protocol design and implementation was that people would be reasonable; to assume otherwise runs the risk of hobbling it in just the way the proprietary networks were hobbled. The cybersecurity problem defies easy solution, because any of the most obvious solutions to it will cauterize the essence of the Internet and the generative PC.[101]

So, we are stuck with an insecure global network where self-organisation has almost become a given. While it is clear that governments and regulators are attempting to exert some form of control in cyberspace, these efforts appear to be doomed to failure because of the very architectural composition of the network. Born as a distributed system, a network of such vastness cannot really be controlled efficiently, at least not in a way that will make cybercrime disappear entirely. Or can it? Would it be possible to re-invent the Internet and turn it into a system that is not self-organising? This may seem like a loaded question, but it is a real choice. Despite what some may think, emergence and self-organisation are not always present in complex systems. While self-organisation is simply the way in which a complex system can bring order to chaos, it is possible for a complex system to remain chaotic, or for highly-ordered systems to emerge as a result of top-down control, or other statistical circumstances.[102] In theory, a tightly controlled network could be organised in such a manner that self-organisation rarely occurs, or it results in a network that responds to network design choices. There is a small but vocal number of researchers that insist that the apparent self-organising nature of the Internet is the result of design decisions, and therefore future networks that are properly designed should display a directed self-organisation. Alderson and Willinger note that:

> If recent experience with the wired Internet is an indication, network self-organization in the form of management simplicity will be a critical objective, but will likely be the result of deliberate and well-designed protocols rather than a feature that emerges out of randomness.[103]

---

101. Zittrain, supra note 21, p.61.
102. Ball P, *Critical Mass: How One Thing Leads to Another*, London: Arrow Books (2004), p.301.
103. Alderson and Willinger, supra note 98, p.99.

If it is possible to reorganise cyberspace, then perhaps it might be desirable to forego the self-organising Internet, and replace it with a more centralised, closed and controlled version. Remove self-organisation, and you may have a global communications network that operates in a less chaotic manner. Is it time to reset the Internet? Internet 2.0, if you may.

The year 2010 will probably be seen in the future as a watershed moment in the history of the Internet because there are two very different models of how the global network will evolve. On the one hand, we have a vocal number of advocates for maintaining openness in the system despite its pitfalls. On the other hand, we have a number of businesses that are pushing for a more controlled and closed architecture where a few companies act as filtering gateways. In the first camp there is an array of "Web loyalists" comprising software engineers, bloggers, open source advocates and net neutrality proponents. In the other camp we encounter a puzzling coalition of governments and companies like Apple that want to create a closed Internet which relies less on the browser, and more on applications and clients that connect to the Internet in order to provide one specific function or service, and where users browse a limited version of the Web via pre-approved programs.

We are currently presented with two very different ideas about what the Internet should be, a choice that I call the dilemma of the open Web versus the closed Internet.[104] For the Web loyalists, the status quo should remain; the open Web has given us peer-production, blogs, social networks, free email and an amount of information that grows exponentially.[105] For the proponents of change, the current Internet is a bloated network filled with superfluous data, porn, viruses and all sorts of unsavoury material. The solution is to create a more centralised network where users access information through filtered channels that will offer a safer and cleaner environment. Precisely like a gated

104. Guadamuz A, "The Open Web vs the Closed Internet", *TechnoLlama Blog* (22 August, 2010), http://www.technollama.co.uk/open-web-vs-closed-internet.
105. Zhang G-Q et al, "Evolution of the Internet and its Cores", 10 *New Journal of Physics* 123027 (2008).

community. Steve Jobs is perhaps the most vocal proponent of this version of the future. His vision is one of a more free Web. In a now legendary email exchange with one of the editors of the Gawker blog, Jobs commented that his vision is one of a free Internet, but one free from all of the hassles of the current one:

> Yep, freedom from programs that steal your private data. Freedom from programs that trash your battery. Freedom from porn. Yep, freedom. The times they are a changing', and some traditional PC folks feel like their world is slipping away. It is.[106]

This statement needs some background, as it lies at the heart of the current debate about the future of the Internet. Apple has become the leading proponent of the Internet 2.0, a place that by the time you are reading this might already be prevalent. The World Wide Web is just a small part of what constitutes the Internet and it is the most visible aspect of the network – you connect to it via your browser, surf pages, watch videos and may even download content, legal or not. However, there is a competing version of the Web taking shape. As more people browse the Internet on their mobile phones, MP3 players, e-book readers and tablet computers (like Apple's iPad), the relevance of the WWW is diminished. For example, on my own Android mobile phone, I am constantly connecting to the Internet, but not to the browser-based Web. I connect to Twitter via an application (app) called Twidroid; I connect to Google Maps, and browse local events and places using an augmented reality browser called Layar. All of these are clients; they use the Internet, but not the HTML-based Web. This is not an isolated occurrence; as more and more users rely on their mobile devices to connect to the Internet, the app will become a more important part of our daily interaction with the global network. This has prompted influential thinkers such as Chris Anderson to declare the death of the Web.[107]

---

106. Email conversation reproduced in: Tate R, "Steve Jobs Offers World 'Freedom From Porn'", *Gawker: Valleywag Blog* (May 15, 2010), http://gawker.com/5539717/steve-jobs-offers-world-freedom-from-porn.
107. Anderson C and Wolff M, "The Web Is Dead. Long Live the Internet", 18.09 *Wired* (August 17, 2010), http://www.wired.com/magazine/2010/08/ff_webrip/all/1.

These apps have to be installed wilfully by the user, so only those apps that have been approved by the device maker can run on the device. This closes down the Internet, and also adds a layer of centrality that did not exist in the old Web. Undoubtedly, the Jobsian model of the closed and centralised Internet has certain appeal; it restricts self-organisation and it also allows regulators to monitor closely what application people are running. Controversial, illegal and even potentially liable programs will be filtered out in favour of bland, mass-appeal apps.

However, this version of the future has two main problems that stress the importance of network theory. The obvious one is that by concentrating application delivery into a few marketplaces, there is an added risk of creating a centralised network that would become a prime target for cybercriminal attacks. Just recently, hackers managed to hack into PayPal accounts and charge millions of dollars' worth of Apple iTunes content.[108] The second problem with the Jobsian Internet is that perhaps it underestimates the power of self-organisation. The current Internet has become very adept at propagating a specific model of communication delivery to the world. While there are those who tell us that this model has to change, we have to accept the possibility that the emergent nature of the Web cannot be contained any more. It is possible that the future of the Internet has already been written in the protocols that gave it life. Once again, similes involving genies, bottles, boxes and apple trees apply.

The future of the Internet may already be written.

108. Shiels M, "Web Scam Hits iTunes and Paypal Users", *BBC News* (24 August, 2010), http://www.bbc.co.uk/news/technology-11065301.

8. Conclusion

Everything in nature is the result of fixed laws.
Charles Darwin, *The Autobiography of Charles Darwin: 1809-1882*[1]

1. A TALE OF TWO INTERNETS

Two events in the last few months surrounding the completion of this work have served to set the stage of the opposing philosophies regarding Internet regulation. These I believe serve as a good illustration of the main conclusions to be drawn from this book.

On November 28 2010, the whistleblowing site Wikileaks began releasing some of the more than 250,000 diplomatic cables from USA embassies around the world, in a coordinated exercise with five major international newspapers,[2] but the bulk of the release was conducted through the Wikileaks website. The cables contained embarrassing details both to the United States and to various governments around the world, and in some cases, even some sensitive data that has sparked political unrest in various fronts.

From the very beginning, there were calls from numerous parties within the United States to try to shut down Wikileaks.[3] What followed was almost a textbook case study on Internet resilience, and just how difficult it is to police the Internet.

---

1. Darwin C, *The Autobiography of Charles Darwin: 1809–1882*, London: W. W. Norton & Company (1993).
2. Leigh D, "How 250,000 US Embassy Cables Were Leaked", *The Guardian* (28 November, 2010), http://goo.gl/Azhcm.
3. Sarah Palin wrote on Twitter that "Inexplicable: I recently won in court to stop my book "America by Heart" from being leaked, but US Govt can't stop Wikileaks 'treasonous act?'" http://twitter.com/SarahPalinUSA/status/9251635779866625.

To explain the regulatory attempts to shut down Wikileaks, it is important to remember some of the concepts seen in Chapter 4 about Internet architecture. If you wanted to reach Wikileaks with your Internet browser of choice (then identified as www.wikileaks.org), you had to know its address, or you could enter "wikileaks" into a search engine. The result would be that the Domain Name system would translate wikileaks.org into a computer IP address, and would direct your browser to the server hosting that content. The actual Wikileaks website was housed in several hosting services, mostly in Sweden and France, but they had also bought hosting space in the cloud computing web services offered by Amazon.com. The wikileaks.org domain name was assigned by Californian domain name registrar EveryDNS.net, which also provided free DNS services for the site.[4] By 1st December 2010, just a couple of days after the initial leaks, Amazon had dropped the service alleging breach of its Terms of Use, and EveryDNS.net had revoked the DNS registration alleging damage to its servers from coordinated cyber-attacks. By the end of that week, several payment systems which took donations for the Wikileaks (including Visa, MasterCard and PayPal) had also dropped the organisation. Bereft of hosting, routing and monetary channels, one would have thought that Wikileaks would simply be forced to disappear.[5] However, if there is one thing that we have learnt about the Internet from network science is that it is incredibly resilient.

In any other type of architecture, such a massive attack on the entire Wikileaks operational infrastructure would spell its demise. However, as it has been repeatedly stated throughout this book, there is something at which the Internet is really good at, it takes censorship as an attack to its infrastructure and reroutes services to avoid the affected area. Just a few minutes after Wikileaks had its DNS services removed the fact was advertised to the world via Twitter and Facebook.[6] Because the site was still being

4. Guadamuz A, "Wikileaks: So, This Is What Cyberwar Looks Like", *TechnoLlama* (3 December, 2010), http://goo.gl/fv2fm.
5. Ibid.
6. http://twitter.com/wikileaks and http://facebook.com/wikileaks respectively.

hosted in a computer connected to the Internet, it was still possible to access the content via an IP address despite the fact that writing Wikipedia.org into a browser would take you nowhere.[7] Similarly, several mirrors[8] and new DNS registrations started popping up everywhere – social media services were used to retransmit the latest IP address as they became available. Wikileaks even managed to get other domains.[9] Moreover, Wikileaks made available an encrypted torrent file through The Pirate Bay which allegedly contained all of the cables as a manner of online insurance against complete disconnection.[10] In short, this was Internet resilience at its best.

The so-called Cablegate incident made abundantly clear just how difficult it can be to regulate large distributed networks such as the Internet. One of the most important lessons taught by network science is precisely that a scale-free network is resilient in the extreme. Even large co-ordinated attacks are unlikely to bring down the entire network, particularly when knowledgeable and determined agents are working within the very same Internet architecture to spread information. The lack of centrality within the wider Internet makes it almost impossible to shut down a website such as Wikileaks. Evidence of this is that even after considerable public and private efforts were used to remove Wikileaks from the web, the site is still running at the time of writing.

7. For example, the site could be accessed during December 2010 at http://213.251.145.96 and http://46.59.1.2, amongst other sites.
8. In Internet architecture terms, a mirror is an exact copy of another site.
9. For example, http://wikileaks.ch and http://wikileaks.info.
10. http://thepiratebay.org/torrent/5723136/WikiLeaks_insurance.

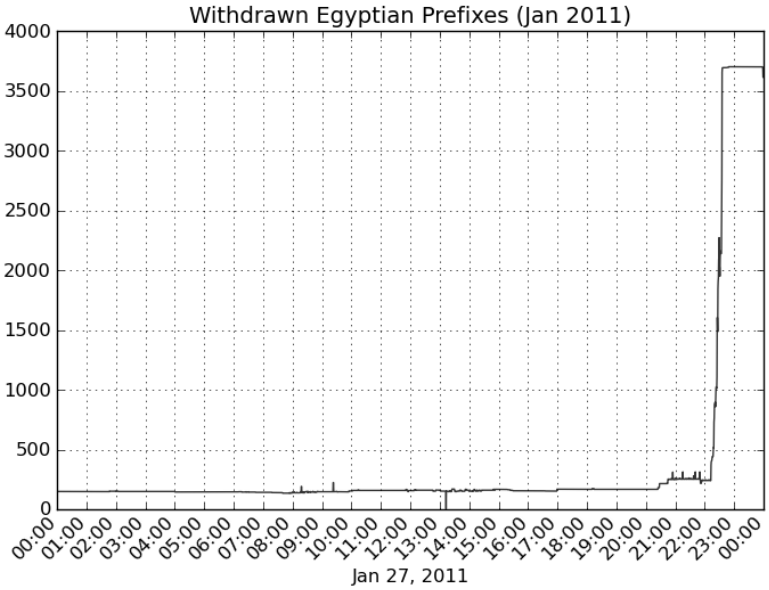


Figure 8.1 The Egyptian Internet shuts down[11]

Here is where the second story comes to play. Right after the Wikileaks Cablegate scandal (and some have even suggested that because of it), the Arab world erupted in civil unrest. From Tunisia to Yemen, populations across the region began a series of street protests that resulted in the fall of several regimes. When the conflict reached Egypt in January 2011, a large part of the protests were coordinated using the Internet, particularly through the use of social media sites.[12] It may be too much to suggest that the Internet caused the revolution in Egypt, but it certainly helped protesters to organise and stay ahead of the authorities. It also was vital in mobilising large numbers of people to specific locations, such as Tahrir Square in Cairo.[13] What is certain is that the Egyptian government considered that the Internet posed a threat to their interests, so

[11] Source: http://www.renesys.com/blog/2011/01/egypt-leaves-the-internet.shtml.

12. Herrera L, "Egypt's Revolution 2.0: The Facebook Factor", *Jadaliyya* (12 February, 2011), http://goo.gl/qlkEd.
13. Gustin S, "Social Media Sparked, Accelerated Egypt's Revolutionary Fire", *Wired* (11 February, 2011), http://goo.gl/bz6J2.

they did something that had never done before to such extent: they shut down the Internet.

On January 27 2011, at around 10.30 GMT, the entire Egyptian Internet was disconnected from the rest of the world.[14] This was possible because Egypt, just as many other countries in the Middle East, has a national firewall consisting of an extra layer of Internet servers that intermediate all traffic in and out of the country through servers running the adequately named Border Gateway Protocol (BGP). Egyptian authorities managed to shut down simultaneously 3,500 BGP routes into the country, which meant that more than 90 percent of all traffic in and out of the country could not get through (Figure 8.1).[15]

What the Egyptian case illustrates is an excellent example of the dual nature of Internet architecture. At the larger scale, the Web is a scale-free network, entirely distributed and remarkably robust. At the national level, the Internet is increasingly centralised, and therefore more likely to suffer from large cascading local failures. The more centralised the system, the easier it is to regulate.

This is therefore the conundrum currently presented to regulators around the world, and all is consistent with the empirical and theoretical evidence provided by network theory. It is possible to control the Internet, but to do so it must stop being decentralised. Higher levels of centrality allow for more control, but this in change translates into a less open system.

2. SELF-ORGANISATION THEORY OF INTERNET REGULATION

When dealing with the subject of Internet regulation in Chapter 4, a hypothesis was presented, this forms the central part of this work. Assuming that the Internet is a

14. Williams C, "How Egypt Shut Down the Internet", *The Telegraph* (28 January, 2011), http://goo.gl/j5PTU.
15. Greenemeier L, "How Was Egypt's Internet Access Shut Off?" *Scientific American* (28 January, 2011), http://goo.gl/CCE08.

complex adaptive system subject to self-organisation, then it is possible to postulate that any attempt to regulate specific elements within the network will have to take into account this important emergent attribute of the global communication system. Moreover, it is the main assertion of the present book that it is not possible to adequately regulate online environments that display self-organising characteristics without some knowledge of the empirical and theoretical features of such environments.

Throughout the work, several examples have been presented as evidence that the Internet is indeed a self-organising system. The network is made up of nodes and links that grow according to power laws. Older links in the network accumulate more links, and those successful nodes in turn tend to accumulate more links themselves, creating a "rich get richer" situation. The resulting hubs serve as important connectors within the network, which explain in turn the seemingly ordered nature of the system. The nodes themselves often cluster into small world networks where the intervening pathways between nodes tend to be short.

The scale-free nature of the network makes the Internet resilient to random attacks. However, this also means that other undesired networks which exist within cyberspace are also robust, such as P2P file-sharing networks, or cybercrime rings. Similarly, because of architectural decisions early on, the network displays high levels of centrality at the national scale.

All of these features, amongst others, offer strong confirmation that there are self-organising forces online. Any regulatory effort that ignores this fact is faced with severe difficulties, as the same self-organising forces that shape the Internet's architecture are also at work to undermine and even defeat regulatory action.

When presented with autopoietic systems, regulation theories have two possible strategies. One could accept that the network responds to its own self-organising elements, and therefore cannot be governed. If this is the case, then regulation is not possible. This work has adopted the opposite view, that self-regulation need not mean that governance of the system is impossible. While this may be optimistic, it is the only

viable avenue to take if one is willing to undertake regulatory efforts. Not to do this would be to fall prey to an anarchic and/or libertarian view of governance, where everything is left to the self-organising powers of the system. Even in the face of contradictory evidence we will adopt the optimistic view of regulation, and will assume that some form of order outside of the regulatory effort is possible.

Within the optimistic regulatory philosophy, we could try to build the system to fit the regulatory goals. Following the idea presented in Lessig's *Code*,[16] regulation strategies can be built into the system assuming that this will seed the elements around which self-organisation will occur. Complex systems will usually order themselves at fitness peaks of higher order. If we know how self-organisation works within the network, then we can try to code situations that will constitute fitness peaks in the overall landscape.

There are two examples presented in the work that can represent opportunities for engineered self-organisation. Firstly, in the fight against P2P file-sharing, it seems evident that the networks are robust self-organising entities. But what would happen if one built network architecture that specifically targets such networks? While there have been some attempts to attack the networks in this manner, perhaps more strict legislation that tackles not the infringers, but the architecture, would have more chance of success. Secondly, some forms of cybercrime rely heavily on the current open and centralised Internet architecture. A more tightly regulated network, with more gateways and intermediaries, may sacrifice the Web's dynamic nature, but it may seriously hinder some forms of cybercrime, particularly denial of service attacks, spam and phishing.

The optimistic view of regulation also presents opportunities for smarter regulatory efforts by informing decision-makers and stakeholders about the way in which the target system operates. Any attempt to legislate in the areas covered by Internet regulation, such as privacy, copyright and cybercrime, has to consider the emergent traits of cyberspace. At some point policymakers will realise that their regulatory efforts are

16. Lessig L, *Code Version 2.0*, 2nd ed, New York: Basic Books (2006).

having no effect, and hopefully they will look at some of the research highlighted in this work in search of evidence.

I am aware that this may sound arrogant; it is not my intention to be the eccentric person in the street holding a placard stating that "The End Is Nigh". There are many scholars who are already looking at the theories of complexity for possible answers to regulatory conundrums. The goal of this work is to point interested readers to the empirical studies that may explain regulatory failure.

To recap, the self-organisation theory of Internet regulation therefore is as follows: the Internet is a complex system that displays self-organisation. In order to efficiently and successfully regulate the digital environment, it is imperative that one understands how it is organised, what characteristics are present, what elements act as fitness peaks and how architectural decisions affect its emergent features.

One of my personal heroes is Edward Tufte. In his seminal book *The Cognitive Style of Powerpoint*,[17] he commented that "Bullet outlines dilute thought". I was tempted to make an outline list of the various salient points of the theory, but I believe that the above explanation should suffice without having to dilute the conclusion.

3. FUTURE RESEARCH

This work has used three fields of Internet regulation as the case studies of the theory of complexity presented above. These include topics such as copyright policy, online copyright infringement, free and open source software, user-generated content and cybercrime. These were chosen for two reasons. Firstly, by reading through the literature on complex theory and network science, it became clear that these fields were more developed, and where it would be possible to obtain more supporting evidence for the ideas of self-organisation that are at the centre of the work. Secondly, these have been

17. Tufte ER, *The Cognitive Style of Powerpoint: Pitching out Corrupts Within*, 2nd ed, Cheshire, CT: Graphics Press (2006).

some of my main areas of research for the past eight years. It was hoped that the familiarity with the legal topic would allow me to make stronger connection to the wealth of research into networks that has been highlighted in the work. These three main case studies, however, are just some of the various areas of Internet law that could be subject to similar analysis.

There are three potential topics where I believe future research could be conducted. The first, and perhaps more obvious, is the subject of online privacy within social networks. At the time of writing, Facebook boasts 500 million active users.[18] At the same time as this figure was reached, Facebook was immersed in several privacy scandals about what it does with the information collected on its users.[19] Almost by definition, such systems are practically tailor-made for disciplines such as social network analysis. Therefore, network theory could try to look at some questions about privacy concerns. What constitutes a user's closest social network? How much of the information made available to "friends" can be mined for other purposes? Is it possible to create a social network where privacy concerns are minimised?

Another possible topic ripe for analysis is network neutrality. According to Marsden:

> In short, net neutrality is about the rules of the road for Internet users, and about the relationship between the owners of those roads and the users. Government is asked to make a decision as to which users have priority and whether road charging should be introduced, ostensibly to build wider and faster roads in future.[20]

This is a highly-politicised and controversial topic, particularly in the United States, where the choice of network provider in rural areas is limited. The current debate hinges on the question of whether ISPs and bandwidth providers should charge for higher speeds. Currently, the Internet rests on the assumption that all packets are created equal.

---

18. Zuckerberg M, "500 Million Stories", *The Facebook Blog* (21 July, 2010), http://blog.facebook.com/blog.php?post=409753352130.
19. Quigley R, "Facebook Privacy Fears for 100m Users as Their Personal Details Are Published on File-Sharing Site", *Mail Online* (29 July, 2010), http://is.gd/eTGqX.
20. Marsden CT, *Net Neutrality: Towards a Co-Regulatory Solution,* London: Bloomsbury Academic (2010), pp. 2–3.

By favouring some content over another based on price, this model would be under threat. It is therefore easy to see why a network theory analysis would be favourable in this area. Would a change in the architectural structure of bandwidth provision affect the network as a whole? Is network neutrality possible or will different packet speeds emerge as a fitness solution in the system?

The other topic is that of intermediary liability. While this topic was covered when discussing copyright infringement, this is a much richer legal subject that involves areas such as defamation, electronic commerce and pornography.[21] Because ISP liability deals mostly with the distribution of content through the network, the topic lends itself to network analysis because content placed online has incoming and outgoing links. It might be possible to try to analyse the centrality of content within a network to ascertain potential damages, and also to try to identify social networks and replication pathways within the system.

These are only three examples of the various subjects that could be analysed in light of the study of the architecture of networks. It is hoped that the present work will inspire fellow Internet Law colleagues to look into some of the tools described in previous chapters.

4. A FINAL WORD ON REDUCTIONISM

One of the most neglected stories of the 2008 global credit crunch has been the partial responsibility of physicists and mathematicians in creating the crisis. In 1999, the specialised magazine *Physics World* ran an editorial commenting on the growing phenomenon of physicists leaving academia to become quantitative analysts, better

21. See for example: Goldstein MP, "Service Provider Liability for Acts Committed by Users: What You Don't Know Can Hurt You", 18:3 *The John Marshall Journal of Computer & Information Law* 52 (2000).

known in the industry as "quants".[22] The reason for this migration was twofold. Firstly, mathematicians and physicists were hired by financial services to provide models to make sense of the chaotic nature of markets and share prices. Secondly, several physicists were hired to come up with mathematical models and software programs that would be at the heart of complex financial instruments called derivatives, which many have blamed as one of the causes of the credit crunch.[23]

Derivatives are investment packages that do not have inherent value; their relative worth is tied to the value of other trade items, including shares, currencies, commodities and even aggregated credit packages (hence the name). As these instruments depend entirely on the linked tradable goods, they tend to be exceptionally complex, so much so that only a few mathematicians and physicists were said to understand them. Despite their seeming complexity, derivatives became highly sought after because they produced high returns for the initial investment. The problem seems to be that the instruments were so complex that nobody actually understood them, and therefore an entire market rested on the assumption that some people actually knew what they were doing, when they almost certainly did not. Derivatives rose in value far higher than they were actually worth, and they were often tied with large insurance-like packages called credit default swaps. When the faulty nature of the packages was unearthed due to the collapse of several credit schemes, the pyramid-like house of cards inevitably tumbled down.

The results of this staggering display of hubris is well known, and at the time of writing global markets are still reeling from the acts of folly displayed by financiers and bankers. With the benefit of hindsight, it seems clear that entrusting mathematicians with the keys of the City and Wall Street placed too much faith on the exactitude of maths in detriment of the unpredictability of human nature.

---

22. "'Rocket Science': The Facts", *Physics World* (June 3 1999), http://physicsworld.com/cws/article/print/1081.
23. Even back in 2003, financier Warren Buffet had called derivatives "financial weapons of mass destruction". See: "Buffett Warns on Investment 'Time Bomb'" *BBC News* (4 March, 2003), http://news.bbc.co.uk/1/hi/business/2817995.stm.

The reason why I highlight this case is to serve as a word of caution about the reach of the theories expounded in this work. While it is true that it is assumed that mathematics and physics do have something to tell us about social systems, one should never lose sight of the fact that it is possible to go too far in this approach. It is not my wish to replace Internet regulation theories with mechanistic network analysis that is only interested in charting nodes and links into logarithmic tables and pretty visualisations of networks. The data tells us a part of the story; what we decide to make of the information is decidedly our own responsibility.

The application of complexity theory research described throughout this work may generate unease amongst some readers. This may be because the use of physical formula to understand human behaviour has had a mixed history, as was explained in the Introduction. The implication of such a deterministic outlook of the world has had negative implications, so it is usually suspected by default. But despite its dubious history, modern physics has been demonstrating that there could be an application of physical models to social interactions.[24] Formulas used to describe how magnets achieve their orientation, or how gases condense, can also be used to chart how businesses grow, how crime rates fluctuate, or how crowds flow.[25]

It would be easy to dismiss the trends cited, and particular the emerging science of networks, as another doomed attempt to explain social complexity with mathematics, or a way of deleting free will to convert the human experience into a set of equations. However, to view power laws as deterministic does not really address the fact that this is not an exact science; it is a descriptive tool of how networks operate.[26] Humans still retain free agency, while the network itself could be deterministic and react in predicable ways.

---

24. Ball P, "The Physical Modeling of Human Social Systems", 1 *ComPlexUs* 190 (2003).
25. Ibid, pp.198–200.
26. It must be stressed that the term "deterministic nature" has other implications in the research. It is another mathematical model to describe network growth. See: Barabási A-L, Ravasz E and Vicsek T, "Deterministic Scale-Free Networks", 299(3) *Physica A* 559–564 (2001).

The best way to understand the potential deterministic nature of networks is to conduct simple thought experiments about how people actually interact with one another in a social gathering. We would generally like to think that we are free agents, and therefore social networks should respond to the very random nature of human experience. Yet, we are constantly responding and acting according to physical and social constraints. Imagine that you are at a busy conference coffee break. If you are observant, you will probably notice that people have gathered in small groups, some people will work the room while others will remain with the same group, and perhaps there may be a person standing by the coffee table on their own. You will rarely see a person shouting across the room, or an extremely large group where nobody can interact. If you map the number of links made during such breaks, you will start to see certain patterns emerging. These patterns are not deterministic in the sense that they completely erase agency from those present; you can still choose to move around the room, or not to talk to anyone else, but the pattern made by the collection of conducts provides a good example of the apparently deterministic nature of social networks. People will act freely, but the constraints of social norms and the laws of physics will mean that social networks will produce certain results. Smaller groups will have less deterministic value because the action of one individual will have a larger effect, while the larger group will tend to absorb the random individual behaviour.

This same phenomenon is precisely what has been mapped by the research conducted so far on all sorts of networks. Large scale-free networks seem to follow certain rules that respond to those same physical constrains. Meaningful links, nodes and hubs serve to explain the larger picture, but not the individual choices.

It is only natural that grand theories of everything should be met with scepticism. Attempting to explain complex systems with a few theories may seem like unforgivable reductionism; an attempt to apply materialistic ideals to social relations where they do not fit. However, if there is sound evidence that certain network environments like the Internet act in predictable ways, then all the research into this behaviour should be taken into consideration when attempting to analyse the underlying trends that govern such

patterns, even if it is an analysis that belongs to the physical sciences and not to the social ones.

It can be argued that we are on the threshold of better understanding complex systems like the Web thanks to the predictable nature of the science of networks, but it is important to make sure that such enthusiasm is tempered by the scale of the task of mapping such large structures. All predictive models of cyberspace should take into consideration that it is a changing environment. As Barabási argues:

> It is far from us to suggest that the scale-free model introduced above describes faithfully the topology of the www. Naturally, the www has a much richer structure that cannot be captured by such simple ingredients. For example, the links are not invariant in time, they constantly change, being either eliminated or rewired to other documents. Similarly, the www documents are not stable, they are often removed, and change address. Furthermore, the web pages are structured in domains, that by themselves have a rather complex hierarchical structure.[27]

Research into networks should then be released with the caveat that the descriptive and predictive features given to power laws are to be taken as tools, not as absolute predictions. This has to be stressed because it would be plausible to read the extensive research presented so far and complain that we are talking about a form of technological determinism.[28] The reader can rest assured that such a goal is not intended, and that the tools put forward should not be construed as deterministic in any way, just like gravity is not deterministic.

This work originated from one paper presented in various stages to diverse audiences.[29] The last slide of the presentation features a wonderful picture of the footpaths in a public park in Stuttgart University that was shown at the end of Chapter 3. The picture shows the designed path by whoever built the space, a stylish crossing X

---

27. Barabási A-L, Albert R and Jeong H, "Scale-Free Characteristics of Random Networks: The Topology of the World Wide Web", 281 *Physica A* 69–77 (2000), p.75.
28. For more about the topic of technological determinism, see: Smith MR and Marx L, *Does Technology Drive History? The Dilemma of Technological Determinism*, Cambridge, MA: London: MIT Press (1994).
29. To mixed results, ranging from outraged to enthusiastic.

through the roughly square lawn. However, one can clearly see another path in the picture, one that was not designed. This path has been made by people walking from one building to another in a direct line, which does not follow the official pathway. This exemplifies nicely what the present work is trying to achieve. We may plot paths through cyberspace; we may attempt to regulate the space in various ways. But is this regulation really considering the paths that will be chosen almost inevitably by the inhabitants of the new space? Network science provides a descriptive tool to make better decisions when building the paths.

Critical Review

ABSTRACT

Complexity theory as a subject has gained increasing prominence across numerous disciplines including physics, biology, sociology and economics. Large interconnected systems such as the Internet display a number of inherent architectural characteristics deeming them well-suited to the study of complex dynamic networks. The book uses various network science-based tools to explore the contentious issue of Internet regulation.

The book demonstrates that the Internet as a global communications space is a self-organising entity that has proven problematic for regulators, and that in order to regulate cyberspace, one must first understand how the network operates. In order to illustrate how the WWW operates, the author presents case studies in copyright policy, peer-production and cybercrime, providing in-depth analyses of the challenges posed by the Internet's complex dynamic networks. The book concludes that regulatory efforts that ignore empirical evidence will ultimately encounter serious problems.

The book introduces network theory to legal audiences and applies some of the characteristics of large distributed self-organising networks to the topic of Internet regulation. As such, this fascinating book will prove invaluable to researchers, academics and students in the fields of Internet regulation and policy, intellectual property law and information technology law.

1. SUMMARY

One effect of the emergence of the vast online environment that we know as the internet is that it has pushed forward research into networks. The science of networks is an important field of Mathematics that charts the emergence and characteristics of networks, and it also offers some understanding of the behaviour of the various links and hubs within a network. Network science has implications to a large number of disciplines, as network structures can be used to describe things like cities, brains, and the economy. The physics of networks has been largely descriptive, but thanks to the Internet, many assumptions that it makes have been tested. Although the web is vast, its growth and reach allows researchers to map and test several previously untested ideas about how networks interact. With the Web, we now have the tools to test the organisational structures of networks, their architecture, their growth, and even allows predictions about their behaviour, strengths and vulnerabilities.

Complexity theory is a subset of Network science, and as such it is a subject that has been gaining prominence in various disciplines, including physics, biology, sociology and economics. Large interconnected systems such as the Internet display a number of inherent architectural characteristics that make them well-suited to the study of complex dynamic networks. The Internet as a global communications space is a self-organising entity that has proved to be problematic for regulators. This book uses various analytical tools found in network science and complexity theory and applies them to the subject of Internet regulation, arguing that in order to regulate Cyberspace; one must also understand how the network operates.

The scope of the book therefore is both theoretical and practical. While a modest objective of the book is to serve as an introduction to the wider legal audience to some of the theories of complexity and networks, the main objective is more ambitious in scope. By looking at the application of complexity theory and network science in various areas of Internet regulation, namely copyright infringement, peer-production, and cybercrime, the book tries to provide enough evidence to postulate a theory of

Internet regulation based on network science. This theory is twofold. First, the theory states that the Internet is a self-organising system, and as such, any regulatory strategy must take this into consideration, otherwise any solution to perceived and actual problems is doomed to failure.

This work advances the knowledge in the subject in one important way. As far as I am aware, no other work has attempted to create a regulation theory of complex networks in the way this book is doing. As stated, the book treats the Internet as a complex system that displays self-organising properties. In order to efficiently and successfully regulate the digital environment, it is imperative that one understands how it is organised, what characteristics are present, what elements act as self-organising elements, and how architectural decisions affect its emergent features. Moreover, the book offers an important step in the ongoing debate with regards to the nature and future direction of Internet regulation.

2. AIMS AND OBJECTIVES

The main aims and objectives of the work are to:

- Analyse whether complexity theory is useful within the context of jurisprudence.
- Propose a legal theory of complex systems.
- Explore how network theory can help understand better specific legal topics.
- Look at the role of complexity in the field of Internet regulation with the aim of acquiring empirical evidence as to how networks operate. This can then be used by policymakers to produce better informed policy in various fields.
- Contribute to the ongoing debate about the nature of regulation of the Internet, namely, whether to push for an open or closed Internet.
- Introduce to legal audiences to network and complexity theories

3. METHODOLOGY

Trying to define a methodology for this work has been a challenge, particularly because it is trying to bridge different areas of study that traditionally do not interact with one another, namely studies from the physical sciences, and the law.

In the introductory section of the book, some space is spent describing the methodological separation that has occurred between the physical and social sciences. The split has made it more difficult to use studies from say, Physics and Mathematics, in the law. It is postulated in the book that this is a recent event, and that the development of disciplines that have social relevance should be translated into legal studies.

The methodology from the sciences has been a straightforward literature review[1] of the prevalent research from the highest impact journals. There has been a considerable amount of secondary work conducted on the areas of network theory and complexity, so having access to the most cited and important works becomes easier when specialists in the field put together article collections, reading lists, and websites detailing the best research out there.

Similarly, the legal part of the work has also been following familiar paths, mostly by using legal research methodology, namely looking at case law and legislation where it exists, but mostly it has been doctrinal in nature.[2]

Putting both methods together is slightly more difficult, but it is not completely unheard of.[3] There is a growing tradition of empirical legal research,[4] in which the present work is inspired. Although with a few exceptions the book does not conduct its own empirical

---

[1] Hart C, *Doing a Literature Review: Releasing the Social Science Research Imagination*", London: SAGE (1998).
[2] Kumar R, *Research Methodology: A Step-By-Step Guide for Beginners*, London, SAGE (2010).
[3] Walker L, "Social facts: Scientific methodology as legal precedent", 76:4 *California Law Review* 877 (1988).
[4] Revesz R L, "A Defense of Empirical Legal Scholarship", 69:1 *The University of Chicago Law Review* 169 (2002).

research as such, many of the tools to analyse empirical results are already in place, and therefore this is the preferred solution.

3.1. Future research

The work conducted in the book and in this short paper is a starting point, and there is a considerable potential in conducting further research in the area to further exploit the power of the study of complex systems into the Internet regulation arena in particular, and legal scholarship in general.

The empirical nature of future studies would be conducted building on the body of work that has been presented in the book and here. This would be done by looking for adequate datasets which describe Internet architecture in detail, with emphasis on data which can inform a study about vulnerability, centrality, cascading failures and unintended downstream effects. For that purpose we have identified datasets with rich Internet topology data. These include:

- The CAIDA dataset of Distributed Denial of Service (DDoS) attacks.[5] This looks interesting and is relevant for future analysis because it may help to determine possible vulnerabilities in centralised systems. I am particularly interested in trying to ascertain if the Internet's Root Servers could be vulnerable to some sort of DDoS attack. As stipulated above, I believe that this is not the case, but I would like to have empirical data to prove this assertion.
- The DIMES data on Internet topology.[6] This is a very interesting project that uses software agents installed in computers around the world and sends pings and traceroutes commands to them and back again, sort of conducting a sonar map of the Internet. These give an accurate idea of the way in which information moves around the Web.

[5] http://www.caida.org.
[6] http://www.netdimes.org/new/?q=node/54.

- PREDICT, the Protected Repository for the Defense of Infrastructure Against Cyber Threats,[7] seems to have some really interesting datasets about Internet topology and IP packet headers, which could be useful to map virus attacks and other types of vulnerability within the system.

This is a part of future studies that will require more assistance from colleagues that are more knowledgeable in empirical research. While it is possible at the moment to have access to the datasets, and use many of the available open source software to perform analysis on the data, I still feel that in order to conduct deeper studies any future work could profit from having direct contact with people who are specialists in the subject.

4. CONCLUSIONS

Assuming that the Internet is a complex adaptive system subject to self-organisation, then it is possible to postulate that any attempt to regulate specific elements within the network will have to take into account this important emergent attribute of the global communication system. Moreover, it is the main assertion of the book that it is not possible to adequately regulate online environments that display self-organising characteristics without some knowledge of the empirical and theoretical features of such environments.

Throughout the book several examples are presented as evidence that the Internet is indeed a self-organising system. The network is made up of nodes and links that grow according to power laws.[8] Older links in the network accumulate more links, and those successful nodes in turn tend to accumulate more links themselves, creating a "rich get

[7] https://www.predict.org/.

[8] A power law is a mathematical expression that happens "when the probability of measuring a particular value of some quantity varies inversely as a power of that value". See: Newman MEJ, "Power Laws, Pareto Distributions and Zipf's Law", 46:5 *Contemporary Physics* 323 (2005), p.323.

richer" situation. The resulting hubs serve as important connectors within the network, which explain in turn the seemingly ordered nature of the system. The nodes themselves often cluster into small world networks where the intervening pathways between nodes tend to be short. The network is fractal in nature, in other words, it has the same architectural features be it at large or short scale, hence the definition that it is scale-free.

The scale-free nature of the network makes the Internet resilient to random attacks. However, this also means that other undesired networks which exist within cyberspace are also robust, such as P2P file-sharing networks, or cybercrime rings. Similarly, because of architectural decisions early on, the network displays high levels of centrality at the national scale.

All of these features, amongst others, offer strong confirmation that there are self-organising forces online. Any regulatory effort that ignores this fact is faced with severe difficulties, as the same self-organising forces that shape the Internet's architecture are also at work to undermine and even defeat regulatory action.

The father of self-organisation studies in social systems is Niklas Luhmann with his influential theory of autopoiesis[9]. In its broadest sense, Luhmann's theory of autopoiesis matches what we have witnessed online, as he defines it as social systems that respond to internal stimuli instead of relying on external elements; these elements come together to generate stability in the system. It is a common misunderstanding that self-organisation causes chaos,[10] when in reality it most chaotic systems tend to stability in the long run, much in line with what is known as fitness landscapes.[11] If we think of the Internet as an autopoietic system, then we should conclude that it becomes organised because of the interaction of its parts favours clustering and stability in order to manage complexity.

---

[9] Literally meaning self-creation. Luhmann N, *Social Systems,* Stanford, CA: Stanford University Press (1995), p.22.

[10] Chaos in the strict mathematical sense, meaning that rendering long-term prediction is impossible in general. See: Alligood KT, *Chaos: An Introduction to Dynamical Systems*, New York: Springer-Verlag (1997).

[11] Kauffman SA and Weinberger EW, "The NK model of rugged fitness landscapes and its application to maturation of the immune response", 141:2 *Journal of Theoretical Biology* 211 (1989).

When presented with autopoietic systems, regulation theories have two possible strategies. One could accept that the network responds to its own self-organising elements, and therefore cannot be governed. If this is the case, then regulation is not possible. The book has adopted the opposite view, that self-regulation need not mean that governance of the system is impossible. While this may be optimistic, it is the only viable avenue to take if one is willing to undertake regulatory efforts. Not to do this would be to fall prey to an anarchic and/or libertarian view of governance, where everything is left to the self-organising powers of the system. Even in the face of contradictory evidence we will adopt the optimistic view of regulation, and will assume that some form of order outside of the regulatory effort is possible.

Within the optimistic regulatory philosophy, we could try to build the system to fit the regulatory goals. Following the idea presented in Lessig's *Code*,[12] regulation strategies can be built into the system assuming that this will seed the elements around which self-organisation will occur. As stated, complex systems will usually order themselves at fitness peaks of higher order. If we know how self-organisation works within the network, then we can try to code situations that will constitute fitness peaks in the overall landscape.

There are two examples presented in the book that can represent opportunities for engineered self-organisation. Firstly, in the fight against P2P file-sharing, it seems evident that the networks are robust self-organising entities. But what would happen if one built network architecture that specifically targets such networks? While there have been some attempts to attack the networks in this manner, perhaps more strict legislation that tackles not the infringers, but the architecture, would have more chance of success. Secondly, some forms of cybercrime rely heavily on the current open and centralised Internet architecture. A more tightly regulated network, with more gateways and intermediaries, may sacrifice the Web's dynamic nature, but it may seriously hinder some forms of cybercrime, particularly denial of service attacks, spam and phishing.

---

[12] Lessig L, *Code Version 2.0*, 2nd ed, New York: Basic Books (2006).

The optimistic view of regulation also presents opportunities for smarter regulatory efforts by informing decision-makers and stakeholders about the way in which the target system operates. Any attempt to legislate in the areas covered by Internet regulation, such as privacy, copyright and cybercrime, has to consider the emergent traits of cyberspace. At some point policymakers will realise that their regulatory efforts are having no effect, and hopefully they will look at some of the research highlighted in this work in search of evidence.

To recap, the self-organisation theory of Internet regulation therefore is as follows: the Internet is a complex system that displays self-organisation. In order to efficiently and successfully regulate the digital environment, it is imperative that one understands how it is organised, what characteristics are present, what elements act as fitness peaks and how architectural decisions affect its emergent features.

5. CONTRIBUTION

One of the main ideas with this book is to serve as an interface between legal research and network theory. This may seem like a mundane and even undemanding task, but I am of the strong opinion that this is of vital importance because the law increasingly has to deal with vast and complex networked environments. The structure of the book reflects what we know from Systems Theory[13] that communication between different systems of thought is possible, but difficult and often requires a translation of ideas from one system into a vocabulary that makes sense of it in terms of the normative structure of the recipient system. This can often introduced systematic and necessary "errors in translation", as the recipient system tries to approximate the ideas from the sender system. Therefore, an interface is necessary, a concept analogous to that of computer interfaces.

---

[13] The theory that tries to create a systematic body of principles applicable to all fields of research. See: Weinberg G*, An Introduction to General Systems Thinking*, New York, NY: Dorset House, 1975 (2001).

The main contribution of this work, therefore, is to both inform and educate the legal scholar to an important new area of study, and to offer several examples of the possible application of analytical network tools to legal practice. The work is therefore a "Grenzstellen", literally meaning, “border crossings”.[14]

A legal interface into network theory is needed at this time more than ever. Network theory makes several conclusions and predictions that arise from empirical research and theoretical analysis. These have been ignored somehow by the legal research mainstream (with various exceptions that are described in Chapter 3 of the book). Much of the current interest in networks can be traced back to a series of popular science books dedicated to publicising the latest developments in this area of research. Titles of note are *Linked* by Albert-Laszlo Barabási,[15] *The Tipping Point* by Malcom Gladwell,[16] *Critical Mass* by Philip Ball[17] and *Six Degrees* by Duncan J Watts.[18] These “pop science” credentials could make those unfamiliar with the literature suspicious about the validity and reliability of network theories, but this scepticism would be misplaced, as most of these books have sound peer-reviewed research behind them, and in most instances they have been written by the primary investigators themselves. The book goes through the main theories, and applies them in specific areas of Internet regulation studies. This is where the relevance of the work can be found.

To measure this contribution, I will first concentrate on stressing the importance of evidence-based policymaking. After that has been established, we will apply some of the theories and network tools discussed in the book to events relevant to Internet regulation that have taken place near or after the book’s completion, and therefore were not included in the work. These include privacy, centrality and resilience examples. These will be offered to try to stress the viability of the tools described in the book in various

---

[14] Teubner G, *Autopoietic Law: A New Approach To Law And Society*, Berlin: Walter de Gruyter (1988).
[15] Barabási A-L, *Linked: The New Science of Networks*, Cambridge MA: Perseus Pub. (2002).
[16] Gladwell M, *The Tipping Point: How Little Things Can Make a Big Difference*, London: Abacus (2002).
[17] Ball P, *Critical Mass: How One Thing Leads to Another*, London: Arrow Books (2004).
[18] Watts DJ, *Six Degrees: The Science of a Connected Age*, London: Vintage (2004).

areas of Internet-related policy issues. Finally, we will place this in the wider context of the current debate on the direction of Internet regulation.

5.1 The importance of evidence-based policy-making

There may be need to stress the importance of having strict and effective policy-making in Internet regulation. The following examples offer cases where a healthy dose of evidence might have been useful.

In March 1996, the European Union adopted the Directive 96/9/EC on the legal protection of databases (Database Directive),[19] which created a new sui generis right protecting against unauthorised extraction of information contained in databases. The stated goal of the Directive was to harmonise national practice in this topic in order to protect investment and to foster the creation of a European database market. This was done because, according to the opening paragraphs of the Directive:

> "[A]t present a very great imbalance in the level of investment in the database sector both as between the Member States and between the Community and the world's largest database-producing third countries".

The response to this perceived imbalance was to create a hitherto inexistent right with practically no studies explaining the need to undertake this action, and with no evidence whatsoever that it was needed and/or required. It was assumed by the framers of the Database Directive that by creating a new type of protection, database designers and creators would somehow flock to the EU countries, giving European businesses a competitive advantage in this new market. The problem was that this action was undertaken almost entirely on faith. Professor James Boyle famously described the process of enacting this new right like this:

> "Imagine a process of reviewing prescription drugs which goes like this: representatives from the drug company come to the regulators and argue that their

[19] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJL 077 27/03/1996.

> drug works well and should be approved. They have no evidence of this beyond a few anecdotes about people who want to take it and perhaps some very simple models of how the drug might affect the human body. The drug is approved. No trials, no empirical evidence of any kind, no follow-up."[20]

Needless to say, it works out that enacting new legislation with such flimsy premises did not produce the desired effects. To its credit, the European Commission conducted a review of the impact of the new right, and found that it had no effect whatsoever in fostering the creation of a new sector in the European economy. In 1996, the United States had the largest share of the global database market, with 56%, while European share was 22%. While this share increased between 1996 and 2001, it had dropped again to 24% by 2004, while the U.S. share went back to its previous levels.[21] This is strong indication that the sui generis right did not have any noticeable effect in strengthening the European database market. In an indicting comment on policy based on lobbying and guesswork, the Commission's report said:

> "Nevertheless, as the figures discussed below demonstrate, there has been a considerable growth in database production in the US, whereas, in the EU, the introduction of "sui generis" protection appears to have had the opposite effect. With respect to "non-original" databases, the assumption that more and more layers of IP protection means more innovation and growth appears not to hold up."[22]

Another example has been the debate over the term extension for performers of sound recorders in Europe. For years there has been a discrepancy in terms of protection for performers in sound recordings between the United States and Europe (95 and 50 years respectively). On July 2008 the European Commission decided to support term

---

[20] Boyle J, "A Natural Experiment", *Financial Times* (November 22 2004), http://is.gd/Ye0PhW.
[21] European Commission, *First Evaluation of Directive 96/9/EC on the Legal Protection of Databases*, DG Internal Market Working Paper, http://is.gd/DsY3XV.
[22] Ibid, p.24.

extension for copyright for performers,[23] a policy that was then adopted into a Directive in 2011 (Directive 2011/77/EU).[24] Justifying their position, the Commission stated that:

> "The extended term would benefit performers who could continue earning money over an additional period. A 95-year term would bridge the income gap that performers face when they turn 70, just as their early performances recorded in their 20s would lose protection. They will continue to be eligible for broadcast remuneration, remuneration for performances in public places, such as bars and discotheques, and compensation payments for private copying of their performances."[25]

The problem with this argument is that the evidence does not back up the assumptions behind it. For example, the Gowers Review of Intellectual Property[26] came strongly against term extension for sound recordings after commissioning a report dealing specifically with the economic evidence for and against extension.[27] The report concluded that "*the case for an extension of the copyright term in sound recordings to be weak*." More importantly, the report found that increasing term extension would be detrimental for the UK's balance of trade, and it would increase costs to consumers between £240 and £480 million GBP.

Similarly, the European Commission paid for another report from the IVIR Centre in Amsterdam.[28] The report answered the arguments put forward by content owners one by one: extending terms further than 50 years will not encourage more production; it will not make any difference to investments by the record industries; and it will erode the public domain. The report concluded that "*[t]he authors of this study are not convinced by the arguments made in favour of a term extension*."

---

[23] European Commission, *Intellectual Property: Commission adopts forward-looking package*, IP/08/1156 (July 2008).
[24] Directive 2011/77/EU of The European Parliament and of the Council of 27 September 2011 amending Directive 2006/116/EC on the term of protection of copyright and certain related rights, OJ L 265/1.
[25] European Commission, see supra note 23,
[26] HM Treasury, *Gowers Review of Intellectual Property*, (2005), E.10. http://bit.ly/gwg0tu.
[27] Centre for Intellectual Property and Information Law, *Review of the Economic Evidence Relating to an Extension of the Term of Copyright in Sound Recordings*, University of Cambridge (2005), http://bit.ly/zODygW.
[28] Hugenholtz B et al, *The Recasting of Copyright & Related Rights for the Knowledge Economy*,

There are several more examples like these, where intellectual property policy is shaped by the whims of whoever is in power at the time, but mostly seems to be geared towards serving narrow sectors of the economy. As new business models come into play, and new industries rise in economic importance, the old alliances seem to be crumbling, and one would expect that new policies and legislative proposals would reflect the changing reality.

Nonetheless, if anything, things appear to be getting worse. The content industries still pull a lot of power in policymaking circles, they are able to push and lobby for legislative proposals and treaties like the Stop Online Piracy Act (SOPA)[29] in the US, and the Anti-Counterfeiting Trade Agreement (ACTA)[30] at an international level. These proposals pit the technology sector against the content industries, and evidence is once more a casualty in the proceedings.

The rather haphazard manner with which legislation such as SOPA and ACTA were drafted are symptomatic of a malaise that has plagued IP policymaking for years, particularly in technological areas. But while the actual decision makers draft legislation without the use of evidence, the same bodies are busy giving lip service to the importance of evidence-based decision-making processes. The very same European Commission that increased term protection released a 2007 Green Paper to support the importance of evidence in the digital economy. They stated:

> "The Commission and national authorities have wide experience with market and sector monitoring. This needs to be further developed. Competition sector inquiries, the identification of lead markets and the development of joint technology initiatives provide a large part of the answer. But in a number of areas, more evidence needs to be gathered through effective feedback from the operation of the single market on the ground. Better account also needs to be taken of the consumer, SMEs and the global dimension as well as social and environmental impacts."[31]

---

[29] H.R.3261, http://thomas.loc.gov/cgi-bin/bdquery/z?d112:h.r.03261:.
[30] http://www.mofa.go.jp/policy/economy/i_property/pdfs/acta1105_en.pdf.
[31] European Commission, *A single market for 21st century Europe: Accompanying Communication*, COM(2007) 724 final (2007).

Similarly, the UK government has commissioned some studies that appear to follow this line of reasoning in its own IP-related studies. The aforementioned Gowers Review had a very strong stance with regards to the importance of evidence in its policy analysis of the intellectual property markets in the United Kingdom. The Gowers Review of Intellectual Property states:

> "The Review takes an evidence-based approach to its policy analysis and has supplemented internal analysis by commissioning external experts to examine the economic impact of changes to the length of copyright term on sound recordings, and the question of orphan works." [32]

The Review then goes on to cite evidence in specific areas of study, and based on those made its recommendations. The Gowers Review was generally well received by most sectors and stakeholders, yet it is perhaps ironic that an evidence-based review of this calibre was later ignored by the same government that commissioned it when it drafted changes to IP legislation, namely the Digital Economy Act 2010.

In a similar vein, the UK Intellectual Property Office commissioned another report, this one specifically on the subject of IP and the digital economy. The Hargreaves Report states:

> "Government should ensure that development of the IP System is driven as far as possible by objective evidence. Policy should balance measurable economic objectives against social goals and potential benefits for rights holders against impacts on consumers and other interests. These concerns will be of particular importance in assessing future claims to extend rights or in determining desirable limits to rights."[33]

The Hargreaves Review makes probably the strongest case yet for evidence as the basis for policymaking, it both encourages its adoption and uses it to draft a comprehensive arrange of recommendations. However, as James Boyle quipped, "as

---

[32] HM Treasury, see supra note 26.
[33] Hargreaves I, Digital Opportunity: A review of Intellectual Property and Growth, (2011), http://www.ipo.gov.uk/ipreview-finalreport.pdf.

opposed to what, you might ask. Astrology-based?"[34] This is sadly accurate because something as straightforward as evidence should require three separate official documents supporting it, and yet be ignored when those policies are adopted.

Let us be optimistic and imagine that future governments around the world will begin to adopt these recommendations eventually, and evidence will become the norm. This is precisely where there is dire need of well-informed and relevant studies into all sorts of issues, but especially within digital environments. Without proper understanding of how the Internet operates, how can policymakers expect to be able to draft appropriate legislation and implement effective regulation to the electronic domain?

Specifically, it is the contention of the book that some of the most important regulatory failures in recent years, namely the difficulty in curbing copyright infringement and the failure to tackle cybercrime, can be blamed in part to a systemic lack of understanding of how the Internet works as a complex system.

In the following sections, we will provide further examples of how network theory can help Internet regulation by giving specific evidence about what is at stake.

5.2 Centrality and copyright enforcement

On January 18, 2012, thousands of sites around the world were either blacked out or operated with protest notices on display due to a couple of pieces of legislation being discussed in the United States, the Stop Online Piracy Act and the Protect IP Act (SOPA[35] and PIPA[36] respectively). The high profile of the protesters –a list that includes Wikipedia, Google, Reddit, and Wired amongst others– translated into a level of coverage hardly seen for a technology story.

SOPA is a good example of Internet regulation that was drafted without any consideration with regards to evidence, and it is one area where some understanding of network theory would have been useful. SOPA quickly became unpopular with

[34] Boyle J, "An Intellectual Property System for the Internet Age", *Financial Times* (May 18, 2011).
[35] See supra note 21.
[36] http://thomas.loc.gov/cgi-bin/query/z?c112:S.968:.

important sectors of the technology community. The bill proposed powers to US government agencies to be able to force Internet Service Providers (ISPs) and search engines to block access to specific sites.[37] But something that went unreported was that there was more to worry about SOPA and PIPA than the difficulty created by enhanced powers to filter content. The most controversial norm in SOPA was contained in s102, which reads:

> "A service provider shall take technically feasible and reasonable measures designed to prevent access by its subscribers located within the United States to the foreign infringing site (or portion thereof) that is subject to the order, including measures designed to prevent the domain name of the foreign infringing site (or portion thereof) from resolving to that domain name's Internet Protocol address. Such actions shall be taken as expeditiously as possible, but in any case within 5 days after being served with a copy of the order, or within such time as the court may order."

This would have established a filtering responsibility for ISPs and other intermediaries against alleged copyright infringers, and it would have had serious effects in the United States. Looking at such a filtering provision using network theory, its danger to countries around the world becomes clearer because as drafted, it is very possible that SOPA and PIPA could have serious extraterritorial consequences. The reason for this is that the existing network architecture of the Internet is centred heavily on the United States, and any legislation that affects the core infrastructure in that country could have cascading consequences elsewhere.

As is explained in more detail in the book, network theory is the systematic study of any netlike or complex system or collection of interrelated things; networks are broken into their basic elements (namely nodes and links), and studied to discern patterns.[38] In network theory, there is a concept called centrality which measures the importance of a

---

[37] EFF, Stop the Internet Blacklist Bills, (2011), http://blacklist.eff.org/.
[38] Guadamuz A, *Networks, Complexity and Internet Regulation: Scale-Free Law*, Cheltenham, UK: Edward Elgar (2011), p.15.

node in any given network.[39] This is calculated by the number of links a node has to neighbouring nodes, the shortest number of paths to other nodes in the network, and the average shortest path. A node is said to be central in a network if it is linked to a large number of other nodes, if it can be connected to other nodes quickly (the six degrees of separation phenomenon),[40] and if the average distance to other nodes is short. When plotting charts describing networks, central nodes can be sometimes easily identified as in figure 7.1, where darker shades indicate more central nodes.

Node and hub centrality is an important indication that there is a power law at work in a network, as high concentration of centrality in some nodes may give rise to a scale-free network, where some nodes are more important than others.[41] While the concept of scale-free networks is explored in more detail in the book, suffice it to say that it is one in which some nodes have considerably more links than could be expected by average, so these types of networks result in hubs and even super-hubs that act as important connectors in the system's structure. The Internet is a scale-free network,[42] so centrality comes into play in two ways. First we have the physical network, the wires, routers and hubs that make up its physical architecture. Second there is the logical level of centrality, which consists of websites, links, hyper-links, but also include the Domain Name System (DNS) and Internet governance structures. It should be no surprise to anyone to learn that any way you look at the Internet, the United States is extremely central. Take for example this picture of the global submarine cable network (Figure 9.1):

---

[39] Freeman LC, "Centrality in Social Networks: Conceptual Clarification", 1:3 *Social Networks* 215 (1979).
[40] Milgram S, "The Small World Problem", 2 *Psychology Today* 60–67 (1967).
[41] A scale-free network is a system in which the same distribution of relationships exists at any scale, see: Ravasz E and Barabási A-L, "Hierarchical Organization in Complex Networks", 67.
[42] Huberman BA, *The Laws of the Web : Patterns in the Ecology of Information*, Cambridge, Mass.: MIT Press (2001).

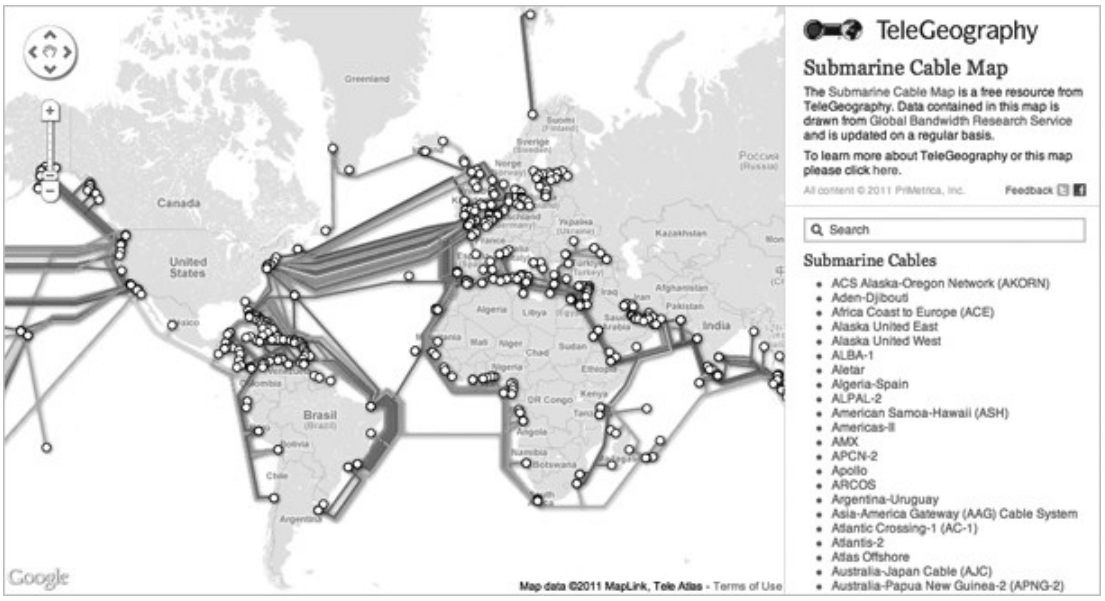


Figure 9.1 Mapping physical centrality[43]

You will notice that the number of cables going in and out of the US is more than any other country, and while there are other important hubs (such as the UK), that country's central role in the physical backbone of the Web is considerably high. This is just one aspect of the large dominance that the US has in the Internet's infrastructure. Things get even more interesting when you look at the logical architecture, where the US has managed to remain considerably ahead of other countries. While anyone can become an Internet server by just installing the appropriate software into any computer connected to the Web, you need a registrar if you want a domain name that resolves in the system (such as google.com, or facebook.com). Most top level domains are registered in the United States (.com, .org, .net, .biz), and statistics show that the US is the country with the most domain names registered under its jurisdiction, with 78,453,258 as of January 2012.[44] The closest second country is Germany with over six million registrations. In

[43] http://www.submarinecablemap.com/.
[44] http://www.webhosting.info/domains/country_stats/.

fact, not even combining all of the other countries in the world can you reach the total of domains registered in the US.

Table 9.1 Country-wise Domains Distribution: Domain Names by Country of Purchase.

Rank	Country	Domains
1	United States	78,453,258
2	Germany	6,481,160
3	United Kingdom	4,617,854
4	China	4,502,381
5	Canada	3,869,783
6	France	3,271,896
7	Japan	2,483,667
8	Australia	2,405,261
9	Spain	1,589,942
10	The Netherlands	1,372,323

A similar picture emerges with regards to hosting, that is, where content is actually placed in a server. If we look at the number of Internet users per region, Asia has 44% of the world's Net population, while Europe has reached 22%, and North America has only 12%.[45] However, large amounts of content are still hosted in companies based in the US. In fact, 9 out of the top 10 hosting companies are American, and of these, the largest host in the world is GoDaddy (WildWestDomains in Figure 9.2).

[45] http://www.internetworldstats.com/stats.htm.

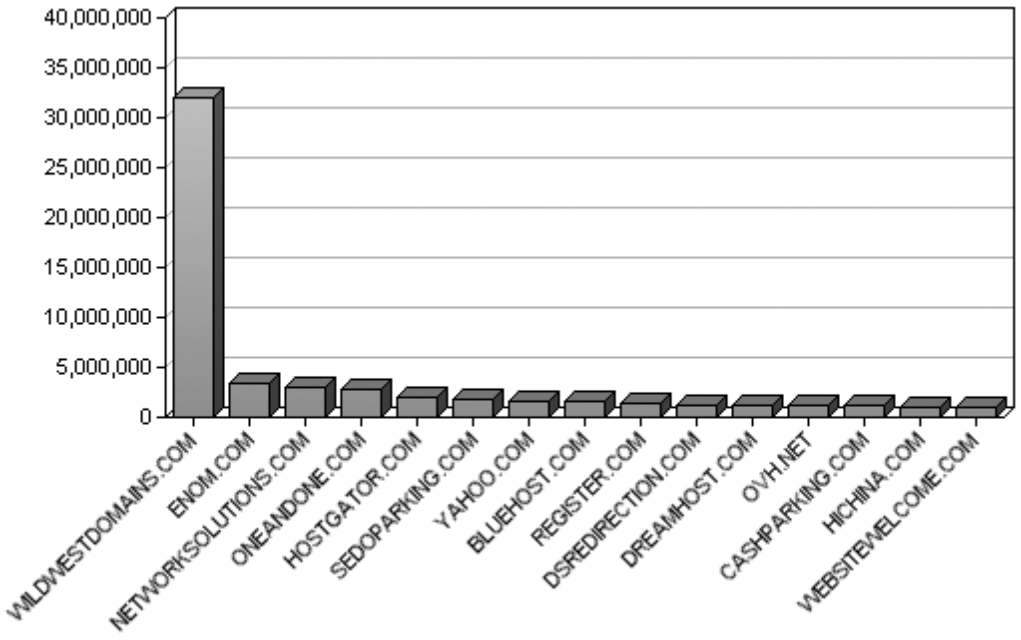


Figure 9.2 Top hosting companies in the world[46]

The end result is a skewed map of the world, where whatever happens in the United States disproportionately affects the rest of the Internet. Here is where a study of network centrality can be useful, as legislation that would affect both the physical and logical infrastructure of the Web would not only be national, but it would affect large sectors of the world's users.

In the book we highlighted studies into the dangers of the current state of network centrality in the global network.[47] Researchers conducted a survey of country centrality to try to determine the potential downstream negative effects of country-wide censorship of the Internet. The objective of the paper was to establish an analytical framework for determining the influence of each country within the flow of international traffic. The researchers collected traceroute data between countries trying to determine the paths

[46] http://www.webhosting.info/webhosts/tophosts/Country/US.
[47] Karlin J, Forrest S and Rexford J, *Nation-State Routing: Censorship, Wiretapping, and BGP*, arXiv Working Paper (2009), http://arxiv.org/abs/0903.3218v1.

taken by information in the global network. This produced high levels of centrality consistent with the other studies highlighted in previous sections. What is novel about the approach of this study is that it also calculated what they call "strong country centrality" (SCC). They assumed that under some circumstances there may be other paths to information that do not go through one country. SCC would take place when all other viable paths led through that country as well; in other words, data had no other way of getting from A to B other than through that country. In findings consistent with other centrality studies, they found that the United States, the UK and Germany were the most central countries on the Internet, but also displayed high levels of SCC.

This problem can be seen from another perspective to prove just how vital centrality is when dealing with Internet filtering. Bloem et al conducted a study trying to determine which would be the optimal point to filter malicious software and viruses within a network.[48] They looked at the rate in which malware spreads in large networks such as the Internet, and wanted to calculate whether it would be possible to do so by deploying filtering software in specific nodes. They discovered that the more central the node, the easier it would be to sift the undesired computer programs from the system, and inversely, filtering would become more difficult if the server was not a central one. Applying these findings to a large scale, any sort of block placed in a central part of the network has higher possibility of affecting downstream users.

It is then easy to see how SOPA and PIPA, or any other future law that proposes some sort of Internet filter, would have global consequences. As written, SOPA could perfectly trickle downstream to other physical and logical clients elsewhere, which would mean that SOPA would be used to filter content to all of us. This is not as far-fetched as it may sound. So far we have not witnessed too many incidents regarding downstream filtering because most of those practices take place in countries that are not central at all to the Web's inner workings. Seen from an architectural perspective,

[48] Bloem M et al. "Malware Filtering for Network Security Using Weighted Optimality Measures", *IEEE International Conference on Control Applications* (2007).

countries like China, Syria, Egypt and Pakistan, where national filtering takes place, are mostly large Intranets, which is why it was so easy for Egyptian authorities to shut down all Web access during that country's revolution. The US has an entirely different role to the network, so anything that is filtered there could end up being filtered in places that have never heard of SOPA, be it justified or not.

While SOPA and PIPA have been shelved at the time of writing, due in large part to the scale of opposition that the bills encountered, their existence serves to prove the point that legislation is currently being drafted by people who are unaware of the Internet's basic workings. Examples such as these abound in copyright policymaking, and it is hoped that this work will help to make people aware of a body of evidence that has gone largely ignored until now.

5.3 Resilience

Another area where knowledge of networks could serve to better inform policy is the understanding of resilience of scale-free networks. The subject of resilience is covered in some detail in the book, especially dealing with the robustness of networks which can lead to copyright infringement, such as BitTorrent, and that of cybercrime networks and botnets. To recap the concept, networks that display power law characteristics tend to be resilient to random attacks.[49] This is because some vital elements of the network have more links than other, and any random attack will probably not knock down the important nodes within the system.

Resilience has been indirectly in the news recently, although the reports were not aware that what was being discussed was precisely about that subject. First, the director of the US National Security Agency (NSA) was quoted by the Wall Street Journal

---

[49] Albert R, Jeong H and Barabási A-L, "Error and Attack Tolerance in Complex Networks", 406 *Nature* 378-382 (2000); see also Chassin D P and Posse C, "Evaluating North American Electric Grid Reliability Using the Barabási–Albert Network Model" 355:2 *Physica A: Statistical Mechanics and its Applications* 667 (2005).

implying that the hacker collective Anonymous had the capability of attacking and seriously affecting that country's power grid. According to the report:

> "That threat was described to lawmakers at a hearing last week. "A near-peer competitor [country] could give cyber malware capability to some fringe group," said Gen. Martin Dempsey, chairman of the Joint Chiefs of Staff. "Some hacker, next thing you know, could be into our electrical grid. We have to get after this.""[50]

Almost at the same time, officials have also expressed concerned about another alleged potential hacker attack to the Internet's top level infrastructure, the 13 root name DNS servers that are located around the world. On February 12 2012, a person identifying himself as member of Anonymous posted a document online detailing an operation planned for March 2012 in which hacktivists will attempt to bring down the global network (at the time of writing, the attach has not taken place).[51] The message gives the IP addresses of the root servers, and threatens to deploy a concerted Distributed Denial of Service (DDoS) to those computers.

Anonymous has denied both threats, but as there is no one who speaks for these groups in an official manner, we have to go on hints and guesses. Is there a real threat, or are hackers and cyber threats the new "weapons of mass destruction"? The question is easier to answer from a network theory perspective.

Let us assume that the threats themselves are real. Could Anonymous knock down the US power grid or the Internet? The answer to the first one is maybe, and to the second is a resounding negative.

Interestingly, power grids have been the subject of various studies on vulnerability and resilience in networks, partly because there is a lot of data about them, they do not move, but also because of their importance. The seminal work on the subject is a study

---

[50] Gorman S, "Alert on Hacker Power Play", *Wall Street Journal* (February 2012), http://on.wsj.com/zR9MLi.

[51] http://pastebin.com/NKbnh8q8.

by Kinney et al,[52] which looks at the North American grid in particular because it is one of the most complex networks of this nature. They concluded that the failure of highly connected nodes in the network could knock down as much as 25% of the entire system due to what is known as a cascading failure. This is an effect that happens in scale-free networks that have a high degree distribution, in other words, systems that rely on highly connected and/or vital elements, so anything that happens to such a hub would affect its downstream tributaries. For contrast, a similar study[53] was conducted in European power grids, and it found that most of the studied grids lacked scale-free characteristics in degree distribution, so they relied less on highly-connected nodes, and therefore would result in more resilient networks, and failure of a node would be less likely to result in cascading failures.

It would then be theoretically possible for a well-orchestrated attack to knock down significant parts of the US power grid, but not of the ones in Europe. However, a big misconception about hacking is that people tend to mistake a DDoS with an attack on a target's computer infrastructure. In reality, most hacking attacks against websites manage to knock out pages from the Internet, but do not affect the actual computer system behind it. In the case of power grids, a concerted attack would have to be considerably more virulent than the average incident that we are used to. So while security agencies should be asking questions about possible hacking vulnerabilities of national power grids, the reality is that these concerns should not be a priority until it is proven that random hackers have actual access to the operating systems behind the electrical infrastructure.

---

[52] Kinney R et al, "Modeling Cascading Failures in the North American Power Grid", 46:1 *The European Physical Journal B* 101 (2005).

[53] Rosas M, Valverde S and Solé R V, "Topological Vulnerability of the European Power Grid Under Errors and Attacks" 17:1 *International Journal of Bifurcation and Chaos* 2465 (2007).

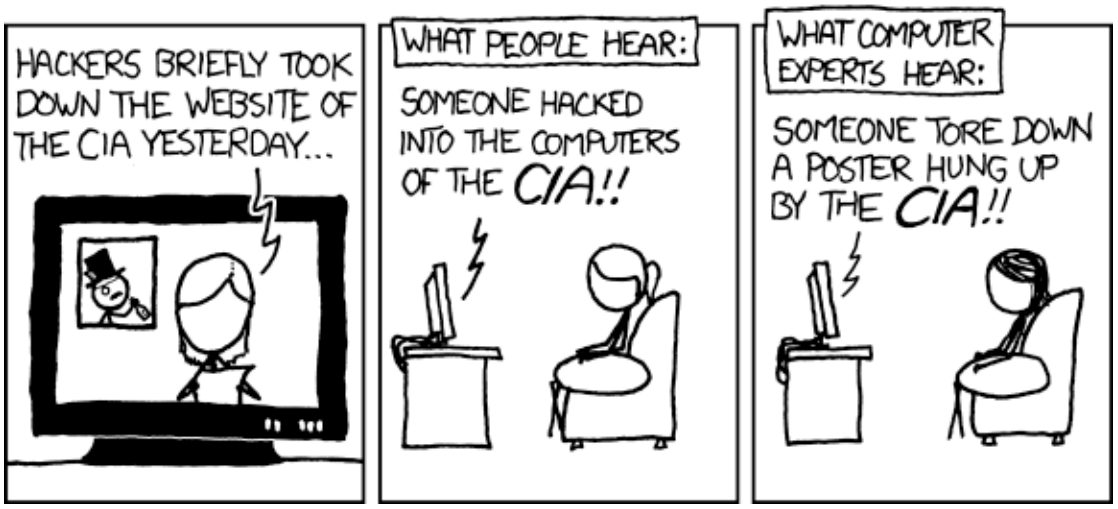


Figure 9.3 Websites are not the same as backbone systems[54]

The DNS root servers are a different proposition altogether. The Internet is at its most basic an interconnection of computers using IP addresses, but as it is difficult to remember numbers, the Web relies on a hierarchical naming system that translates numbers into domain names (such as google.com). The root name servers are at the top of the hierarchy, so any change in any domain name connected to the Internet has to go through those computers. In other words, if the root name system crashed, a browser would not know how to find the computer that hosts amazon.com, or yahoo.co.uk. As mentioned above, there are 13 root servers; and each is assigned a specific letter:

Table 9.2 Root Name Servers.

Letter	IPv4 address	Operator
A	198.41.0.4	Verisign
B	192.228.79.201	USC-ISI
C	192.33.4.12	Cogent Communications
D	128.8.10.90	University of Maryland

[54] Xkcd, *CIA*, http://xkcd.com/932/.

E	192.203.230.10	NASA
F	192.5.5.241	Internet Systems Consortium
G	192.112.36.4	Defense Information Systems Agency
H	128.63.2.53	U.S. Army Research Lab
I	192.36.148.17	Autonomica
J	192.58.128.30	Verisign
K	193.0.14.129	RIPE NCC
A	198.41.0.4	ICANN
B	192.228.79.201	WIDE Project
C	192.33.4.12	Verisign

Such centrality would lead one to believe that the system is highly vulnerable to attacks, but this is actually not the case. The first element is historical, as the root servers were already subject to a DDoS strike in 2002 and on February 2007.[55] In the latest and largest strike, only 6 of the 13 servers were subjected to serious hits, and of these only two were slowed down in any noticeable manner, mostly due to not having updated software. The result was that the Internet did not suffer as a result.

Similarly, any study of the root name server system will indicate that it is actually not scale-free at all, and it has considerable distribution of nodes. While in theory there are 13 servers, the reality is that these are distributed in 260 locations around the world, which builds redundancies into the system (Figure 9.4).

[55] ICANN, *Factsheet: Root server attack on 6 February 2007*, http://bit.ly/hcpBun.

Figure 9.4 The location of the "thirteen" root servers[56]

Moreover, there are dozens of studies looking into the resilience of complex computer networks in general,[57] and the DNS system in particular.[58] The result is that the distributed nature of the network has been able to make it less likely to be brought down even by a well-coordinated strike. The importance of this for policymaking is that it allows authorities to allocate resources away from imaginary threats and place them where they can be useful.

The study of network robustness can also help to elucidate when a website that is being targeted by law enforcement agencies will be brought down, and when it will survive. There are three notable examples in recent years, The Pirate Bay, Wikileaks and Megaupload.

---

[56] http://www.root-servers.org/.

[57] Wosinska L et al, "Network Resilience in Future Optical Networks", in Hutchison D et al (eds), *Lecture Notes in Computer Science*, Berlin: Springer (2009).

[58] Chassin D P and Posse C, "Evaluating North American Electric Grid Reliability Using the Barabási–Albert Network Model" 355:2 *Physica A: Statistical Mechanics and its Applications* 667 (2005).

The Pirate Bay is already covered in the book in more detail, and the failure to shut it down is, in my opinion, a strong indication of the robustness of scale-free networks.

Wikileaks provides another interesting study in resilience of online content. Wikileaks is not in itself a scale-free network, but it displays the robustness of the entire Internet as a complex adaptive system itself, and shows just how difficult it is to knock down content from the Web. On November 28 2010, the whistleblowing site WikiLeaks began releasing some of the more than 250,000 diplomatic cables from USA embassies around the world, in a coordinated exercise with large newspapers from around the world.[59] The main release was done through the then Wikileaks website. The cables contained embarrassing details both to the United States and to various governments around the world, and in some cases, even some sensitive data that has sparked political unrest in various fronts.

From the very beginning, there were calls from various parties within the United States to try to shut down Wikileaks.[60] What followed was almost a textbook case study on Internet resilience, and just how difficult it is to police the Internet. The actual Wikileaks website (www.wikileaks.org) was housed in several hosting services, mostly in Sweden and France, but they had also bought hosting space in the cloud computing web services offered by Amazon.com. The Wikileaks domain name (wikileaks.org) was assigned by California domain name registrar EveryDNS.net, which also provided free DNS services. By December 1st 2010, just a couple of days after the initial leaks, Amazon had dropped the service alleging breach of its Terms of Use, and EveryDNS.net revoked the DNS registration alleging damage to its servers from co-ordinated cyber-attacks. By the end of that week, several payment systems which took donations for the site (including Visa, MasterCard and PayPal) had also dropped the organisation. Bereft of hosting, routing and monetary channels, one would have thought that Wikileaks

---

[59] Leigh D, "How 250,000 US embassy cables were leaked", *The Guardian* (28 November 2010), http://goo.gl/Azhcm.

[60] Sarah Palin wrote on Twitter that "Inexplicable: I recently won in court to stop my book "America by Heart" from being leaked, but US Govt can't stop Wikileaks' treasonous act?" http://twitter.com/SarahPalinUSA/status/9251635779866625

would simply disappear. However, network science shows us that the Web is incredibly resilient.

There is something at which the Internet is really good at, it takes censorship as an attack to its infrastructure, and reroutes services to avoid the affected area. Just a few minutes after Wikileaks had its DNS services removed; the fact was advertised to the world via Twitter and Facebook. Because the site was still hosted somewhere, it was still possible to access the content via an IP address (at the time of writing, it was hosted at http://88.80.2.31, in Sweden). Similarly, several mirrors and new DNS registrations started popping up everywhere, and social media was instrumental in making users know where the content could be reached. By tweeting and retweeting the latest IP addresses where it could be found, Wikileaks managed to survive through the crisis.

Wikileaks shows that distributed architectures, which are a hallmark of robust scale-free systems, will make a site considerably more difficult to remove from the Internet. Contrast that to what happened with the digital locker website Megaupload.

On January 19 2012, just a day after several sites had protested against SOPA, the file-sharing site Megaupload was the subject of an international law enforcement operation by U.S. authorities, who managed to have six men arrested in New Zealand and charged with running an international criminal operation engaged in copyright infringement. At the same time, the FBI also managed to shut down the site through technical means by ordering their domain registrar to seize the names, so the addresses megaupload.com and related domains do not resolve in the system.

Megaupload was huge; it accounted for 4% of Internet traffic and received an estimated 50 million visitors per day. It used up more bandwidth than any other digital locker combined. Megaupload operated both a free and a subscription service, but what really seems to have played a big part in its demise is the fact that it was run as a business with income estimated at $150 million USD in subscription fees and $25 million USD from advertising.

What happened to Megaupload is completely different to the experiences of The Pirate Bay and Wikileaks because it was a considerably centralised service. While the company was registered in Hong Kong, and most of its operations ran from New Zealand, its .com domain was registered to an American company named DotRegistrar. Similarly, the site had also hired some of its hosting services to companies in the States, where at some point it had leased more than 1,000 servers to companies like Carpathia Hosting and Cogent Communications. This state of affairs opened up Megaupload to enforcement by American authorities, which is precisely what took place.

Megaupload then comes to exemplify that some sort of Internet regulation is possible, and that it is indeed feasible to remove websites from the Internet. However, the case also illustrates that such enforcement seems only possible with centralised services. This is a story that has been repeated throughout the relatively short history of the fight against piracy. Centralised systems are easily shut down (such as Napster and now Megaupload), while distributed networks survive almost anything that law enforcement can throw at them.

This is why both cybercrime and copyright infringement are areas where wider understanding of network science would help shaping enforcement strategies in the future.

5.4 Privacy

An area of growing interest that was not covered in the book, but that can benefit from the analytical tools borrowed from the study of complex systems is that of online privacy. This has been the subject of growing regulatory concern given the increasing threats from mainstream services with regards to user privacy. Just recently, Google was caught bypassing user no-tracking preferences in iPhone devices,[61] while Facebook has

[61] Angwin J and Valentino J, "Google's iPhone Tracking", Wall Street Journal (February 17, 2012), http://on.wsj.com/ytg5EB.

been criticised for rolling out a new interface that may uncover more about a user than they wish to make public.[62]

On January 25 2012, the European Commission proposed a complete overhaul of its data protection and digital privacy framework in order to respond to these and other threats to consumers in the European Union. The Commission is proposing two main regulatory changes, a new Data Protection Directive[63] and the General Data Protection Regulation,[64] both of which will be tailored for the protection of the consumer in a digital environment. The main changes proposed include several far-reaching provisions that will affect Internet privacy, including the creation of an obligation to notify authorities of serious data breaches as soon as they take place; and more importantly, the proposed creation of a "right to be forgotten", which will make it easier for data subjects to amend or erase data held about them if it is no longer necessary.[65]

These changes to European privacy law are immensely welcome as Internet privacy has become such a hot topic, but I am concerned that they may prove ineffective as they have not been drafted with any evidence of how information actually flows in the network. The Commission did undertake a comprehensive array of reports and impact assessments of the 1995 Data Protection Directive and the existing privacy framework,[66] but none of this looked directly at any sort of network theory. This is understandable, as these theories are not yet part of the mainstream, but this would have been a primary area in which to make use of complexity.

---

[62] Dembosky A, "Facebook Timeline Ads Plan Raises Fresh Privacy Fears", *Financial Times* (February 9, 2012), http://on.ft.com/xQHIRj.

[63] Proposal for a Directive of the European Parliament And of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data, COM/2012/010.

[64] Proposal for a Regulation of the European Parliament And of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, COM(2012) 11 final.

[65] Art. 17 of the Regulation.

[66] The conclusions can be found here: http://bit.ly/xrP1pi.

Just how relevant can network theory be in the field of online privacy? One recent court case may help us to illustrate just how much is the law struggling with online environments. The case is *AMP v Persons Unknown.*[67] The case is interesting because it features BitTorrent, but not in a copyright context, but rather a privacy one. The question at the heart of the case is whether it is possible to remove one specific torrent file from the Internet through legal means.

In June 2008, a female university student from Notthingham lost or had her mobile phone stolen while travelling in that city's public transport.[68] The phone contained images "*of an explicit sexual nature which were taken for the personal use of her boyfriend at the time*".[69] Shortly after the theft, the images were copied from the phone and uploaded to a picture-sharing site with her name and a link to her Facebook page attached. Someone warned her of this fact, and an email was sent to the hosting website, which promptly removed the images. However, once in digital format the content is more likely to replicate, and this case was not the exception to that rule; the images were bundled into a torrent file and uploaded to The Pirate Bay under the title "Sexy Rich Chick Mobile Phone Found by IRC Nerdz". At the same time another person contacted her on Facebook and threatened to have her exposed unless she friended him. Similarly, her parent's company was contacted with blackmail threats. While this was happening, the torrent file spread around the various tracker sites, where it is still available at the time of writing.

The claimant's family, as it turns out, is considerably wealthy, so her parents hired the assistance of law firms and computer experts to try and have the images removed from the Web. The lawyers filed a Digital Millennium Copyright Act (DMCA) takedown notice[70] to Google in the US with the intention of having the search engine remove links

---

[67] *AMP v Persons Unknown* [2011] EWHC 3454 (TCC).
[68] Her identity, for reasons that will become clear, has been kept hidden both in the ruling and here.
[69] Para 5.
[70] Urban J M, "Efficient Process or Chilling Effects-Takedown Notices under Section 512 of the Digital Millennium Copyright Act" 22 *Santa Clara Computer & High Technology Law Journal* 621 (2005-2006).

to the torrent files from searches on copyright grounds.[71] They then filed for an injunction in the High Court of England and Wales "*to prevent transmission, storage and indexing*" of the pictures based on the claimant's right to privacy under Article 8 of the European Convention on Human Rights, and under Section 3 of the Protection from Harassment Act 1997. This is not the place to go into a detailed analysis of the substantive law in this case; suffice to say that the court makes a very good application of existing UK privacy law, and continues to set a high standard of privacy protection in online environments. Ramsey J delivers a thoughtful and well considered ruling in that regard. In short, privacy law does not affect freedom of expression in cases like this, and the possible damage done to the claimant's enjoyment of a private life outweighs other considerations, and therefore should preclude the publication of the images through any media.

The ruling is truly ground-breaking in the fact that for the first time a court in the UK has been asked to serve an injunction against the publication of a specific torrent file. The case was resolved with the judge issuing an injunction result of the case is that the judge decided to issue a blanket injunction against anyone who is eventually found to be seeding the file in the future. In order to ascertain if this was possible, Ramsey J had to look in detail at how BitTorrent works. Based on expert testimony, the court defines BitTorrent like this:

> "BitTorrent is a peer to peer file sharing protocol used for distributing large amounts of data over the internet. The BitTorrent protocol is used to download files quickly by reducing the server and network impact of distributing large files. Rather than downloading a file from a single source server, as is the case with the conventional HyperText Transfer Protocol (HTTP), the BitTorrent protocol allows users to join a "swarm" of users to download and upload from each other simultaneously."[72]

The ruling then describes how the BitTorrent protocol breaks up the file, and several people in the swarm are sharing pieces at the same time, either serving its entirety as

[71] See the notice here: http://www.chillingeffects.org/notice.cgi?sID=73152.
[72] Para 10.

"seeders", or sharing it while downloading it as "leechers". This is an accurate and useful definition of BitTorrent; and it is always useful to have a court define and understand the underlying technology. However, things start going a bit amiss in the decision when the court tries to determine if it is possible to stop a torrent file from being shared. According to the experts consulted by the court, such a thing is possible. The ruling says:

> "[To] prevent the transmission, storage and indexing of the relevant ".torrent" files it is necessary to identify the users who have downloaded the files using the BitTorrent protocol. The relevant files can then be deleted by these users and, in addition, these users can be prevented from acting as seeders of parts of the file which will prevent them distributing the images which are the subject of the current claim. […] each seeder can be identified by way of their Internet Protocol Address ('IP Address') while they are seeding. […] it would therefore be possible to obtain the IP Address of every seeder in the swarm and identify from that address their physical location, name and address from their Internet Service Provider. […] it would be possible to identify the IP address of each computer seeding a particular ".torrent" file and details of the person allowing the seeding to take place. They could therefore be served with an order requiring them to take steps to stop their account from being used."[73]

This assessment of the capabilities of checking IP addresses is not accurate. Sure, it is perfectly possible to obtain IP addresses of those sharing a specific file at any given time, but this soon becomes a game of whack-a-mole because nothing prevents others from creating other torrent files and sharing them online. Similarly, IP address identification is not an exact science, and anonymisation through the use of technological tools is perfectly possible. Moreover, and this cannot be stressed enough times, an IP address never identifies a person, it may identify a household, but there is no way of knowing if that address is being used in a public place, or by a person who has their wireless network open, etc. In other words, a court cannot possibly expect to be able to serve an injunction to those actually seeding the file. It might identify some, but it is equally likely that it will have false positives and send the injunction to the wrong people.

---

[73] Paras 14-16.

The books goes into a lot of detail of the network characteristics of BitTorrent sharing system, and a simple look at these would lead one to conclude that shutting down the sharing of a single file is almost impossible. There is enough evidence to indicate that the BitTorrent protocol has scale-free characteristics,[74] and therefore is both resilient and viral. Moreover, each torrent file being shared could be considered its own isolated network. In other words, shutting down the actual network of those sharing a specific .torrent file with the infringing pictures would be next to impossible, as even if one person is left seeding the content means that it can replicate again, and the network survives. An injunction of the type that is being sought in this case could be equated with a random attack on a scale-free network, and such strikes are unlikely to knock-out the network.

To illustrate this point, even after the ruling, a simple Google search led me to the torrent file in question in The Pirate Bay, which indicates that the injunction has failed in its stated purpose already. One look at the file with the appropriate file displays a log of the IP address of those sharing the file at any given time. By looking up those addresses with network analysis tools, one can find out that the seeders were located in Canada, the United States and Sweden. It would be necessary for the claimant to try to get courts in all of those countries to issue similar injunctions to that which was granted here. Even with large monetary resources, this seems both impractical and futile.

As sad as this case is, it serves as another example of why policymakers, experts and judges dealing with online environments should be at least a little bit familiar with network theories, and particularly with concepts like resilience. In certain online piracy debates one should always take into account the fact that distributed complex networks are remarkably robust, and therefore any enforcement effort should take that into account.

---

[74] Dale C et al, "Evolution and Enhancement of BitTorrent Network Topologies", 16th *International Workshop on Quality of Service* (2008).

Similarly, online privacy should take into consideration a phenomenon that is also related to network theories. One interesting characteristic of scale-free topologies is what is known as the "rich-get-richer" phenomenon.[75] This is a concept that postulates that some links that are already being linked in the system have greater opportunity of being linked to in the future, and sites tend to acquire new connections in proportional relation to those it already has. In other words, the more incoming links a node has, the more likely it is to accumulate more links. This helps to explain skewed usage figures in every level of granularity in the Web. What seems to happen is what researchers call preferential attachment, where people link more to things that are already popular.[76] This is relevant for privacy because if there are already links to some form of content, e.g. the pictures in the AMP case, then it is likely that drawing attention to their existence by initiating legal proceedings tends to make matters worse. This is known as the Streisand effect, after the singer sued to have some images removed, but that made them go viral on the Internet. When it comes to privacy in online environments, the cure may be worse than the medicine.

The spread of information in social networks offers another tangible area of study arising from network theory that could help to shape policy and case-law in the future. Social networks are practically tailored to facilitate data-mining and analysis of social interaction, and such evidence could be used to shape more realistic and effective policies.

An example of this is the way in which users interact with one another in those environments and how information is shared within a network. Understanding those interactions could perhaps assist a judge in a case like AMP in better targeting injunctive relief if it is still possible to do something. For example, a recent study has been looking at the way in which people share information in the social recommendations site

---

[75] Fabrikant A, Koutsoupias E and Papadimitriou C H, "Heuristically Optimized Trade-Offs: A New Paradigm for Power Laws in the Internet Paradigm for Power Laws in the Internet", 2380 *Lecture Notes in Computer Science* 781 (2002).

[76] Capocci A et al, "Preferential Attachment in the Growth of Social Networks: The Case of Wikipedia", 74 *Physical Review E* 036116 (2006).

digg.com.[77] We like to think of the Internet as a vast network where people from all over the world can exchange information, but what is emerging from research into social networks is that there is strong homophily, that is, people tend to connect with other like-minded people and share similar items. The same authors then looked specifically at how people consumed information in Digg and in the CNN iReport network.[78] Unlike some networks like BitTorrent, the researchers found highly-clustered groups which relied heavily on few interconnecting and super-connected hubs, which is consistent with their earlier findings about the homiphily present in online environments. These groups display power law characteristics, but the high clustering and the low level of degree distribution would tend to suggest that these groups would be highly affected if only one of the important hubs was removed. They point out that:

> "The only significant damage could be done if one of the "power users" in the fat tail were to be removed from the network. This effect would certainly be damaging for these networks since the observed clustering coeficients are unusually low, meaning that the neighbors of most hubs are generally unconnected. In this way, the loss of a single hub would mean the disconnection of entire groups of hub neighbors."

The study also found that information travelling through the network also relied on the connected hubs, and that most of those consuming information did so from second hand, that is, the hubs not only served as the glue holding together the groups, but also served as prime informers within the system. The implications for privacy policy seem clear. Imagine some form of privacy breach similar to the AMP case, but this time the data is not shared through BitTorrent, but within a social network. By conducting social network analysis of the system one could easily identify the hubs, and then attempt to

---

[77] Lussier J T, Raeder T and Chawla N V, "User Generated Content Consumption and Social Networking in Knowledge-Sharing OSNs", 6007 *Advances in Social Computing* 228 (2007).

[78] Lussier J T, Raeder T and Chawla N V, *Digging up the Dirt on User Generated Content Consumption*, Interdisciplinary Center for Network Science & Applications (ICeNSA), Working Paper, http://cse.vnit.ac.in/comad2010/ResearchTrack/paper%2057.pdf.

stop the data from spreading at those points, as these serve as hubs through which all of the information flows.

As with other possible applications of network theory tools, the use of concepts like centrality, resilience, small worlds and scale-free systems is in its early stages, but it is hoped that their usefulness to the topic of online privacy is evident from the above paragraphs.

6. THE DEBATE BETWEEN AN OPEN AND CLOSED INTERNET

While it is only stated partially at the end of Chapter 5, one of the biggest themes of the book with regards to Internet regulation is that network science has considerable relevance to the debate about whether the Web should be kept open or closed.

As it is explained in Chapter 4, the early history of Internet regulation was plagued with cyber-libertarianism, the idea that the Internet cannot and/or should not be controlled.[79] This view shifted steadily as more governments and international bodies attempted to place legislative restrictions to online behaviours. While some of these efforts failed, or were at best ineffective, some approaches were able to create some form of control. Particularly, the regulation at the choke-points, as Wu and Goldsmith called it,[80] as deployed in national firewalls and other similar practices have proven to be a surprisingly effective manner of control. By exercising tight control over the access points to the wider network, some countries have managed to somewhat tame the Internet. But this success comes at a price; the resulting network is not exactly the robust and distributed open environment that it was supposed to be, by closing the network, the Internet looks like something completely different.

---

[79] Johnson DR and Post DG, "Law and Borders: The Rise of Law in Cyberspace", 48 *Stanford Law Review* 1367 (1996).

[80] Goldsmith JL and Wu T, *Who Controls the Internet? Illusions of a Borderless World*, Oxford: Oxford University Press (2006), p.58.

The modern regulatory dichotomy is therefore between a more closed and controllable Internet, and an open and chaotic environment. While the open vs closed struggle is as old as regulation itself, this time there is an interesting twist to the story, as the battle is being fought both at the public and private levels.

In the public sphere the choice is deceivingly clear: governments can choose (or not) to exercise a tight control over Internet traffic in their territory through the deployment of technical means such as filtering software or firewalls. However, the book has explored that this dichotomy is not entirely straightforward, as countries that seem to support the open Internet model can also deploy specific policies that have as a result the creation of a system that resembles the closed models advocated by restrictive regimes. This arses from architectural choices in the regulation of specific areas, which end up having the opposite effect, namely, closing the Web. This is clear in the United States, where Secretary of State Hilary Clinton can give speeches which clearly advocate for an open Internet,[81] while at the same time the government and the US Congress pass legislation such as the Cyber Intelligence Sharing and Protection Act (CISPA),[82] which have the potential of affecting the Net's infrastructure negatively.

In the private market, the choice between openness and closedness is increasingly due to technical concerns. While the distributed nature of the network has been an important part of its growth since 1995, security concerns have also plagued the network. Nothing exemplifies this better than Apple. From a regulatory perspective, Apple favours the closed and controlled model that is akin to a walled garden in which their operating system for devices (iOS) and the iTunes store act as the filter through which you perceive the network. The walled garden also acts like a forbidden city that is intended to keep out the unsafe, unacceptable and/or unsavoury content from ever reaching the user. In this model, Apple operates as the ultimate censor because all content has to be pre-approved. So Apple users' experience is of a system in which decisions on what type

[81] Hilary Rodham Clinton, *Remarks on Internet Freedom*, (2010) http://www.state.gov/secretary/rm/2010/01/135519.htm.
[82] H.R.3523 , http://thomas.loc.gov/cgi-bin/query/z?c112:H.R.3523:.

of content the user wants to consume has already been taken. The assumption is that the walled garden is safer. Contrast that to the Google model based on openness. Both the Google mobile operating system (Android), and the Google Market (recently renamed Google Play) operate as open spaces with few limits on the type of applications allowed. This is precisely the model of the open Internet that was favoured for many years. Apple's argument is that such a model has given us spam, phishing, malware and botnets, so it makes sense to close development and to have a gatekeeper approving what gets on a device. This supposedly makes user experience more secure.

While it is the author's opinion to favour openness, this is not openly expressed in the book on purpose. The question is left open, but the choices are clear. They are architectural, and as stated above, rely on clear network science areas of study. Scale-free networks tend to be open, distributed and resilient. Centralised systems are more closed and fragile. It is clear that in some instances the move towards centralised systems has been a conscious decision from public and private players, eg. China and Apple. However, it is imperative that policymakers and developers understand that even if they favour openness, some technical choices will turn the system towards centrality, in which case, it will also become more vulnerable to certain types of attacks.

The choice therefore is not only a policy one, but a technological one. Even as a consumer we can advocate an open Internet, but buy Apple products that follow a Jobsian idea of the Internet as a closed and secure system, instead of the open, chaotic and vibrant place that we have grown used to.

Bibliography

"'Rocket Science': The Facts", *Physics World* (June 3 1999), http://physicsworld.com/cws/article/print/1081.

"Buffett Warns on Investment 'Time Bomb'" *BBC News* (4 March, 2003), http://news.bbc.co.uk/1/hi/business/2817995.stm.

"Danish ISPs to Fight the Pirate Bay Block", *Torrent Freak*, (5 February, 2009), http://bit.ly/nND8fF.

"Grandmother piracy lawsuit dropped", *BBC News* (25 September, 2003), http://news.bbc.co.uk/1/hi/entertainment/music/3140160.stm.

"Hundreds killed in Hajj stampede" *BBC News* (12 January 2006), http://news.bbc.co.uk/ 1/hi/world/middle_east/4606002.stm.

"Phishing Indictment Includes More Than 100 Defendants", *Computing Now* (7 October, 2009), http://bit.ly/r5G4n2.

"South Korea's 'Three-Strikes' Law Takes Effect", *Zeropaid* (23 July, 2009), http://bit.ly/ooHU77.

"Spotify: The UK Stats", *Music Ally* (15 October, 2009), http://bit.ly/ao9U66.

"Superpower: Visualising the Internet", *BBC News* (January 2010), http://news.bbc.co.uk/1/ hi/technology/8562801.stm.

"TWAIN'S PLAN TO BEAT THE COPYRIGHT LAW; Will Run Autobiography in New Editions of His Old Works TO PUT PIRATES TO ROUT His Task as a Lobbyist Finished, So He Will Return to New York To-day", *New York Times*, (12 December, 1906), http://bit.ly/9bYh1G.

"UK seven 'were ready to start bombing'", *The Guardian* (21 March, 2006), http://www.guardian.co.uk/terrorism/story/0,,1736228,00.html.

Adamic L and Huberman B, "Zipf's law and the Internet", 3 *Glottometrics* 143 (2002).

Adamic LA and Glance N, "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog", *Proceedings of the 3rd International Workshop on Link Discovery* 36 (2005).

Adamic LA, "The Small World Web", 1696 *Lecture Notes in Computer Science* 443 (1999).

Adams D, *The Hitchhiker's Guide to the Galaxy*, New York: Pocket Books (1981).

Albert R, Jeong H and Barabasi A-L, "Diameter of the World-Wide Web", 401:6749 *Nature* 130 (1999).

Albert R, Jeong H and Barabási A-L, "Error and Attack Tolerance in Complex Networks", 406 *Nature* 378 (2000).

Alderson D and Willinger W, "A Contrasting Look at Self-Organization in the Internet and Next-Generation Communication Networks", 43:7 *Communications Magazine, IEEE* 94 (2005).

Alligood KT, *Chaos: An Introduction to Dynamical Systems*, New York: Springer-Verlag (1997).

Alvestrand H, *A Mission Statement for the IETF*, RFC3935 (2004), http://www.ietf.org/rfc/ rfc3935.txt.

Amor JJ et al, "Measuring Etch: The Size of Debian 4.0", *Debian Conference* (2007), http://bit.ly/pLCyXW.

Andersen PB, "WWW as a Self-organizing System", 5:2 *Cybernetics & Human Knowing* 5 (1998).

Anderson C and Wolff M, "The Web Is Dead. Long Live the Internet", 18.09 *Wired* (August 17, 2010), http://bit.ly/9BpHmC.

Anderson C, "More Long Tail Debate: Mobile Music No, Search Yes", *The Long Tail Blog* (November 8 2008), http://bit.ly/oXfQPr.

Anderson C, *Free: The Future of a Radical Price*, London: Random House Business (2009).

Anderson C, *The Long Tail FAQ*, (2005) http://www.thelongtail.com/about.html.

Anderson C, *The Long Tail: The Revolution Changing Small Markets into Big Business*, New York: Hyperion (2006).

Anderson N, "Pirate Party Hosting Pirate Bay in Pro-P2P Political Gesture", *ArsTechnica* (May 2010), http://bit.ly/d6MmYE.

Anderson P, "Complexity Theory and Organization Science", 10:3 *Organization Science* 216 (1999).

Anti-Phishing Working Group, *Phishing Activity Trends Report* (2009), http://www.antiphishing.org/reports/apwg_report_Q4_2009.pdf.

Ariely D, *Predictably Irrational: The Hidden Forces That Shape Our Decisions*, London: HarperCollins (2009).

Arthur B, Durlauf S and Lane DA, "Introduction: Process and Emergence in the Economy", in Arthur B, Durlauf S and Lane DA (eds), *The Economy as an Evolving Complex System II*, Reading, MA: Addison-Wesley (1997).

Arthur WB, "Inductive Reasoning and Bounded Rationality" 84 *American Economic Review* 406 (1994).

Arthur WB, *The Nature of Technology: What It Is and How It Evolves*, London: Allen Lane (2009).

Asimov I, *Foundation*, London: Octopus Books (1983).

Asimov N, Kim R and Fagan K, "Sabotage attacks knock out phone service", *San Francisco Chronicle* (10 April, 2009), http://bit.ly/q5iOd.

Asser M, "Hajj Perils, Ancient and Modern", *BBC News* (5 March 2001), http://news.bbc.co.uk/1/hi/world/middle_east/1203697.stm.

Ayres I and Baker KK, "A Separate Crime of Reckless Sex", 72 *University of Chicago Law Review* 599 (2005).

Backx P et al, "A Comparison of Peer-To-Peer Architectures", *EURESCOM Summit* (2002), http://goo.gl/UdCEs.

Bak P, Tang C and Wiesenfeld K, "Self-Organized Criticality: An Explanation of 1/*f* Noise", 59 *Physical Review Letters* 381 (1987).

Baldwin R and Cave M, *Understanding Regulation: Theory, Strategy, and Practice*, Oxford: Oxford University Press (1999).

Ball P, "The Physical Modeling of Human Social Systems", 1 *ComPlexUs* 190 (2003).

Ball P, *Critical Mass: How One Thing Leads to Another*, London: Arrow Books (2004).

Barabási A-L, Albert R and Jeong H, "Scale-Free Characteristics of Random Networks: The Topology of the World Wide Web", 281 *Physica A* 69–77 (2000).

Barabási A-L, *Bursts: The Hidden Pattern Behind Everything We Do, New York*, NY: Dutton (2010).

Barabási A-L, Jeong H, Ravasz R, Néda Z, Vicsek T and Schubert A, "On the Topology of the Scientific Collaboration Networks" 311 *Physica A* 590-614 (2002).

Barabási A-L, *Linked: The New Science of Networks*, Cambridge MA: Perseus Pub. (2002).

Barabási A-L, Ravasz E and Vicsek T, "Deterministic Scale-Free Networks", 299(3) *Physica A* 559–564 (2001).

Barlow JP, *A Declaration of the Independence of Cyberspace*, (1996), http://homes.eff.org/ ~barlow/Declaration-Final.html.

Barnett GA and Sung E, "Culture and the Structure of the International Hyperlink Network", 11:1 *Journal of Computer-Mediated Communication* 217 (2005).

Barnhardt S, "The Long Tail of Travel", *Travalution* (19 April, 2007), http://bit.ly/qDF7lV.

Baxter H, "Autopoiesis and the 'Relative Autonomy' of Law", 19:6 *Cardozo Law Review* 1987 (1998).

Beinhocker ED, *The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics*, Boston MA: Harvard Business School Press (2006).

Bellair PE, "Social Interaction and Community Crime: Examining the Importance of Neighbor Networks", 35:4 *Criminology* 677 (1997).

Beller M, "The Sokal Hoax: At Whom Are We Laughing?" 51:9 *Physics Today* 7 (1998).

Benkler Y, "Coase's Penguin, or, Linux and the Nature of the Firm", 112:3 *The Yale Law Journal* 78 (2002).

Benkler Y, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven, CT ; London: Yale University Press (2006).

Benoliel D, "Copyright Distributive Injustice", 10 *Yale Journal of Law & Technology* 45 (2007).

Berger N et al, "On the Spread of Viruses on the Internet", *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithm*s 301 (2005).

Berners-Lee T and Cailliau R, *WorldWideWeb: Proposal for a HyperText Project*, internal CERN memorandum (1990), http://bit.ly/8MKiit.

Bernoff J and Li C, *Groundswell: Winning in a World Transformed by Social Technologies*, Boston MA: Harvard Business School Press (2008).

Bernstein A and Ramchandani R, "Don't Shoot the Messenger! A Discussion of ISP Liability", 1:2 *Canadian Journal of Law and Technology* 1 (2002).

Bernstein RJ, *The Restructuring of Social and Political Theory*, Philadelphia: University of Pennsylvania Press (1978).

Bianconi G and Barabási A-L, "Bose–Einstein Condensation in Complex Networks", 86(24) *Physical Review Letter*s 5632–5635 (2001).

Bildstein B, "Finding and Quantifying Australia's Online Commons", 4:1 *SCRIPTed* 8 (2007).

BitTorrent.org, *Protocol Specifications*, (2006), http://www.bittorrent.org/protocol.html.

Blackburn D, *On-line Piracy and Recorded Music Sales*, Working Paper, Department of Economics, Harvard University (2004), http://www.katallaxi.se/grejer/blackburn/blackburn_fs.pdf.

Boase J, "A Plague of Viruses: Biological, Computer and Marketing", 49:6 *Current Sociology* 39 (2001).

Boffetta G, Paladin G and Vulpiani A, "Strong Chaos without the Butterfly Effect in Dynamical Systems with Feedback", 29:10 *Journal of Physics A: Mathematical and General* 2291 (1996).

Bollobás B and Riordan O, "Robustness and Vulnerability of Scale-Free Random Graphs", 1(1) *Internet Mathematics* 1–35 (2003).

Bonneau J et al, "Eight Friends Are Enough: Social Graph Approximation via Public Listings", *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, Nuremberg, Germany (2009).

Bowrey K, *Law and Internet Cultures*, Cambridge, UK; New York, NY: Cambridge University Press (2005).

Boyle J, "A natural experiment", *Ft.com* (November 2004), http://on.ft.com/rs8CBD.

Boyle J, "Foucault in Cyberspace: Surveillance, Sovereignty, and Hard-Wired Censors", 66 *University of Cincinnati Law Review* 177 (1997).

Boyle J, "The Second Enclosure Movement and the Construction of the Public Domain", 66:1 *Law and Contemporary Problems* 42 (2003).

Boyle J, *Shamans, Software, and Spleens: Law and the Construction of the Information Society*, Cambridge, MA: Harvard University Press (1996).

Boyle J, *The Public Domain: Enclosing the Commons of the Mind*, New Haven, CT; London: Yale University Press (2008).

Brenner S, "At Light Speed: Attribution and Response to Cybercrime/Terrorism/Warfare" 97 *Journal of Criminal Law and Criminology* 379 (2007).

Brin S and Page L, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", 30(1) *Computer Networks and ISDN Systems* 107 (1998).

British Phonographic Industry, *Impact of Illegal Downloading on Music Purchasing*, BPI Paper, http://bit.ly/nxe0aX.

Broder A et al, "Graph Structure in the Web", 22 *Computer Networks* 309 (2000).

Broder A et al, "Graph Structure in the Web", 33 *Computer Networks* 30 (2000).

Brown I, Edwards L and Marsden C, "Information Security and Cybercrime", in Edwards L and Waelde C, *Law and the Internet*, 3rd ed, Oxford ; Portland, OR: Hart (2009).

Brynjolfsson E, Hu YJ and Simester D, *Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales*, SSRN Research Paper Series (2007), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=953587.

Brynjolfsson E, Hu YJ and Smith MD, "From Niches to Riches: The Anatomy of the Long Tail", 47:4 *Emerald Management Reviews* 67 (2006).

Buchan J, *The Authentic Adam Smith: His Life and Ideas*, New York: W.W. Norton (2006).

Buchanan M, *Small World: Uncovering Nature's Hidden Networks*, London: Phoenix (2003).

Burgess J, "'All your chocolate rain are belong to us?' Viral Video, YouTube and the dynamics of participatory culture", in Geert Lovink and Sabine Niederer (eds), *Video Vortex Reader: Responses to YouTube,* Institute of Network Cultures: Amsterdam, (2008).

Callaway DS et al, "Network Robustness and Fragility: Percolation on Random Graphs", 85:25 *Physical Review Letters* 5468 (2000).

Camazine S et al, *Self-Organization in Biological Systems*, Princeton NJ: Princeton University Press (2001).

Capocci A et al, "Preferential Attachment in the Growth of Social Networks: The Internet Encyclopedia Wikipedia", 74:3 *Physical Review E* 036116 (2006).

Carrat F et al, "A 'small-World-Like' Model for Comparing Interventions Aimed at Preventing and Controlling Influenza Pandemics", 4 *BMC Medicine* 26 (2006).

Castells M, *Communication Power*, Oxford ; New York: Oxford University Press (2009).

Castells M, *The Internet Galaxy: Reflections on Internet, Business, and Society*, Oxford: Oxford University Press (2001).

Chandler S, "The Network Structure of Supreme Court Jurisprudence", *International Mathematica Symposium 2005*, The University of Western Australia, Perth (August 2005).

Chattoe E and Hamill H, "It's Not Who You Know – It's What You Know About People You Don't Know That Counts: Extending the Analysis of Crime Groups as Social Networks", 45:6 *The British Journal of Criminology* 860 (2005).

Chau M and Xu J, "Mining Communities and Their Relationships in Blogs: A Study of Online Hate Groups", 65:1 *International Journal of Human-Computer Studies* 57 (2007).

Cheliotis et al, "Taking Stock of the Creative Commons Experiment Monitoring the Use of Creative Commons Licenses and Evaluating Its Implications for the Future of Creative Commons and for Copyright Law", *35th Research Conference on Communication, Information and Internet Policy* (2007), http://bit.ly/olvPPc CreateCommExp.pdf.

Cheliotis G, "From Open Source to Open Content: Organization, Licensing and Decision Processes in Open Cultural Production", 47:3 *Decision Support Systems* 229 (2009).

Cheliotis G, *Remix Culture: Creative Reuse and the Licensing of Digital Media in Online Communities*, Participatory Media Lab Working Paper (2008), http://bit.ly/oxRxz4.

Chen H and Xu J, "Intelligence and Security Informatics", in Cronin B (ed), *Annual Review of Information Science and Technology*, Volume 40, Medford NJ: Information Today (2006).

Chen J, "Introducing Jurisdynamics, a New Blog on Law", *Jurisdynamics Blog* (July 14, 2004), http://bit.ly/4sqPu.

Chen J, "Webs of Life: Biodiversity Conservation as a Species of Information Policy", 89 *Iowa Law Review* 495–608 (2003).

Cheng X, Dale C and Liu J, "Statistics and Social Network of YouTube Videos", *16th International Workshop on Quality of Service* 229 (2008), http://bit.ly/q70Urw.

Chevalier J and Goolsbee A, "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com", 1:2 *Quantitative Marketing and Economics* 203 (2003).

Choi J, Barnett GA and Chon B-S, "Comparing World City Networks: A Network Analysis of Internet Backbone and Air Transport Intercity Linkages", 6:1 *Global Networks* 81 (2006).

Christakis NA and Fowler JH, *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives,* New York: Little, Brown and Co. (2009).

Christin N, Weigend AS and Chuang J, "Content availability, pollution and poisoning in file sharing peer-to-peer networks*", Proceedings of the 6th ACM conference on Electronic Commerce*, Vancouver, Canada (2005).

Churchill W, *The World Crisis, 1911–1918*, Nel Mentor ed, London: New English Library (1968), p75.

Clark D, "The Design Philosophy of the DARPA Internet Protocols", *Symposium Proceedings on Communications Architectures and Protocols*, Stanford, CA (1988).

Clarke R, "Asimov's Laws Of Robotics: Implications for Information Technology" 26:12-27:1 *IEEE Computer* (1993-1994).

Coase R, "The Nature of the Firm", 4:16 *Economica* 386 (1937).

Coase R, "The Problem of Social Cost", 3:1 *The Journal of Law and Economics* 1 (1960).

Cohen B, "Incentives build robustness in BitTorrent", *Proceedings of Workshop on Economics of Peer-to-Peer systems* (2003), http://bit.ly/9SeJIW.

Cohen R et al, "Resilience of the Internet to Random Breakdown", 85 *Physical Review Letters* 4626 (2000).

Coles N, "It's Not What You Know – It's Who You Know That Counts. Analysing Serious Crime Groups as Social Networks", 41:4 *The British Journal of Criminology* 580 (2001).

Connected: The Power of Six Degrees (2008), Science Channel Documentary.

Connolly M and Krueger A, "Rockonomics: The Economics of Popular Music", in Ginsburgh VA and Throsby D, *Handbook on the Economics of Art and Culture*, Amsterdam: Elsevier (2006).

Coombe RJ, *The Cultural Life of Intellectual Properties: Authorship, Appropriation, and the Law*, Durham: Duke University Press (1998).

Corning PA, "The Re-Emergence of "Emergence": A Venerable Concept in Search of a Theory", 7:6 *Complexity* 18 (2002).

Creative Commons, *Defining "Noncommercial": A Study of How the Online Population Understands "Noncommercial" Use*, (2009), http://wiki.creativecommons.org/Defining_Noncommercial.

Crowston K et al, "A Structurational Perspective on Leadership in Free/Libre Open Source Software Teams", *Proceedings of the First International Conference of Open Source Systems* (2005), http://oss2005.case.unibz.it/Papers/68.pdf.

Crowston K et al, "Self-Organization of Teams for Free/Libre Open Source Software Development", 49:6 *Information and Software Technology* 564 (2007).

Cukier KN, "Bandwidth Colonialism? The Implications of Internet Infrastructure on International E-Commerce", *INET99 Conference*, San Jose CA (June 1999), http://bit.ly/qino2l.

Curtis A, *The Trap – What Happened to our Dream of Freedom*, BBC (2007).

Danchev D, "Coordinated Russia vs Georgia cyber attack in progress", *ZDNet* (11 August, 2008), http://zd.net/aKJese.

Darwin C, *The Autobiography of Charles Darwin: 1809–1882*, London: W. W. Norton & Company (1993).

Davis J, "Hackers Take Down the Most Wired Country in Europe", 15.09 *Wired* (21 August, 2007), http://bit.ly/WXG0v.

Davis MD, *Game Theory: A Nontechnical Introduction,* Rev. ed, London: Dover Publications, Constable (1997).

De Vany A and Walls W, "Motion Picture Profit, the Stable Paretian Hypothesis, and The Curse Of The Superstar", 28 *Journal of Economic Dynamics and Control*, 1035 (2004).

Deazley R, *On the Origin of the Right to Copy: Charting the Movement of Copyright Law in Eighteenth-Century Britain (1695–1775)*, Oxford: Hart (2004).

Dellarocas C and Narayan R, *Tall Heads vs. Long Tails: Do Consumer Reviews Increase the Informational Inequality Between Hit and Niche Products?* Robert H. Smith School of Business Research Paper No. 06-056 (2007), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1105956.

Dennett DC, *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, London: Penguin (1996).

Department for Business, Enterprise and Regulatory Reform, *Digital Britain: The Interim Report* (2009), http://www.culture.gov.uk/images/publications/digital_britain_interimreportjan09.pdf.

Department of Business, Innovation and Skills, *Digital Britain Report* (2009), http://interactive.bis.gov.uk/digitalbritain/report/.

Dessai S and Walter M, *Self-Organised Criticality and the Atmospheric Sciences: Selected Review, New Findings and Future Directions*, NCAR Extreme Events workshop, Boulder, CO (June 2000).

Dewan S and Ramaprasad J, *Impact of Blogging on Music Sales: The Long Tail Effect*, Paul Merage School of Business Working Paper (2007), http://www.citi.uconn.edu/cist07/1b.pdf.

Dezso Z and Barabási A-L, "Halting viruses in scale-free networks", 65 *Physical Review E* 055103 (2002).

Dezso Z et al, "General Methods of Statistical Physics – Dynamics of Information Access on the Web", 73:6 *Physical Review. E* 69 (2006)

Di Lorenzo V, "Complexity and Legislative Signatures: Lending Discrimination Laws as a Test Case", 12 *Journal of Law & Politics* 637 (1996).

Di Lorenzo V, "Legislative Chaos: An Exploratory Study", 12 *Yale Law & Policy Review* 425 (1994).

Dibbell J, "A Rape in Cyberspace", *The Village Voice* (21 December, 1993), http://www.juliandibbell.com/texts/bungle_vv.html.

Digital Ethnography, *YouTube Statistics*, (2008), http://ksudigg.wetpaint.com/page/YouTube+Statistics?t=anon.

Domain Tools, *IP Counts by Country* (July 2010), http://www.domaintools.com/internet-statistics/country-ip-counts.html.

Drake C, Oliver J and Koontz E, "Anatomy of a Phishing Email", *Conference on Email and Anti-Spam* (2004), http://bit.ly/coZCQa.

Dugdale JS, *Entropy and its Physical Meaning*, 2nd ed, London: Taylor and Francis (1996).

Duguid F, "Limits of self-organization: Peer production and 'laws of quality'", 11:10 *First Monday* (2006), http://is.gd/ePW1U.

Dulong de Rosnay M, *Creative Commons Licenses Legal Pitfalls: Incompatibilities and Solutions*, Report for the Institute for Information Law, Amsterdam (2010), http://www.ivir.nl/creativecommons/ CC_Licenses_Legal_Pitfalls_2010.pdf

Dumitriu D et al, "Denial-Of-Service Resilience in Peer-To-Peer File Sharing Systems", 33:1 *ACM SIGMETRICS Performance Evaluation Review* 38 (2005).

Durham Y, Hirshleifer J and Smith VL, "Do the Rich Get Richer and the Poor Poorer? Experimental Tests of a Model of Power", 88:4 *The American Economic Review* 14 (1998).

Durrett R, *Random Graph Dynamics*, Cambridge: Cambridge University Press (2007).

Dusollier S, "The Master's Tools v the Master's House: Creative Commons V Copyright", 29:3 *Columbia Journal of Law & the Arts* 271 (2007).

Edwards L and Waelde C, *Online Intermediaries and Liability for Copyright Infringement*, WIPO briefing paper WIPO/IIS/05/1, (2005), http:// papers.ssrn.com/sol3/papers.cfm?abstract_id=1159640.

Edwards L, "Dawn of the Death of Distributed Denial of Service: How to Kill Zombies", 24:1 *Cardozo Arts & Entertainment Law Journal* 23 (2006).

Edwards L, *Cybercrime 2009: The Legal Perspective*, RUSI CyberSecurity Conference, London (October 2009).

Elberse A, "Should You Invest in the Long Tail?" 86:7/8 *Harvard Business Review* 88 (2008).

Elbersei A and Oberholzer-Gee F, *Superstars and Underdogs: An Examination of the Long-Tail Phenomenon in Video Sales*, Harvard Business School Working Paper (2006), http://www.people.hbs.edu/aelberse/papers/hbs_07-015.pdf.

Elkin-Koren N, "What Contracts Can't Do: The Limits of Private Ordering in Facilitating a Creative Commons", 74:2 *Fordham Law Review* 375 (2005).

Elmer-Dewitt P, "First Nation in Cyberspace", 49 *Time Magazine* (3 December, 1996), http://www.chemie.fu-berlin.de/outerspace/Internet-article.html

Emison GA, "The Potential for Unconventional Progress: Complex Adaptive Systems and Environmental Quality Policy", 7 *Duke Environmental Law & Policy Forum* 167 (1996).

Engel C, *Governing the Egalitarians from Without: The Case of the Internet*, Max Planck Preprint Paper 2003/10 (2003).

Engels F and Dühring EK, *Anti-Dühring: Herr Eugen Dühring's Revolution in Science,* London: Progress Publishers, (1954).

Erdós P and Rényi A, "On the Evolution of Random Graphs", 6 *Bulletin of the Institute of International Statistics* 261 (1961).

Euler L, "Seven Bridges of Königsberg", in Newman JR (ed), *The World of Mathematics*, Vol. 1, Mineola, NY: Courier Dover Publications, (2000).

Faloutsos M, Faloutsos P and Faloutsos C, "On Power-Law Relationships of the Internet Topology", 29 *Computer Communications Review* 251 (1999).

Farber DA, "Probabilities Behaving Badly: Complexity Theory and Environmental Uncertainty" 37 *U.C. Davis Law Review* 145 (2003).

Farkas I, Helbing D and Vicsek T, "Social Behaviour: Mexican Waves in an Excitable Medium", 419 *Nature* 131 (2002).

Febvre L and Martin H-J, *The Coming of the Book: The Impact of Printing, 1450–1800*, 2nd Edition, London: Verso Classics (1997).

Fiat A and Saia J, "Censorship Resistant Peer-to-Peer Content Addressable Networks," *Symposium on Discrete Algorithms* (2002).

Findlay-Shirras G and Craig JH, "Sir Isaac Newton and the Currency", 55:218 *The Economic Journal*, (1945).

Fisher WW, *Promises to Keep: Technology, Law, and the Future of Entertainment*, Stanford, CA: Stanford Law and Politics (2004).

Fraud Action UK, *Financial Fraud Action UK Announces Latest Fraud Figures*, Press Release (7 October, 2009), http://bit.ly/nNLwke.

Freeman L, *The Development of Social Network Analysis*, Vancouver: Empirical Press (2006).

Freeman LC, "Centrality in Social Networks: Conceptual Clarification", 1:3 *Social Networks* 215 (1979).

Fuchs C, "The Internet as a Self-Organizing Socio-Technological System", 11:3 *Cybernetics & Human Knowing* 57 (2005).

Gang Y et al, "Epidemic Spread in Weighted Scale-Free Networks", 22:2 *Chinese Physics Letters* 510 (2005).

Gardner M, "Mathematical Games: The Fantastic Combinations of John Conway's New Solitaire Game 'Life'", 223 *Scientific American* 120 (1970).

German DM and Hassan AE, "License Integration Patterns: Addressing License Mismatches in Component-Based Development", *International Conference on Software Engineering* (2009).

Geu TE, "Chaos, Complexity, and Coevolution: The Web of Law, Management Theory, and Law Related Services at the Millennium", 65 *Tennessee Law Review* 925 (1998).

Geu TE, "The Tao of Jurisprudence: Chaos, Brain Science, Synchronicity, and the Law", 61 *Tennessee Law Review* 933 (1994).

Ghosh RA, *Study on the Economic Impact of Open Source Software on Innovation and the Competitiveness of the Information and Communication Technologies (ICT) Sector in the EU*, European Commission Report ENTR/04/112 (2006).

Giles DE, "Increasing Returns to Information in the US Popular Music Industry", 14:4 *Applied Economics Letters* 327 (2007).

Giot L et al, "A Protein Interaction Map of Drosophila Melanogaster", 302:5651 *Science* 1727 (2003).

Girvan M and Newman MEJ, "Community Structure in Social and Biological Networks", 99:12 *Proceedings of the National Academy of Sciences* 7821 (2002).

Gladwell M, *The Tipping Point: How Little Things Can Make a Big Difference*, London: Abacus (2002).

Goffman C, "And What Is Your Erdos Number?" 76:7 *The American Mathematical Monthly* 791 (1969).

Goldberg DS, "And the Walls Came Tumbling Down: How Classical Scientific Fallacies Undermine the Validity of Textualism and Originalism", 39 *Houston Law Review* 463 (2002).

Goldsmith JL and Wu T, *Who Controls the Internet? Illusions of a Borderless World*, Oxford: Oxford University Press (2006).

Goldstein J, "Emergence as a Construct: History and Issues", 1:1 *Emergence* 49 (1999).

Goldstein MP, "Service Provider Liability for Acts Committed by Users: What You Don't Know Can Hurt You", 18:3 *The John Marshall Journal of Computer & Information Law* 52 (2000).

González MC, Hidalgo CA and Barabási A-L, "Understanding Individual Human Mobility Patterns", 453 *Nature* 779 (2008).

Gottfried K, "Opinion – Was Sokal's Hoax Justified?" 50:1 *Physics Today* 5 (1997).

Gould SJ and Eldredge N, "Punctuated Equilibrium Comes of Age", 366 *Nature* 223 (1993).

Gowers A, *Gowers Review of Intellectual Property*, HM Treasury, (2006).

Granic I and Lamey V, "The Self-Organization of the Internet and Changing Modes of Thought", 18:1 *New Ideas in Psychology* 93 (2000).

Gray E et al, "Trust Propagation in Small Worlds", 2692 *Lecture Notes in Computer Science* 239 (2003).

Greenemeier L, "How Was Egypt's Internet Access Shut Off?" *Scientific American* (28 January, 2011), http://goo.gl/CCE08.

Greenleaf G, "An Endnote on Regulating Cyberspace: Architecture vs Law?" 21:2 *University of New South Wales Law Journal* (1998), http://www.austlii.edu.au/au/journals/UNSWLJ/1998/52.html.

Greenwald B and Stiglitz JE, "Externalities in Economies with Imperfect Information and Incomplete Markets", 101 *Quarterly Journal of Economics* 229 (1986).

Grewal R, Lilien GL and Mallapragada G, "Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems", 52:7 *Management Science* 1043 (2006).

Guadamuz A, "Censorship UK", *TechnoLlama Blog* (8 December, 2008), http://www.technollama.co.uk/censorship-uk

Guadamuz A, "Critique of the ICC's Report on the Digital Economy in Europe", *TechnoLlama* (19 March, 2010), http://bit.ly/bIFGQy.

Guadamuz A, "Free and Open Source Software", in Edwards L and Waelde C (eds), *Law and the Internet*, 3rd ed, Oxford: Hart (2009).

Guadamuz A, "How Will ACTA Affect UK Copyright Law?" *TechnoLlama* (July 15, 2010), http://bit.ly/dlEGfz.

Guadamuz A, "ISP Liability to Get ECJ Hearing", *TechnoLlama* (12 February, 2010), http://bit.ly/pYOfK3.

Guadamuz A, "Open Science: Open Source Licences for Scientific Research", 7:2 *North Carolina Journal of Law and Technology* 321 (2006).

Guadamuz A, "Public Rights Licences", *WorldLII* (2008), http://www.worldlii.org/int/other/PubRL/.

Guadamuz A, "Scale-Free Law: Network Science and Copyright", 70:4 *Albany Law Review* 1297 (2008).

Guadamuz A, "The License/Contract Dichotomy in Open Licenses: A Comparative Analysis", 30:2 *University of La Verne Law Review* 101 (2009).

Guadamuz A, "The open Web vs the closed Internet", *TechnoLlama Blog* (22 August, 2010), http://bit.ly/8YHY3D.

Guadamuz A, "The Software Patent Debate", 1(3) *Journal of Intellectual Property Law & Practice* 196 (2006).

Guadamuz A, "Viral Contracts or Unenforceable Documents? Contractual Validity of Copyleft Licenses", 26:8 *European Intellectual Property Review* 331 (2004).

Guadamuz A, "Wikileaks: So, This Is What Cyberwar Looks Like", *TechnoLlama* (3 December, 2010), http://goo.gl/fv2fm.

Guernsey L, "Welcome to the Web. Passport, Please?" *New York Times* (15 March 15, 2001), http://nyti.ms/pfgrVQ.

Guo H and Cheng H, "Computer Virus Propagation in Social Networks", *ICIS 2007 Proceedings* (2007), http://aisel.aisnet.org/icis2007/124.

Gustin S, "Social Media Sparked, Accelerated Egypt's Revolutionary Fire", *Wired* (11 February, 2011), http://goo.gl/bz6J2.

Habermas J, *Knowledge and Human Interests*, 2nd [English] ed, London: Heinemann Educational (1978), Chapter 3.

Hall BH, Jaffe AB and Tratjenberg M, *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools*, NBER Working Paper 8498 (2001).

Ham F and van Wijk J, "Interactive Visualization of Small World Graphs", *Proceedings of the IEEE Symposium on Information Visualization* (2004).

Hamelink CJ, "Did WSIS Achieve Anything at All?" 66:3-4 *International Communications Gazette* 281 (2004).

Hars A and Ou S, "Working For Free? Motivations of Participating in Open Source Projects", *Proceedings of the 34th Annual Hawaii International Conference on System Sciences* (2001).

Hart K, *Postmodernism: A Beginner's Guide*, Oxford: Oneworld (2004).

Hayes AW, "An Introduction to Chaos and Law", 60 *University of Missouri at Kansas City Law Review* 751 (1992).

Haythornthwaite C, "Social Networks and Internet Connectivity Effects", 8:2 *Information, Communication & Society* 125 (2005).

Helbing D et al, "How individuals learn to take turns: Emergence of alternating cooperation in a congestion game and the prisoner's dilemma", 8 *Advances in Complex Systems* 871 (2005).

Helbing D, Johansson A and Al-Abideen HZ, "The Dynamics of Crowd Disasters: An Empirical Study", 75 *Physical Review E* 046109 (2007).

Henderson LF, "The Statistics of Crowd Fluids", 229 *Nature* 331 (1971).

Herrera L, "Egypt's Revolution 2.0: The Facebook Factor", *Jadaliyya* (12 February, 2011), http://goo.gl/qlkEd.

Hettinger EC, "Justifying Intellectual Property", 18:1 *Philosophy and Public Affairs* 31 (1989).

Hirshleifer J, *The Dark Side of the Force: Economic Foundations of Conflict Theory*, Cambridge: Cambridge University Press (2001).

Hobbes T, *Leviathan*, New York: Barnes & Noble Publishing (2004).

Hofstadter DR, *Godel, Escher, Bach: An Eternal Golden Braid*, 20th-anniversary ed, London: Penguin (2000).

Hofstadter DR, *Metamagical Themas: Questing for the Essence of Mind and Pattern*, London: Penguin (1986).

Huang S-Y et al, "Network-Induced Nonequilibrium Phase Transition in the 'Game Of Life'", 67:2 *Physical Review E* 026107 (2003).

Huberman BA and Adamic LA, "Growth Dynamics of the World-Wide Web", 401:6749 *Nature* 131 (1999).

Huberman BA and Adamic LA, "Power-Law Distribution of the World Wide Web", 287 *Science* 2115 (2000).

Huberman BA, *The Laws of the Web: Patterns in the Ecology of Information*, Cambridge, MA: MIT Press (2001).

Hucknall M, "Fundamental socialism", *The Guardian* (23, November 2006), http://www.guardian.co.uk/ commentisfree/2006/nov/23/comment.music.

Hunt F and Johnson P, "On the Pareto distribution of Sourceforge projects", *Proceedings of the Open Source Software Development Workshop* (2002).

ICANN, *Geographic Distribution Of Registrars* (2010), http://forms. icann.org/idashboard/public/.

IFPI, *IFPI Digital Music Report*, (2010). http://www.ifpi.org/content/ library/DMR2010.pdf.

Intelligence and Security Committee, *Could 7/7 Have Been Prevented?* Cabinet Office Review of the Intelligence on the London Terrorist Attacks on 7 July 2005 (2008), http://goo.gl/9kTKV

Internet Engineering Task Force, *Requirements for Internet Hosts: Communication Layers*, RFC 1122 (1989), http://bit.ly/qcS6No.

Internet World Stats, *World Internet Usage and Population Statistics* (2010), http://www.internetworldstats.com/stats.htm.

Ioannides Y and Overman HG, "Zipf's Law for Cities: An Empirical Examination", 33:2 *Regional Science and Urban Economics* 127 (2003).

Jaffe A and Lerner J, *Innovation and Its Discontents*, Princeton NJ: Princeton University Press (2004).61–69.

Jagatic TN et al, "Social Phishing", 50:10 *Communications of the ACM* 94 (2007).

Jakobsson M, "Modelling and Preventing Phishing Attacks", *Lecture Notes In Computer Science* (2005), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.64.1926.

Janczewski L and Colarik AM, *Cyber Warfare and Cyber Terrorism,* Hershey: Information Science Reference (2008).

Jeong H et al, "The Large-Scale Organization of Metabolic Networks", 407 *Nature* 651–654 (2000).

Jin X et al, "A Topological Analysis of the Open Source Software Development Community", *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, (2005).

Johnson DR and Post DG, "Law and Borders: The Rise of Law in Cyberspace", 48 *Stanford Law Review* 1367 (1996).

Johnson S, *Emergence: The Connected Lives of Ants, Brains, Cities and Software*, London: Penguin (2002).

Jondet N, "The French Copyright Authority (HADOPI), the Graduated Response And the Disconnection Of Illegal File-Sharers", *BILETA 2010*, University of Vienna (March 2010).

Joseph A, *Cybercrime Definition*, Computer Crime Research Center Papers, http://www.crime-research.org/articles/joseph06/.

Jovanović M, Annexstein F and Berman K, "Modeling Peer-To-Peer Network Topologies Through 'small-World' Models And Power Laws", *IX Telecommunications Forum* (2001).

Juran JM, "The Non-Pareto Principle: Mea Culpa", *Quality Progress* (1975).

Kades E, "The Laws of Complexity and the Complexity of Laws: The Implications of Computational Complexity Theory for the Law", 49 *Rutgers Law Review* 403, 452 (1997).

Kansa EC, Schultz J and Bissell AN, "Protecting Traditional Knowledge and Expanding Access to Scientific Data: Juxtaposing Intellectual Property Agendas via a Model", 12:3 *International Journal of Cultural Property* 285 (2005).

Karlin J, Forrest S and Rexford J, *Nation-State Routing: Censorship, Wiretapping, and BGP*, arXiv Working Paper (2009), http://arxiv.org/ abs/0903.3218v1.

Katz DM and Stafford DK. "Hustle and Flow: A Social Network Analysis of the American Federal Judiciary", 71:3 *Ohio State Law Journal* (2010).

Katz DM, Stafford DK and Provins E, "Social Architecture, Judicial Peer Effects and the 'Evolution' of the Law: Toward a Positive Theory of Judicial Social Structure", 24 *Georgia State University Law Review* 977 (2008).

Kauffman SA and Weinberger EW, "The NK model of rugged fitness landscapes and its application to maturation of the immune response", 141:2 *Journal of Theoretical Biology* 211 (1989).

Kauffman SA, "Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets", 22 *Journal of Theoretical* Biology 437 (1969).

Kauffman SA, "The Sciences of Complexity and 'Origins of Order'", 2 *Proceedings of the Biennial Meeting of the Philosophy of Science Association, V Symposia and Invited Papers* 299 (1990).

Kauffman SA, *At Home in the Universe: The Search for Laws of Self-Organization and Complexity,* Oxford: Oxford University Press (1995).

Keyani P, Larson B and Senthil M, "Peer Pressure: Distributed Recovery from Attacks in Peer-to-Peer Systems", *Proceedings of the International Workshop on Peer-to-Peer Computing* 306 (2002).

Khambatti M, Ryu K and Dasgupta P, "Structuring Peer-to-Peer Networks using Interest-Based Communities", *International Workshop on Databases, Information Systems and Peer-to-Peer Computing*, Humboldt University, Berlin, Germany (September 2003).

King M, "The 'Truth' About Autopoiesis", 20:2 *Journal of Law and Society* 218 (1993).

Kirkpatrick M, "Google CEO Schmidt: 'People Aren't Ready for the Technology Revolution'", *ReadWriteWeb Blog* (4 August, 2010), http://is.gd/eNTTN.

Kleinfeld J, "The Small World Problem", 39 *Society* 61-66 (2002).

Kochen M, *The Small World*, Norwood, NJ:Ablex Publishing (1989).

Koertge N, *A House Built on Sand: Exposing Postmodernist Myths About Science*, Oxford: Oxford University Press (1998).

Kolmogorov AN, "Combinatorial foundations of information theory and the calculus of probabilities", 38:4 *Russian Mathematical Surveys* 29 (1983).

Krapivsky PL, Rodgers GJ and Redner S, "Degree Distributions of Growing Networks" 86(23) *Physical Review Letters* 5401–5404 (2001).

Krotoski A, *Social Influence in Second Life: Social Network and Social Psychological Processes in the Diffusion of Belief and Behaviour on the Web*, Ph.D. thesis, University of Surrey (2009), http://alekskrotoski.com/media_files/SocialInfluenceInSecondLife.pdf.

Krugman P, *Incidents from my Career*, http://web.mit.edu/krugman/www/incidents.html.

Landes WM and Posner RA, "An Economic Analysis of Copyright Law", 18:2 *The Journal of Legal Studies* 325 (1989).

Lanzara GF and Morner M, "The Knowledge Ecology of Open-Source Software Projects", *Proceedings of the 19th EGOS Colloquium*, Copenhagen (2003).

Latora V and Marchiori M, "A Measure of Centrality Based on Network Efficiency", 9 *New Journal of Physics* 188 (2007).

Latour B, *Reassembling the Social: An Introduction to Actor-Network-Theory,* Oxford: Oxford University Press (2005).

Lee E, "Warming up to User-Generated Content", 5 *University of Illinois Law Review* 1459 (2008).

Lehrer J, "The Buddy System: How Medical Data Revealed Secret to Health and Happiness", 17.10 *Wired* (September 2009).

Lehrer J, *The Decisive Moment: How the Brain Makes up Its Mind*, Edinburgh: Canongate (2009).

Leibniz GWF, "Freedom and Possibility", in *Philosophical Essays,* Indianapolis, IN: Hackett Publishing (1989).

Leibniz GWF, *Monadology and Other Philosophical Essays*, Indianapolis, IN: Bobbs-Merrill Company (1965).

Leigh D, "How 250,000 US Embassy Cables Were Leaked", *The Guardian* (28 November, 2010), http://goo.gl/Azhcm.

Leiner B et al, *A Brief History of the Internet*, Internet Society Paper (2000).

Lessig L, *Code Version 2.0*, 2nd ed, New York: Basic Books (2006).

Lessig L, *Code: and Other Laws of Cyberspace,* New York, NY: Basic Books (1999).

Lessig L, *Remix: Making Art and Commerce Thrive in the Hybrid Economy*, London: Bloomsbury Academic (2008).

Lessig L, *The Future of Ideas: The Fate of the Commons in a Connected World,* New York: Random House (2001).

Levinson P, *Digital McLuhan: A Guide to the Information Millennium*, London: Routledge (2001).

Li L et al, "Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications", 2:4 *Internet Mathematics* 431 (2005).

Litman J, *Digital Copyright: Protecting Intellectual Property on the Internet*, Amherst, NY: Prometheus Books (2001).

Locke J, "Some Considerations of the Consequences of the Lowering of Interest and the Raising the Value of Money", in Medema SG and Samuels WJ (eds), *The History of Economic Thought: A Reader*, London: Routledge (2003).

Loguinov D et al, "Graph-Theoretic Analysis of Structured Peer-to-Peer Systems: Routing Distances and Fault Resilience", *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (2003).

LoPucki LM, "The Systems Approach to Law", 82 *Cornell Law Review* 479, 480-82 (1997).

Luhmann N, "Law as a Social System", 83 *Northwestern University Law Review* 136 (1988).

Luhmann N, *Essays on Self-Reference*, New York: Columbia University Press (1990).

Luhmann N, *Social Systems,* Stanford, CA: Stanford University Press (1995).

Mackaay E, "History of Law and Economics", in Bouckaert B and De Geest G, *Encyclopedia of Law & Economics*, Cheltenham UK; Northampton, MA: Edward Elgar (2000).

Madey G, Freeh V and Tynan R, "Modeling the Free/Open Source Software Community: A Quantitative Investigation", in Koch S, *Free/Open Source Software Development*, Hershey, PA: Idea Group (2005).

Mahadevan P et al, "The Internet AS-Level Topology: Three Data Sources and One Definitive Metric", 36:1 *Computer Communications Review* 17 (2006).

Maillart T et al, "Empirical Tests of Zipf's Law Mechanism in Open Source Linux Distribution", 101:21 *Physical Review Letters* 218701 (2008).

Mandelbrot B, "How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension", 156:3775 *Science* 636 (1967).

Mandelbrot B, "The Pareto-Lévy Law and the Distribution of Income", 1:2 *International Economic Review* 79 (1960).

Marco M, "RIAA Bullies College Students with P2PLawsuits.com", *The Consumerist* (1 March, 2007), http://bit.ly/pANX6p.

Markoff J, "Georgia Takes a Beating in the Cyberwar with Russia", *The New York Times* (11 August, 2008), http://nyti.ms/HYPeX.

Marsden CT, *Net Neutrality: Towards a Co-Regulatory Solution,* London: Bloomsbury Academic (2010).

Maturana HR and Varela FJ, *Autopoiesis and Cognition: The Realization of the Living*, London: Reidel (1980).

Matwyshyn AM, "Organizational Code: A Complexity Theory Perspective on Technology and Intellectual Property Regulation", 11 *Journal of Technology Law & Policy* 13 (2006).

Maurushat A, "Zombie Botnets", 7:2 *SCRIPTed* 370 (2010).

McGloin JM, "Policy and Intervention Considerations of a Network Analysis of Street Gangs", 4:3 *Criminology & Public Policy* 607 (2005).

McLuhan M, *The Gutenberg Galaxy; the Making of Typographic Man*, Toronto: University of Toronto Press (1962).

Mechanical-Copyright Protection Society, *Directors' Report and Accounts* (2004), http://www.mcps-prs-alliance.co.uk/aboutus/.

Miles E, "In re Aimster & MGM, Inc. v. Grokster, Ltd.: Peer-to-Peer and the Sony Doctrine", 19 *Berkeley Technology Law Journal* 21 (2004).

Milgram S, "The Small World Problem", 2 *Psychology Today* 60–67 (1967).

Milgram S, *The Individual in a Social World*, New York: McGraw-Hill (1992).

Mill JS, *A* System *of Logic Ratiocinative and Inductive*, London: John W. Parker and Son (1872).

Miller JH and Page SE, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, Princeton, NJ: Princeton University Press (2007).

Mingers J, *Self-Producing Systems: Implications and Applications of Autopoiesis*, London: Plenum Press (1995).

Minowitz P, "Adam Smith's Invisible Hands" 1:3 *Economy Journal Watch* 381 (2004).

Mockapetris PV and Dunlop KJ, "Development of the Domain Name System", 25:1 *ACM SIGCOMM Computer Communication Review* 112 (1988).

Moglen E, *The dotCommunist Manifesto*, (2003), http://emoglen.law.columbia.edu/publications/dcm.html.

Moody G, *Rebel Code: Linux and the Open Source Revolution*, London: Penguin (2002).

Mook N, "IFPI Sues 8,000 P2P File Swappers", *BetaNews* (17 October, 1006), http://www.betanews.com/article/IFPI-Sues-8000-P2P-File-Swappers/1161101823.

Moreno Y, Gomez JB and Pacheco AF, "Instability of Scale-Free Networks under Node-Breaking Avalanches" 58(4) *Europhysics Letters* 630–636 (2002).

Mueller M, *Ruling the Root*, Boston, MA: MIT Press (2004).

Murray A, "Conceptualising the Post-Regulatory (Cyber)State", in Brownsword R and Yeung K (eds), *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes,* Oxford: Hart (2008).

Murray AD, *The Regulation of Cyberspace: Control in the Online Environment*, Milton Park, Abingdon UK; New York, NY: Routledge-Cavendish (2007).

Naughton J, *A Brief History of the Future: The Origins of the Internet*, London: Phoenix (2000).

Negroponte N, "A Bill of Writes", *Wired 3.05* (May 1995), http://web.media.mit.edu/~nicholas/Wired/WIRED3-05.html.

Nevins CH and Keeble A, *Emusic Sales Data Supports "Long Tail" Concept*, Press Release (15 January 2009), http://www.emusic.com/about/pr/PR2009115.html.

Newman F, "One dogma of dialectical materialism", 1 *Annual Review of Critical Psychology* 83 (1999).

Newman MEJ, "Clustering and Preferential Attachment in Growing Networks", 64:2 *Physical Review E* 025102 (2001).

Newman MEJ, "Power Laws, Pareto Distributions and Zipf's Law", 46:5 *Contemporary Physics* 323 (2005).

Newman MEJ, "The Structure of Scientific Collaboration Networks", 98:2 *Proceedings of the National Academy of Sciences of the United States of America* 6 (2001).

Newman MEJ, Barabási A-L and Watts D J, *The Structure and Dynamics of Networks*, Princeton, NJ; Oxford: Princeton University Press (2006).

Nielsen, "Twitter's Tweet Smell Of Success", *Nielsen Wire* (18 March, 2009), http://goo.gl/vBL2m.

Nielsen, *Global Faces and Networked Places*, Nielsen report (March 2009), http://is.gd/eTSFc.

O'Reilly T, "What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software", 1 *Communications & Strategies* 17 (2007).

Oberholzer-Gee F, Strumpf K, "The effect of file sharing on record sales: An empirical analysis", 115:1 *Journal of Political Economy* 1 (2007).

Oestreicher-Singer G and Sundararajan A, *Recommendation Networks and the Long Tail of Electronic Commerce*, Wharton Working Papers (2009), http://papers.ssrn.com/sol3/ papers.cfm?abstract_id=1324064.

Offman C, "Fight the Future!" 8:12 *Wired Magazine* (2000).

Open Knowledge Foundation, *Freedom Defined*, (2008), http:// freedomdefined org/Definition.

Opsahl T, Agneessens F and Skvoretz J, "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths", 32:3 *Social Networks* 245 (2010).

Pareto V, *Manual of Political Economy,* New York: Augustus M. Kelly Publishers (1971).

Park K and Lee H, "On the Effectiveness of Route-Based Packet Filtering for Distributed DoS Attack Prevention in Power-Law Internets", 31:4 *SIGCOMM Computer Communications Review* 15 (2001).

Parker A, *The True Picture of Peer-to-Peer Filesharing*, Report by Cache Logic (2005), http://www.cachelogic.com/research/p2p2004.php.

Particularly in: Marx K, *Wage Labour and Capital*, Whitefish, MT: Kessinger Publishing, (2004).

Percacci R and Vespignani A, "Scale-Free Behavior of the Internet Global Performance", 32 *European Physical Journal B* 411 (2003).

Perens B, "A Cyber-Attack on an American City" *Silicon Alley Insider* (25 April, 2009), http://www.businessinsider.com/a-cyber-attack-on-an-american-city-2009-4.

Perkins F, *Leibniz and China: A Commerce of Light*, Cambridge: Cambridge University Press (2004).

Phillips DE, *The Software License Unveiled: How Legislation by License Controls Software Access*, Oxford: Oxford University Press (2009).

Phillips JD, "Divergence, Sensitivity, and Nonequilibrium in Ecosystems", 36:4 *Geographical Analysis* 369 (2004).

Picker RC, "Simple Games in a Complex World: A Generative Approach to the Adoption of Norms", 64 *University of Chicago Law Review* 1225 (1997).

Pink DH, *Drive: The Surprising Truth About What Motivates Us*, New York, NY: Riverhead Books (2009).

Porter H, "Google is just an amoral menace?" *The Observer* (5 April, 2009), http://www.guardian.co.uk/commentisfree/2009/apr/05/google-internet-piracy.

Post D and Johnson D, *"*Chaos Prevailing on Every Continent": A New Theory of Decentralized Decision-Making in Complex Systems", 73 *Chicago-Kent Law Review* 1055 (1999).

Post DG and Eisen MB, "How Long is the Coastline of the Law? Thoughts on the Fractal Nature of Legal Systems", 29 *Journal of Legal Studies* 545 (2000).

Pouwelse JA et al, "Pirates and Samaritans: A Decade of Measurements on Peer Production and Their Implications for Net Neutrality and Copyright", 32:11 *Telecommunications Policy* 701 (2008).

Prehofer C and Bettstetter C, "Self-Organization in Communication Networks: Principles and Design Paradigms", 43:7 *Communications Magazine, IEEE* 78 (2005).

Prigogine I and Stengers I, *Order Out of Chaos*, New York: Bantam (1984).

Qiu D and Sang W, "Global Stability Of Peer-To-Peer File Sharing Systems", 31:2 *Computer Communications* 212 (2008).

Quigley R, "Facebook Privacy Fears for 100m Users as Their Personal Details Are Published on File-Sharing Site", *Mail Online* (29 July, 2010), http://is.gd/eTGqX.

Radin M, "Humans, Computers, and Binding Commitment", 75 *Indiana Law Journal* 1125 (2000).

RAND, *Paul Baran and the Origins of the Internet*, http://www.rand.org/about/history/baran.html.

Rapoport A and Horvarth V, "A Study of a Large Sociogram" 6 *Behavioral Science* 279 (1961).

Ravasz E and Barabási A-L, "Hierarchical Organization in Complex Networks", 67 *Physical Review E* 026112 (2003).

Raymond ES, *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, Sebastopol, CA: O'Reilly Media (2001).

Redner S, "How Popular Is Your Paper? An Empirical Study of the Citation Distribution", 4:2 *The European Physical Journal B* 131 (1998).

Reed WJ, "The Pareto, Zipf and Other Power Laws", 74(1) *Economics Letters* 15 (2001).

Reidenberg J, "Lex Informatica: The Formulation of Information Policy Rules through Technology", 76 *Texas Law Review* 553 (1998)

Reynolds C, "Flocks, Herds and Schools: A Distributed Behavioral Model", *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques* 25 (1987).

Rietjens B, "Give and Ye Shall Receive! The Copyright Implications of BitTorrent", 2:3 *SCRIPTed* 364 (2005).

Ripenau M, Foster I and Iamnitchi A, "Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design", 6:1 *IEEE Internet Computing Journal* (2002).

Roe MJ, "Chaos and Evolution in Law and Economics", 109 *Harvard Law Review* 641 (1995).

Rosen K and Resnick M, "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy", 8:2 *Journal of Urban Economics* 165 (1980).

Rosen LE, *Open Source Licensing: Software Freedom and Intellectual Property Law*, Upper Saddle River, NJ: Prentice Hall PTR (2004).

Rosen S, "The Economics of Superstars", 71 *American Economic Review* 845 (1981).

Ruhl JB, "The Fitness of Law: Using Complexity Theory to Describe the Evolution of Law and Society and its Practical Meaning for Democracy", 49 *Vanderbilt Law Review* 1407 (1996).

Ryan B, "Communication Breakdown: The Recording Industry's Pursuit of the Individual Music User, a Comparison of U.S. and E.U. Copyright Protections for

Internet Music File Sharing", 25 *Northwestern Journal of International Law & Business* 229 (2004).

Saia J et al, "Dynamically Fault-Tolerant Content Addressable Networks," *IPTPS* (March 2002).

Saroiu S, Gummadi KP and Gribble SD, "A Measurement Study of Peer-to-Peer File Sharing Systems", *Proceedings of Multimedia Computing and Networking 2002* (2002)

Saroiu S, Gummadi KP and Gribble SD, "Measuring and Analyzing the Characteristics of Napster And Gnutella Hosts", 9:2 *Multimedia Systems* 170 (2003).

Savirimuthu A and Savirimuthu J, "Identity Theft and Systems Theory: The Fraud Act 2006 in Perspective", 4:4 *SCRIPTed* 436 (2007).

Scherer FM and Harhoff D, "Technology policy for a world of skew-distributed outcomes", 29:4-5 *Research Policy* 559 (2000).

Scherer FM, "The Size Distribution of Profits from Innovation", 86:49/50 *Annales d'Économie et de Statistique* 495 (1998).

Schoder D and Fischbach K, "Core Concepts in Peer-to-Peer (P2P) Networking", in Subramanian R and Goodman B (eds), *P2P Computing: The Evolution of a Disruptive Technology*, Hershey PA: Idea Group Inc. (2005).

Schroeder R, *The Social Life of Avatars: Presence And Interaction in Shared Virtual Environments*, London: Springer-Verlag (2001).

Schulman LS and Seiden PE, "Statistical Mechanics of a Dynamical System Based on Conway's Game of Life", 19:3 *Journal of Statistical Physics* 293 (1978).

Scott JP, "Social Network Analysis", 22:1 *Sociology* 109 (1988).

Scott RE, "Chaos Theory and the Justice Paradox", 35 *William & Mary Law Review* 329, (1993).

Sen S and Wang J, "Analyzing Peer-To-Peer Traffic across Large Networks", *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement* (2002).

Shannon CE, "The Mathematical Theory of Communication", 27 *Bell System Technology Journal* 379 (1948).

Shavitt Y and Shir E, "DIMES: Let the Internet Measure Itself", 35:5 *Computer Communication Review* 71 (2005).

Shavitt Y and Weinsberg U, "Quantifying the Importance of Vantage Points Distribution in Internet Topology Measurements", *Proceedings of INFOCOM 2009 IEEE* 792 (2009).

Shenker S, "Fundamental Design Issues for the Future Internet", 13:7 *IEEE Journal on Selected Areas in Communications* 1176 (1995).

Shiels M, "Web Scam Hits iTunes and Paypal Users", *BBC News* (24 August, 2010), http://www.bbc.co.uk/news/technology-11065301.

Shilts R, *And the Band Played On: Politics, People, and the Aids Epidemic*, New York, NY: Penguin Books (1989).

Sifry D, "State of the Blogosphere August 2005 Part 5: The A-List and the Long Tail", *Sifry's Alerts Blog* (10 August, 2005), http://www.sifry.com/alerts/archives/2005_08.html.

Simon HA and Bonini CP, "The Size Distribution of Business Firms" 48:4 *The American Economic Review* 607 (1958).

Sinclair D, "Self-Regulation versus Command and Control – Beyond False Dichotomies", 19 *Law & Policy* 529 (1997).

Smith A et al, *An Inquiry into the Nature and Causes of the Wealth of Nations*, Indianapolis, IN: Liberty Press (1981).

Smith A, *The Theory of Moral Sentiments*, New York: A.M. Kelley (1966), IV.I.10.

Smith MR and Marx L, *Does Technology Drive History? The Dilemma of Technological Determinism*, Cambridge, MA; London: MIT Press (1994).

Smith S, "From Napster to Kazaa: The Battle over Peer-to-Peer Filesharing Goes International", *Duke Law & Technology Review* 8 (2003).

Smith TA, "The Web of Law", 44 *San Diego Law Review* 309 (2007).

Sokal AD, "Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity", 14:1-2 *Social Text* 217 (1996).

Sola-Pool I and Kochen M, "Contacts and Influence", 1 *Social Networks* 5 (1978).

Solomonoff R and Rapoport A, "Connectivity of Random Nets", 13 *Bulletin of Mathematical Biophysics* 107 (1951).

Spulber D and Yoo CS, "On the Regulation of Networks as Complex Systems: A Graph Theory Approach", 99 *Northwestern University Law Review* 1687–1722 (2005).

Stigler SM, *Statistics on the Table*, Boston: Harvard University Press (1999).

Stone M, "Formalism", in Coleman JL, Shapiro S and Himma KE (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law*, Oxford: Oxford University Press (2004).

Strahilevitz LJ, "A Social Networks Theory of Privacy", 72 *University of Chicago Law Review* 919–988 (2005).

Strandburg KJ, "Law and the Science of Networks: An Overview and an Application to the 'Patent Explosion'", 21 *Berkeley Technology Law Journal* 1293 (2007).

Streeter CL and Gillespie DF, "Social Network Analysis", 16:1 *Journal of Social Service Research* 201 (1993).

Strogatz S, "Exploring Complex Networks", 410 *Nature* 268 (2001).

Strogatz SH, *Sync: The Emerging Science of Spontaneous Order*, New York: Hyperion (2003).

Sunstein C, "Social Norms and Social Roles", 96 *Columbia Law Review* 903 (1996).

Sunstein CR, *Infotopia: How Many Minds Produce Knowledge*, Oxford: Oxford University Press (2008).

Surowiecki J, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few*, London: Abacus (2005).

Svolkia J, "Forget Citibank – Borrow from Bob" *Harvard Business Review* (2009), http://hbr.org/web/2009/hbr-list/forget-citibank-borrow-from-bob.

Taleb N, *The Black Swan: The Impact of the Highly Improbable*, London: Allen Lane (2007).

Tan TF and Netessine S*, Is Tom Cruise Threatened? Using Net IX Prize Data to Examine the Long Tail of Electronic Commerce*, Wharton Working Paper (2009), http://bit.ly/1WAcr.

Tancer B, *Measuring Web 2.0 Consumer Participation*, Hitwise US Research Note (2007), http://goo.gl/EsCfJ.

Tate R, "Steve Jobs Offers World 'Freedom From Porn'", *Gawker: Valleywag Blog* (May 15, 2010), http://goo.gl/hzh3.

Taylor G, "Never Mind The Billshock", *British Phonographic Industry Blog* (25 January, 2010), http://goo.gl/7Zq5g.

Technorati, *State of the Blogosphere 2008*, http://bit.ly/zoVhS.

Telco 2.0, *The "Long Tail" Interrogated*, (12 November, 2008), http://www.telco2.net/blog/2008/11/exclusive_interview_will_page.html.

TERA Consultants, *Building a Digital Economy: The Importance of Saving Jobs in the EU's Creative Industries*, International Chamber of Commerce Report (2010), http://www. iccwbo.org/bascap/id35360/index.html.

Teubner G and Bankowski Z, *Law as an Autopoietic System*, Oxford: Blackwell Publishers (1993).

Thelwall M, "Social Networks, Gender, and Friending: An Analysis of Myspace Member Profiles", 59:8 *Journal of the American Society for Information Science and Technology* 1321 (2007).

Townsend AM, "Network Cities and the Global Structure of the Internet", 44:10 *American Behavioral Scientist* 1697 (2001).

Towse R, *Creativity, Incentive and Reward: An Economic Analysis of Copyright and Culture in the Information Age*, Cheltenham UK; Northampton, MA: Edward Elgar (2001).

Traynor I, "Russia Accused of Unleashing Cyberwar to Disable Estonia", *The Guardian* (17 May, 2007), http://www.guardian.co.uk/world/2007/may/17/topstories3.russia.

Trujillo B, "Patterns in a Complex System: An Empirical Study of Valuation in Business Bankruptcy Cases", 53 *UCLA Law Review* 357 (2005).

Tu Y, "How Robust is the Internet?" 406 *Nature* 353 (2000).

Tufte ER, *The Cognitive Style of Powerpoint: Pitching out Corrupts Within*, 2nd ed, Cheshire, CT: Graphics Press (2006).

Turing, AM, "On Computable Numbers, with an Application to the Entscheidungsproblem", 42:2 *Proceedings of the London Mathematical Society* 230 (1937).

Ueda HR et al, "Universality and Flexibility in Gene Expression from Bacteria to Human", 101:11 *Proceedings of the National Academy of Sciences* 3765 (2004).

Uetz P et al, "Herpesviral Protein Networks and Their Interaction with the Human Proteome", 311:5758 *Science* 239 (2006).

Uzzi B and Spiro J, "Collaboration and Creativity: The Small World Problem", 111:2 *The American Journal of Sociology* 58 (2005).

Valverde S et al, "Self-Managing Systems – Self-Organization Patterns in Wasp and Open Source Communities", 21:2 *IEEE Intelligent Systems* 5 (2006).

Van Zwol R, "Flickr: Who is Looking?" *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence* (2007), http://www.semedia.org/PubFolder/vanZwol-FlickrWhoLooking.pdf.

Vázquez A, "Growing Network with Local Rules: Preferential Attachment, Clustering Hierarchy, and Degree Correlations", 67:5 *Physical Review E* 056104 (2003).

Vespignani A and Pastor-Santorras R, "Epidemic Spreading in Scale-Free Networks", 86:14 *Physical Review Letters* 3200 (2001).

Viegas FB et al, "Talk Before You Type: Coordination in Wikipedia", *Proceedings of the 40th Annual Hawaii International Conference on System Sciences* 78 (2007).

Vinnicombe T, "Thomas Hobbes and the Displacement of Political Philosophy", 32:8 *International Journal of Social Economics* 667 (2005).

Vliegendhart R, *Top 20 BitTorrent Trackers*, (2009), http://www.tribler.org/trac/wiki/DistributedTracker.

Waelde C et al, *The Common Information Environment and Creative Commons,* Final Report to the Common Information Environment Members of a study on the applicability of Creative Commons Licences (2005).

Waldrop MM, *Complexity: The Emerging Science at the Edge of Order and Chaos*, New York: Penguin (1993).

Walker CW, "Application of the DMCA Safe Harbor Provisions to Search Engines", 9:2 *Virginia Journal of Law & Technology* 1 (2004).

Wasserman S and Faust K, *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press (1994).

Watts DJ and Strogatz SH, "Collective Dynamics of 'small-World' Networks", 393 *Nature* 440 (1998).

Watts DJ and Strogatz SH, "Collective Dynamics of 'small-World' Networks", 393:6684 *Nature* 440 (1998).

Watts DJ, *Six Degrees: The Science of a Connected Age*, London: Vintage (2004).

Weber S, *The Success of Open Source*, Cambridge, MA: Harvard University Press (2004).

Weinberger D, *Everything Is Miscellaneous: The Power of the New Digital Disorder*, New York: Times Books (2007).

Weisstein EW, "Normal Distribution", *MathWorld* (2007), http://mathworld.wolfram.com/NormalDistribution.html.

Westkamp G, "Digital Rights Management, Internet Governance and the Autopoiesis of Modern Copyright Law", 7:4 *Contemporary Issues in Law* 317 (2005).

Wheeler DA, *Linux Kernel 2.6: It's Worth More!* (2004), http://www.dwheeler.com/essays/linux-kernel-cost.html.

Wheeler DA, *More Than a Gigabuck: Estimating GNU/Linux's Size*, (2002), http://www.dwheeler.com/sloc/redhat71-v1/redhat71sloc.html.

Wilkinson DA and Huberman BA, "Assessing the Value of Coooperation in Wikipedia", 12:4 *First Monday* (2007), http://goo.gl/SpPav.

Williams C, "How Egypt Shut Down the Internet", *The Telegraph* (28 January, 2011), http://goo.gl/j5PTU.

Williams RJ et al, "Two Degrees of Separation in Complex Food Webs", 99 *Proceedings of the National Academy of Sciences* 12913 (2002).

Willinger W, Alderson D and Doyle JC, "Mathematics and the Internet: A Source of Enormous Confusion and Great Potential", 56:5 *Notices of the American Mathematical Society* 586 (2009).

Wilson C, "Searching for Saddam: The Social Network That Caught a Dictator", *Slate* (February 22, 2010), http://www.slate.com/id/2245228/.

Wilson C, *Botnets, Cybercrime, and Cyberterrorism: Vulnerabilities and Policy Issues for Congress*, CRS Report for Congress (2008), http://bit.ly/aIzr1k.

Woodmansee M and Jaszi P, *The Construction of Authorship: Textual Appropriation in Law and Literature*, Durham; London: Duke University Press (1994).

World Association of Newspapers, *World Press Trends 2010*, (2010), http://www.wan-press.org/worldpresstrends2010/home.php.

Wu F et al, "Information Flow in Social Groups", 337:1-2 *Physica A: Statistical and Theoretical Physics* 327 (2004).

Wu T, "The Copyright Paradox – Understanding Grokster", *Supreme Court Review* 229 (2005).

Yook S-H, Jeong H and Barabási A-L, "Modelling the Internet's Large-Scale Topology", 99:21 *Proceedings of the National Academy of Sciences* 5 (2002).

Zentner A, *Measuring the Effect of Online Music Piracy on Music Sales*, University of Chicago Working Paper (2004), http://economics.uchicago.edu/download/musicindustryoct12.pdf

Zhang G-Q et al, "Evolution of the Internet and its Cores", 10 *New Journal of Physics* 123027 (2008).

Zhou S, Zhang G-Q and Zhang G-Q, "Chinese Internet AS-level Topology", 1:2 *IET Communications* 209 (2007).

Zipf GK, *Human Behavior and the Principle of Least Effort*, Cambridge MA: Addison-Wesley, (1949).

Zittrain J et al, *Access Denied: The Practice and Policy of Global Internet Filtering*, Cambridge, MA: MIT Press (2007).

Zittrain J, *The Future of the Internet: And How to Stop It*, London: Allen Lane (2008).

Zook MA, "Old Hierarchies or New Networks of Centrality? The Global Geography of the Internet Content Market", 44:10 *American Behavioral Scientist* 1679 (2001).

Zou CC, Towsley D and Gong W, "Email Virus Propagation Modeling and Analysis", *CiteSeer* (2003), http://bit.ly/difAHB.

Zuckerberg M, "500 Million Stories", *The Facebook Blog* (21 July, 2010), http://blog.facebook.com/blog.php?post=409753352130.

REFERENCES FOR THE CRITICAL REVIEW

Albert R, Jeong H and Barabási A-L, "Error and Attack Tolerance in Complex Networks", 406 Nature 378-382 (2000)

Angwin J and Valentino J, "Google's iPhone Tracking", Wall Street Journal (February 17, 2012), http://on.wsj.com/ytg5EB.

Ball P, Critical Mass: How One Thing Leads to Another, London: Arrow Books (2004).

Barabási A-L, Bursts: The Hidden Pattern Behind Everything We Do, New York, N.Y.: Dutton (2010).

Barabási A-L, Linked: The New Science of Networks, Cambridge MA: Perseus Pub. (2002).

Bloem M et al. "Malware Filtering for Network Security Using Weighted Optimality Measures", IEEE International Conference on Control Applications (2007).

Boyle J, “A Natural Experiment”, Financial Times (November 22 2004), http://is.gd/Ye0PhW.

Boyle J, “An Intellectual Property System for the Internet Age”, Financial Times (May 18, 2011).

Brown JS and Duguid P, The Social Life of Information, Boston: Harvard Business School Press (2002)

Capocci A et al, “Preferential Attachment in the Growth of Social Networks: The Case of Wikipedia”, 74 Physical Review E 036116 (2006).

Centre for Intellectual Property and Information Law, Review of the Economic Evidence Relating to an Extension of the Term of Copyright in Sound Recordings, University of Cambridge (2005), http://bit.ly/zODygW.

Chassin D P and Posse C, "Evaluating North American Electric Grid Reliability Using the Barabási–Albert Network Model" 355:2 Physica A: Statistical Mechanics and its Applications 667 (2005).

Chassin D P and Posse C, "Evaluating North American Electric Grid Reliability Using the Barabási–Albert Network Model" 355:2 Physica A: Statistical Mechanics and its Applications 667 (2005).

Dale C et al, "Evolution and Enhancement of BitTorrent Network Topologies", 16th International Workshop on Quality of Service (2008).

Davis MD, Game Theory : A Nontechnical Introduction, Rev. ed, Mineola, N.Y. ; London: Dover Publications, Constable (1997).

Dembosky A, “Facebook Timeline Ads Plan Raises Fresh Privacy Fears”, Financial Times (February 9, 2012), http://on.ft.com/xQHIRj.

EFF, Stop the Internet Blacklist Bills, (2011), http://blacklist.eff.org/.

European Commission, A single market for 21st century Europe: Accompanying Communication, COM(2007) 724 final (2007).

European Commission, First Evaluation of Directive 96/9/EC on the Legal Protection of Databases, DG Internal Market Working Paper, http://is.gd/DsY3XV.

European Commission, Intellectual Property: Commission adopts forward-looking package, IP/08/1156 (July 2008).

Fabrikant A, Koutsoupias E and Papadimitriou C H, "Heuristically Optimized Trade-Offs: A New Paradigm for Power Laws in the Internet Paradigm for Power Laws in the Internet", 2380 Lecture Notes in Computer Science 781 (2002).

Freeman LC, "Centrality in Social Networks: Conceptual Clarification", 1:3 Social Networks 215 (1979).

Gladwell M, The Tipping Point: How Little Things Can Make a Big Difference, London: Abacus (2002).

Gorman S, "Alert on Hacker Power Play", Wall Street Journal (February 2012), http://on.wsj.com/zR9MLi.

Guadamuz A, Networks, Complexity and Internet Regulation: Scale-Free Law, Cheltenham, UK: Edward Elgar (2011), p.15.

Hargreaves I, Digital Opportunity: A review of Intellectual Property and Growth, (2011), http://www.ipo.gov.uk/ipreview-finalreport.pdf.

Hart C, Doing a Literature Review: Releasing the Social Science Research Imagination", London: SAGE (1998).

HM Treasury, Gowers Review of Intellectual Property, (2005), E.10. http://bit.ly/gwg0tu.

Huberman BA, The Laws of the Web : Patterns in the Ecology of Information, Cambridge, Mass.: MIT Press (2001).

Hugenholtz B et al, The Recasting of Copyright & Related Rights for the Knowledge Economy,

ICANN, Factsheet: Root server attack on 6 February 2007, http://bit.ly/hcpBun.

Kauffman SA and Weinberger EW, "The NK model of rugged fitness landscapes and its application to maturation of the immune response", 141:2 Journal of Theoretical Biology 211 (1989).

Karlin J, Forrest S and Rexford J, Nation-State Routing: Censorship, Wiretapping, and BGP, arXiv Working Paper (2009), http://arxiv.org/abs/0903.3218v1.

Kinney R et al, "Modeling Cascading Failures in the North American Power Grid", 46:1 The European Physical Journal B 101 (2005).

Kumar R, Research Methodology: A Step-By-Step Guide for Beginners, London, SAGE (2010).

Leigh D, "How 250,000 US embassy cables were leaked", The Guardian (28 November 2010), http://goo.gl/Azhcm.

Lessig L, Code Version 2.0, 2nd ed, New York: Basic Books (2006).

Luhmann N, Social Systems, Stanford, Calif.: Stanford University Press (1995).

Lussier J T, Raeder T and Chawla N V, "User Generated Content Consumption and Social Networking in Knowledge-Sharing OSNs", 6007 Advances in Social Computing 228 (2007).

Lussier J T, Raeder T and Chawla N V, Digging up the Dirt on User Generated Content Consumption, Interdisciplinary Center for Network Science & Applications (ICeNSA), Working Paper, http://cse.vnit.ac.in/comad2010/ResearchTrack/paper%2057.pdf.

Milgram S, "The Small World Problem", 2 Psychology Today 60–67 (1967).

Newman MEJ, "Power Laws, Pareto Distributions and Zipf's Law", 46:5 Contemporary Physics 323 (2005), p.323.

Ravasz E and Barabási A-L, "Hierarchical Organization in Complex Networks", 67.

Revesz R L, "A Defense of Empirical Legal Scholarship", 69:1 The University of Chicago Law Review 169 (2002).

Rosas M, Valverde S and Solé R V, "Topological Vulnerability of the European Power Grid Under Errors and Attacks" 17:1 International Journal of Bifurcation and Chaos 2465 (2007).

Teubner G, Autopoietic Law: A New Approach To Law And Society, Berlin: Walter de Gruyter (1988).

Urban J M, "Efficient Process or Chilling Effects-Takedown Notices under Section 512 of the Digital Millennium Copyright Act" 22 Santa Clara Computer & High Technology Law Journal 621 (2005-2006).

Walker L, “Social facts: Scientific methodology as legal precedent”, 76:4 California Law Review 877 (1988).

Watts DJ, Six Degrees: The Science of a Connected Age, London: Vintage (2004).

Wosinska L et al, "Network Resilience in Future Optical Networks", in Hutchison D et al (eds), Lecture Notes in Computer Science, Berlin: Springer (2009).

Weinberg G, An Introduction to General Systems Thinking, New York, NY: Dorset House, 1975 (2001).

Index