2016-05-01

Computing Research in Academia: Classifications, Keywords, Perceptions, and Connections

Sung Han Kim
Brigham Young University - Provo

Computing Research in Academia: Classifications,

Keywords, Perceptions, and Connections

Sung Han Kim

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Richard G. Helps, Chair
Derek L. Hansen
Barry M. Lunt

School of Technology

Brigham Young University

May 2016

ABSTRACT

Computing Research in Academia: Classifications,
Keywords, Perceptions, and Connections

Sung Han Kim
School of Technology, BYU
Master of Science

The Association for Computing Machinery (ACM) recognizes five computing disciplines: Computer Science (CS), Computer Engineering (CE), Information Technology (IT), Information Systems (IS), and Software Engineering (SE). Founded in 1947 the ACM is the world's largest society for *computing educators, researchers, and professionals*. While Computer Science has been a degree program since 1962, the other four are relatively new. This research focuses on understanding the graduate research in four of the five ACM disciplines (CS, CE, IT, and IS) using a large body of thesis and dissertation metadata. SE is not found in the metadata and graduate work in SE is not included. IS is no longer officially found in the metadata so its representative ProQuest replacement, Information Science—although not an ACM recognized discipline—is used based on the commonality of the associated ProQuest Classification code.

The research is performed using co-word and graph analysis of author-supplied Classifications, Departments, and keywords.

Similarities and differences between the disciplines are identified. Whether the computing discipline is the primary or the secondary focus of the research makes a large difference in the connections it makes with other academic disciplines. It was found that the Departments from which computing research originates varies widely but the majority come from computing-related Departments. Finally, gaps are apparent from the practitioners' views of the computing disciplines versus the public's view.

Keywords: research, computer science research, information systems research, information science research, computer engineering research, education, thesis, dissertation, computing research, information technology research

ACKNOWLEDGEMENTS

I would like to thank my father and mother, Joseph KiChin Kim and Kelly Kyong Kim whose lives are miracles of hard work, longevity, and determination. Without their vision and incredible labors none of this would have been possible—none of it. They are where it all began.

I would like to thank Professor Hansen and Professor Helps for all their help and encouragement, not just during the course of this thesis but in all of our previous interactions. They are fine examples of men and academics. My gratitude also goes out to Professor Lunt for his courtesy and kindness.

I would like give special thanks to my beautiful wife, Christina, and my adorable children, Alexandria, Catherine, and Kaj, for their support, humor, and encouragement. They were all champs, especially my wife during the finishing moments of this endeavor. They are the end that has no end, to combine with the beginning that had no beginning.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

While *IT* is ubiquitous as a term meaning *computing*, IT, the academic discipline, is still new (SIGITE, *IT Discipline)*. It now stands as a separate academic discipline apart from its related computing disciplines (CD) of Computer Science (CS), Information Systems (ISys), Computer Engineering (CE), and Software Engineering (SE). These are the disciplines recognized by the Association for Computing Machinery (ACM) (*Computing Disciplines & Majors*).

Despite the inherent overlap of these fields, there must be differences that justify the existence of these separate, yet related, disciplines. To further develop the understanding of what each discipline entails, as well as their differences, a large corpus of thesis classifications and keywords is examined in this study. Connections between the computing disciplines, as well as to non-computing disciplines are considered, shedding light on the nature of interdisciplinary research and the place of computing majors in the broader University.

1.1 Nature of the Problem

Computing was once a research area reserved for a select few whose enthusiasm, patience, and intense focus allowed them to spend years tapping into the meager computing power and capabilities that were available 10, 20, and even 40-plus years ago. Today, computing is not only pervasive in society, but most modern research requires computing of some form. Be it a no-longer-simple word processing program or the massive computing used to collect, analyze, and represent billions or trillions of data points, computing resources are used in most, if not all, disciplines.

As computing continues to expand and pervades not just research and academia, but almost every facet of human life, there must be more than a single blanket term for computing and more than a single discipline that covers computing and its uses.

Louis Fein, an early advocate (Fein, 1984) for the creation of a computing discipline to be taught at universities, wrote the following in a seminal 1959 paper—before there were any recognized computing disciplines in any U.S. college or university:

> "An integrated university program is recommended reflecting the conviction that many present activities related to computers will develop into disciplines and as such are the legitimate province of the university scholar."

Fein recognized a need for a computing discipline during modern computing's nascency. Fein's vision of computing disciplines (Fein, 1959) has largely come to pass with the acceptance of Computer Science as an academic discipline, starting in the U.S. as a graduate program at Purdue University in 1962, as well as with the current emergence of other fields addressing hardware, software, theory, and more. Fein continues:

> "We must expect that some of these fields will coalesce and develop into disciplines on their own. These will then almost certainly be universally accepted as the legitimate province of the university scholar. Others may not turn out to be disciplines and will gradually be abandoned by universities."

Medicine and engineering have multitudes of specialities that are disciplines unto themselves. Computing, likewise, needs separate disciplines that differentiate themselves into specific bodies of computing study and research. Again, the ACM currently defines these disciplines as above.

Computing today, however, has a unique quality that makes differentiation more difficult: and that is its approachability. Anyone with an Internet connection and access to a computer can learn about computing at basic and even high levels. Free online courses abound. There are few regulatory barriers to practice. There is no required schooling. At basic levels there is no advanced math; and computing can be abstracted to levels where math is not required, let alone appreciated (Hoyles, 2014). Examples of online courses now available run the gamut from algorithms (Wayne, 2016) and data science (Howe, 2014) to circuits and electronics (Agarwal, 2007).

Because anyone in the world with a computer and an Internet connection can broach the world of computing, any academic discipline can also easily incorporate computing into their curriculum—much more so than they could with medicine, engineering, or any other technical field with barriers to entry. In so doing, graduate research using computing elements, versus research in computing, can muddy the waters as to what constitutes research in a computing discipline.

For example Georgia Tech has an MS degree in *Digital Media*, which was formerly known as *Information Design and Technology* (*MS in Digital Media*). This program focuses on teaching computing to arts and humanities students. Another master's degree offered by the University of Massachusetts Amherst is called *Learning, Media, and Technology* (*Master's Degree: Learning, Media and Technology Concentration*). This program focuses on technology and education.

Can any research be called computing research simply by having a computing element? Does education, or the humanities, if it makes up a large contingent of computing research constitute a separate branch of computing like IS, with its focus on management and organizations? What is the relationship between computing disciplines, both core and peripheral?

1.2 Research Objectives

Use co-word and graph analysis of graduate computing research, along with their associated classifications and keywords, to understand cross-disciplinary connections between the computing disciplines and other university research fields.

This research seeks to show the presence of various graduate academic areas connected to graduate computing research and how they relate to each other and to the computing disciplines. The connections the computing disciplines make with each other will also be researched.

1.2.1 Focus of Research

This thesis will focus on taking both a broad and in-depth survey of the research being done in four ACM computing disciplines: IT, CS, CE, and ISys. Using a large body of thesis classifications and keywords, connections between theses and commonality of keywords are investigated.

The expected insights include a greater understanding of the computing disciplines, how they relate to each other, as well as how they relate to non-computing disciplines. Additionally, this research will examine how these relationships have changed over time.

Limitations for this study include:

- Not using the complete ProQuest dataset, but instead only using a subset based on the relevant years of 2009 through 2014—with the reasoning behind these years to be explained later.
- ProQuest Classifications actually change over time based on ProQuest's desire to

keep Classifications modern. Therefore the same Classifications will not always be present from all-time to the present.

2 LITERATURE REVIEW/BACKGROUND

With the explosion of computing and the emergence of various computing disciplines, it is important to understand the current research being done in the various computing disciplines—as well as the research being done in the disciplines that have, or claim, computing elements. This can help define the scope of computing research in graduate programs and its evolution up to this point.

From foundational computing and electronics communications work by Claude Shannon (NYU, *Claude Shannon*) to pioneering algorithmic research by Donald Knuth (Knuth, *The Art of Computer Programming*); from the simple bubble sort to the much more efficient quicksort; from the abacus invented by the Chinese in the 1300's to Pascal's calculator in the 1600's (Falk, 2014); from the Harvard Mark I in the 1940's to today's Tianhe-2 supercomputer; and from the alleged "640K ought to be enough for anybody" in 1981 (Lai, 2008) to 1.404 Petabytes (2.2 billion times more than 640K) of RAM in the Tianhe-2 supercomputer (Dongarra, 2013) the hardware, processing, theory, and algorithms of computing—and computing itself—has developed over decades and even centuries.

But, in the past 30-plus years (Computer History Museum, *1960 | Timeline*), computing has seen an explosion thanks in large part to the development and easy availability of digital computers, which was made possible by advances in semiconductor technology. Going further back into the recent past, before digital was common, the 1930's and 40's brought innovations such as Telex, Colossus, the Harvard Mark 1, and Eniac, all, respectively, marvels of wiring, vacuum tubes, electro-mechanics, and finally solid-state electronics (Ibid. *1933 | Timeline*; Ibid. *1944 | Timeline*). Hardware advancements and the advent of transistors and the integrated circuit have shrunken what used to fill a large room, like any of the aforementioned, into something that will fit in a hand—yet is over a thousand times faster (AntiqueTech, 2013).

The growth of computing technology has created many disciplines, sub-disciplines, and related fields of computing. Digital Signal Processing, Electronics, Quantum Computing, and Computational Physics are all examples of fields that sprout from, or use, computing, or are a fusion of disciplines.

Of the many fields that can be considered computing or have computing elements, at this time the computing specializations recognized by the Association for Computing Machinery (ACM) are Computer Science, Information Systems, Computer Engineering, Information Technology, and Software Engineering (*Computing Disciplines & Majors*).

The Association for Computing Machinery is the world's largest computing society, boasting over 100,000 members from countries all over the world (*About the ACM Organization*). The ACM fosters relationships between educators, professionals, and researchers to help drive the computing profession and increase computing's impact worldwide (*Association for Computing Machinery*). The ACM includes 37 Special Interest Groups that represent *major areas of the dynamic computing field* (*About SIGS*).

Common ancestry for the computing disciplines entails similarities. Just as engineering requires a common core of math and physics and medicine requires biology, physiology, and chemistry, so too the computing disciplines can be expected to share concepts, key topics, and skills. Such concepts and skills include programming, operating systems, and networking (*Curricula Recommendations*).

Much research has already been done to develop and define these particular, ACM recognized, computing disciplines. And, as new discoveries are made and new methods and uses of computing are discovered, or created, more disciplines will be formed. This trend is seen recently with the emergence of Data Science as a degree (White, 2016). Whether it becomes an ACM recognized specialization on par with the current five is yet to be seen but it is currently a very active area of research (Dhar, 2013), and a related ACM Special Interest Group already exists: SIGKDD, Special Interest Group in Knowledge Discovery and Data Mining (SIGKDD, *About SIGKDD*).

2.1 The Computing Disciplines

It may be argued that the computing disciplines need little publicizing or marketing. This was not the case in 1959 when Louis Fein, even after 12 years of ACM existence, pushed for the creation of computing disciplines in academia (Fein, 1959). Computer Science was not available as a college degree in the United States until 1962, Purdue University being the first to offer it as a graduate degree (Rice and Rosen, 2013). Purdue added a bachelor's degree in CS in 1967.

Regardless of the current popularity of computing disciplines, colleges and educational institutions must refine their course and degree offerings to match the desires of prospective students, which desires are often driven by the job market—as demonstrated by the creation of a Data Science degree at institutions such as the University of San Francisco and the University of Iowa (White, 2016). This refinement of computing disciplines that is happening today was described by Fein 57 years ago (Fein, 1959).

Therefore, with limited resources, it is important for education providers to determine what their offerings will be. It is also important for students, and their families, to know what they will be learning and spending many hours, as well as many thousands of dollars, on. Employers must also know what to expect of graduates in a particular field.

An important question, then, is, “What is to be offered?” Should institutions just pick from one of the five ACM disciplines? Or, will it be a hybrid offering, maybe with a new title? What department should the program be offered from? Engineering? Mathematics? Computing? All these questions were brought up by Fein (1959) and are as relevant today as in Fein’s day.

To understand the types of computing offerings currently available from accredited institutions of higher learning, below are some examples of computing programs recognized by the Accreditation Board for Engineering and Technology (ABET). In Table 2-1 are ABET programs containing the term “compute” and in Table 2-2 are ABET programs containing the term “technology” (*Accredited Program Search*).

Table 2-1: ABET Accredited Programs Containing the Term "Compute"

Computer Science	Computer Engineering
Computer Engineering Technology	Electrical and Computer Technology
Computer Science and Engineering	Electrical and Computer Engineering Technology
Electronic and Computer Engineering Technology	Electronics and Computer Engineering Technology
Computer Systems Engineering	Computer Information Systems
Computer Engineering Option in Electrical Engineering	Computer Technology
Electrical and Computer Option in Engineering Technology	Electronics & Computer Engineering Technology
Computer Systems Engineering Technology	Computer Engineering and Computer Science
Engineering Technology Option in Electronics & Computer Engineering Technology	Electronic(s) and Computer Engineering Technology
Computer Integrated Manufacturing Engineering Technology	Computers Systems Option in Electrical/Electronic(s) Engineering Technology

Table 2-2: ABET Accredited Programs Containing the Term "Technology"

Mechanical Engineering Technology	Electronic(s) Engineering Technology
Electrical Engineering Technology	Civil Engineering Technology
Computer Engineering Technology	Manufacturing Engineering Technology
Architectural Engineering Technology	Construction Engineering Technology
Industrial Engineering Technology	Electrical and Computer Engineering Technology
Electromechanical Engineering Technology	Civil and Construction Engineering Technology
Information Technology	Electronic(s) Technology
Electronic Engineering Technology	Engineering Technology
Architectural Technology	Electronic and Computer Engineering Technology
Electrical Option in Engineering Technology	Mechanical Technology

The ACM describes its five computing disciplines as follows:

- Computer Engineering

  "Typically involves software and hardware and the development of systems that involve software, hardware, and communications" (*Computing Disciplines & Majors*).

- Computer Science

  "Currently the most popular of the computing disciplines; tends to be relatively broad and with an emphasis on the underlying science aspects"

- Information Systems

  "Essentially, this is computing in an organizational context, typically in businesses."

- Information Technology

  "Focuses on computing infrastructure and needs of individual users; tends to involve a study of systems (perhaps just software systems, but perhaps also systems in support of learning, of information dissemination, etc.)."

- Software Engineering

> "Focuses on large-scale software systems; employs certain ideas from the world of engineering in building reliable software systems."

Figure 2-1 is a hierarchy of the computing disciplines based on more detailed ACM descriptions (*Computing Degrees & Careers; Computer Science; Computer Engineering; Information Technology; Information Science; Software Engineering*), with each layer using and requiring the work of the underlying layer.

For example: CE produces new hardware and CS creates and optimizes the software that will run on the new hardware. IT takes the combination of hardware and software and combines them with other hardware and/or software elements to create and administer larger systems, including the setup of their communications links. IS takes these systems and applies them to businesses and organizations, adding features if necessary. SE should be able to develop software at any level, from low-level firmware to high-level business applications and thus covers all four layers. Along with SE, is the possibility of other new fields emerging at, below, or above any of the current fields. More detailed descriptions of the ACM recognized computing disciplines can be found at (*Computing Disciplines & Majors*).

2.2 ProQuest

In order to research the research that is currently being done in the computing disciplines, as well as in the non-computing disciplines, a large repository of academic research needs to be found. As of the time of this writing, there is no larger, or more complete, repository of graduate academic research than ProQuest. This work will examine records from the ProQuest/UMI

Dissertations & Theses Global database to better understand how graduate students classify their computing-related research.

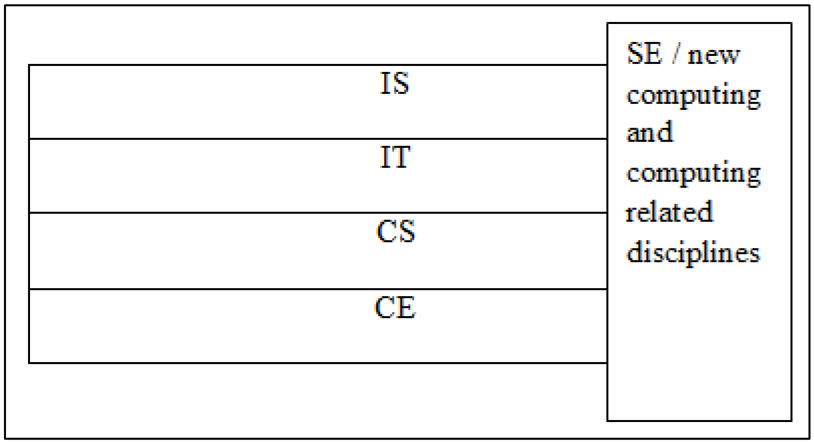


Figure 2-1: The ACM Computing Disciplines and Their Hierarchy

ProQuest/UMI's (University Microfilm Inc.) Dissertations & Theses Global (PQDT) database is *the* most comprehensive collection of theses and dissertations (from here on simply theses, even though there are more dissertations than theses) from around the world. Its database contains over 3.8 million theses from 1743 to the present, with full text available for works after 1997. English language theses equal over 3.4 million with large amounts of international works currently being added (*ProQuest Dissertations & Theses Global*).

However, ProQuest does not represent all available dissertations and theses—not even for the colleges and universities in the United States, although it does represent most dissertations (*Dissertations Abstracts International*). Colleges, universities, and other organizations voluntarily submit their theses to ProQuest and not all choose to do so, while some submit only a portion of their theses (Kelsky, 2011). Brigham Young Unversity is an example of one institution that does not submit its master's theses to ProQuest.

The exact make-up of ProQuest graduate research in terms of numbers of master's theses versus doctoral dissertations is unknown. For the current dataset there are roughly two times more dissertations than theses.

ProQuest specializes in information and data collection and dissemination, and has grown from a firm founded in 1938 that specialized in microfilming books (*History & Milestones*), then known as University Microfilm, to a large conglomeration of at least 15 different companies that operate under a single ProQuest brand (Kaser, 2014).

In 1951 the Association of Research Libraries (ARL), a nonprofit organization that brings together the top research libraries in the US and Canada (*About | Association of Research Libraries*), whose members include Princeton, Harvard, Johns Hopkins, and many other institutions, including Brigham Young University (*List of ARL Members*), allowed ProQuest to provide dissertation services, which ProQuest started as Dissertation Abstracts. In 1998, and officially announced in 1999, ProQuest became the official off-site repository for the U.S. Library of Congress for digital theses and dissertations (*History & Milestones*; Library of Congress, 1999). ProQuest Digital Dissertations, at that time held, over 100,000 theses converted to digital form, beginning with theses from 1997. The agreement with the Library of Congress allowed ProQuest Digital Dissertations to register and deposit digital copies of dissertations and theses in behalf of the U.S. Copyright Office (Library of Congress, 1999).

ProQuest has partnerships with over 700 universities and processes more than 90,000 graduate works every year (*Dissertations and Theses Dissemination and Ordering*). In 1999 it claimed relationships with 99 percent of accredited institutions of higher learning in North America, publishing doctoral dissertations for those institutions (Library of Congress, 1999).

For doctoral dissertations, ProQuest, with its Dissertations Abstracts International (DAI), estimates 95% to 98% of all doctoral dissertations are included in DAI (*Dissertations Abstracts International*).

2.3 Network Analysis and NodeXL

This study will perform keyword analysis (i.e., co-word analysis), which is a subset of the broader field of network analysis and graph theory, using a tool called NodeXL.

Graphs and visuals enable both faster consumption of data, i.e. less mental impedence, and the possibility of seeing insights not readily seen in tables or other representations of data. Network graphs in particular, by showing connections between data points, are well suited to exposing relationships that may not be apparent in lists, tables, or even other forms of visualizations. Network graphs are also well-suited to display co-word information, or the connections between words that share a common trait, which is a thesis in the current study—and the words are the Classifications.

In a graph, vertices (also nodes) are connected by edges (or lines). The edges between vertices can be undirected or directed. If undirected the connections A-B and B-A are considered equal. If a graph is directed then A-B and B-A are distinct and have two separate meanings. In this study, the vertices represent ProQuest Classifications, while an edge connecting two vertices represents a thesis that includes both Classifications together. If a thesis includes many Classifications, all of them would be connected together via edges.

Weighted graphs allow edges to carry a value for a connection. The network analysis conducted in this work is a weighted graph where the weight of an edge is determined by the number of theses that use the same two Classifications that are connected by the edge.

2.3.1 **NodeXL**

NodeXL is an easy-to-use extension to Microsoft Excel that allows for quick creation of network graphs. Calculating rich sets of graph metrics and clustering are all easily done with NodeXL (Hansen, 2011).

2.4 Network Clustering Algorithms

As it is desired to see how the Classifications in the current study relate to each other—beyond the simple connections made from one Classification to another—clustering provides a way to discover semantics not easily seen from nodes and edges alone.

Network clustering algorithms present a way to tie closely connected neighbors to each other, to build a community of data points. In other words, clusters or communities are those nodes with many edges between them, connected to other groups or clusters with fewer edges (Clauset, 2004). A variety of specific network clustering algorithms exist, each with different properties. M.E.J. Newman, at the University of Michigan, has done extensive work on finding communities within network graphs using both CS and social science methods (2006).

One popular algorithm is the Clauset, Newman, and Moore algorithm, originally developed to mitigate the difficulty of the working with large graphs in an efficient manner (Clauset, 2004). The Wakita-Tsurumi algorithm was later developed to mitigate some of the deficiencies in the Clauset, Newman, and Moore algorithm—which did not scale as expected—with the ability to scale to sets of vertices greater than 500,000 (Wakita, 2007). Clustering can bring to light connections, or relationships, not discernable via plain graphs or tabular data (LaSalle and Karypis, 2014).

2.5 Keyword Co-Occurrence Analysis

In the current study, co-words are co-Classifications. They are used in conjuction with the network graphs to analyze the relationships of the Classifications associated with the computing disciplines.

Co-word analysis is a well used technique in academia to understand relationships between concepts (Kim, 2014). As cited in Kim it has been used in software engineering research (Coulter, 1998), in understanding regional innovation systems (Lee and Su, 2010), and in understanding the status and trends of research in library and information science in China (Hu, 2013). Hu is similar to the current research as Hu used co-word analysis to find groups and trends in library and information science research in China, and this current study also seeks to find groups and trends, but in computing research.

Research by Lee and Su (2010) is also particularly relevant as they "analyz[ed the] co-occurrence of keywords specified by authors" using co-word analysis to "present an overview of RIS [Regional Innovation System] research" and to "find the research contexts of RIS." The current study also uses co-word analysis to get an *overview of the research* in the computing disciplines as well *finding contexts, or groupings* of disciplines, in computing research.

Co-word analysis is now an accepted and common research technique. "The method of co-word analysis is a well-known relational bibliometric method" (Lee and Su, 2010). And, "… it has been accepted as a reasonable way to map the relationships among concepts, ideas, and problems" (Hu, 2013). As described above, it performs network analysis on the metadata associated with documents, such as the keywords and Classifications associated with a thesis.

3 METHODOLOGY

ProQuest represents a rich set of metadata from a large body of theses over many years from which to gain insights into graduate research. This study will focus on three data fields: Classifications, Identifier / keywords, and Departments. This study will examine the network connections made by the Classification field, as well as perform various analyses on the Classification, Identifier / keyword, and Department fields. Each field, and the research to be done, will be explained in the following sections. Future research could easily incorporate additional fields, such as Abstracts and Universities.

3.1 Understanding the ProQuest *Classification* Field

The ProQuest instructions for submitting a doctoral dissertation or master's thesis has the following description for *Subject Categories*:

> "The first (primary) subject category that you enter is the one under which your dissertation or thesis will occur in our citation and abstract indices. Using Guide 2: ProQuest Subject Categories (attached), choose the category that most closely corresponds with the field in which you did your graduate research. You may add one or two more secondary subject categories; these will be associated with your work *and may increase its exposure to search engines.* (*ProQuest 2015-2016 Publishing Agreement*)"

A separate ProQuest 2015-2016 guide for *Subject Categories* has similar instructions:

> "The ProQuest Dissertations & Theses (PQDT) database and the ProQuest citation indices are arranged by subject categories. Please select the one category below that best describes

> your field of research or creative work. You may add one or two additional categories in your submission form that will also be associated with your work as secondary subjects" (*ProQuest Subject Categories – 2015-2016 Academic Year*).

The ProQuest *Subject Categories* and their codes found in Guide 2 of the ProQuest submission form and found in the Subject Category Guide instructions, correspond to the Classification field in the ProQuest metadata. It should be noted that there is also a 'Subject' field in the metadata that has no corresponding code but closely mirrors the Classification field.

Classifications are not majors or degrees, but are more akin to Classification of Instructional Program (CIP) codes created by the National Center for Education Statistics (NCES).

> "The CIP titles and program descriptions are intended to be generic categories into which program completions data can be placed, not exact duplicates of a specific major or field of study titles used by individual institutions. CIP codes are standard statistical coding tools that reflect current practice, and are not a prescriptive list of officially recognized or permitted programs" (NCES, *Introduction to the Classification of Instructional Programs: 2010 Edition (CIP-2010)*).

When the Classification name matches the generally accepted name of a degree or discipline, and in this case an ACM recognized computing discipline, it is taken in this study to represent the computing discipline that closely matches the corresponding degree.

The ProQuest instructions above lead to the following categorizations for the theses in this study—based on the position of the Classification on the Classification line. The number of Classifications included in a thesis indicates the extent to which it crosses disciplinary boundaries.

- **Any** (The position of the computing discipline within the Classification line is not important so long as it is present—this is simply the combination of Primary and Secondary below.)
- **Primary** (The computing discipline is the first Classification, indicating what subject category the author feels most closely aligns with their work. Other Classifications may be included as Secondaries (explained below).)
- **Secondary** (The computing discipline is not in the first position and therefore not the primary focus of the thesis; it can be second, third, forth, etc. just not first.)
- **Solo** (The computing discipline is the one-and-only listed Classification and indicates that the author felt the thesis fell within a single discipline with no cross-disciplinary aspects.)

The full list of 411 ProQuest Classifications for the 2015-2016 academic year can be found in the 2015-2016 ProQuest submission form or Subject Category guide. Examples include:

- ACCOUNTING 0272
- ACOUSTICS 0986
- ADULT EDUCATION 0516
- AERONOMY 0367
- AEROSPACE ENGINEERING 0538
- AESTHETICS 0650
- WOOD SCIENCES 0746
- WORLD HISTORY 0506
- ZOOLOGY 0472

The following are the computing disciplines that are the target of this research as found in ProQuest:

- COMPUTER SCIENCE 0984
- COMPUTER ENGINEERING 0464
- INFORMATION TECHNOLOGY 0489
- INFORMATION SCIENCE 0723

ProQuest Classification names will be in all caps in this report to easily discern when a specific ProQuest Classification is being discussed.

The ACM recognized computing discipline of Software Engineering is not present as a ProQuest Classification and is not included in this study. Also, the data shows INFORMATION SYSTEMS is no longer included in the ProQuest list of Classifications as of 2009.

From email correspondence with Carol Wadke, a ProQuest Author/School Relations Specialist, on May 5, 2016 it was discovered that Classification names are reviewed roughly every five years to keep the terms current. In another email on May 2, 2016 Ms. Wadke revealed that Classifications, and keywords, associated with a thesis can be modified by the author at any time, even after submission.

No duplicates of Classification codes were found in the 2015-2016 submission form or Subject Category guide.

3.1.1 **Information Systems Caveat**

INFORMATION SYSTEMS (ISys) was removed as a ProQuest Classification in 2009, the same year that INFORMATION TECHNOLOGY was introduced. In its place INFORMATION

SCIENCE (ISci) is used, which shares the same ProQuest Classification code as INFORMATION SYSTEMS: 0723.

The removal of INFORMATION SYSTEMS is unfortunate as it is a well recognized computing discipline. The removal of INFORMATION SYSTEMS potentially obscures the research performed in the separate disciplines of computing since a well-known discipline must now find a different "home" for its theses when submitting theses to ProQuest.

3.2 Understanding the ProQuest *Indentifier / keyword* Field

The ProQuest 'Identifier / keyword' field offers a broader ability to classify theses in addition to the ProQuest Classification field.

Identifier / keyword is a free-form field in ProQuest that allows an author to add any terms the author thinks will allow for greater visibility to search engines.

From the ProQuest publishing agreement:

> "Adding good keywords is another way to increase the chances that your work will be discovered. For example, geographic locations or specialized terms that do not occur in your title or abstract can increase exposure of your work" (*ProQuest 2015-2016 Publishing Agreement*).

The field is labeled: "Identifier / keyword." Below are samples from the current dataset. Note the presence of *applied sciences* in all the samples. It was found that *applied sciences* is present in 95% of the 33,706 theses in the dataset from 2009 to 2014.

- Applied sciences, Pure sciences, Information, Logic, Metrized entropy, Probabilistic reasoning, Reasoning on metric space, Uncertainty
- Communication and the arts, Applied sciences, Bank website, Credibility, Design elements, Perception of trust, Trustworthy, Website design

- Applied sciences
- Applied sciences, Adaptive circuits, SER tolerance, Dynamic voltage scaling, Reliable computing, Low-power VLSI, Variability compensation
- Applied sciences, Psychology, Task performance, Classifier algorithms, Coaching strategies, Intelligent tutoring systems

Again, from correspondence with Ms. Wadke on May 2, 2016, keywords, like Classifications, can be changed at any time by authors.

3.3 ProQuest Data Criteria

The following bullet points outline, and summarize, the ProQuest data that will be used in the current study. The time-boxing allows for more consistent labeling of Classifications, as the chances of the Classifications of interest, especially the computing Classifications, changing during this six years span of interest will be smaller than over the complete dataset. The complete dataset for the computing disciplines total 109,815 distinct theses. Overlap exists as a thesis containing both IT and CS will be downloaded twice, once for each respective discipline.

- **Time-box:** 2009-2014, with any exceptions noted.
  - The downloaded dataset from ProQuest spans back to 1937 (the year of the first instance of any of the computing disciplines, in this case INFORMATION SYSTEMS), but, the focus of this research is on the six-year period from 2009 to 2014. 2009 was the year IT was introduced by ProQuest, and likely the year CE and ISci were also added.
- **Classifications:**
  - COMPUTER SCIENCE (CS)

- COMPUTER ENGINEERING (CE)
- INFORMATION TECHNOLOGY (IT)
- INFORMATION SCIENCE (ISci, except where indicated)

- **Keywords**
  - All
- **Departments**
  - All

This work builds upon an original work using ProQuest data to graph the Classifications of INFORMATION TECHNOLOGY theses with other disciplines (Kim, 2015).

Classification and Identifier / keyword are present in all records. Department is present in 64% of the theses in the 2009-2014 timeframe. And again, despite calling them theses, roughly two-thirds of the graduate research in the dataset are doctoral dissertations.

3.4 Getting the ProQuest Data

The exact process for downloading ProQuest data can be found in Appendix A. Each of the Classifications matching each of the four ACM disciplines was found using the following general query:

cc("INFORMATION TECHNOLOGY") and la.exact("English") and pd(2009)

Corresponding variations were used to match the different disciplines and years being downloaded. "cc" is the filter for Classification. The "la.exact," or exact language, filter is used to only download English language theses. And, "pd" is the publication date filter.

From email correspondence with Carol Wadke in October 2014 it was found that INFORMATION TECHNOLOGY became an official ProQuest Classification in 2009. From the same set of emails it was found that ProQuest does not know the exact year COMPUTER

SCIENCE became an official ProQuest Classification. ProQuest editors believe it was before 1984.

Quoting an email from ProQuest Specialist Carol Wadke from November 2014 regarding CS and ISci:

> “We don’t have a record of when Computer Science and Information Science were added – the oldest reference we see is 1984; the note indicates they were in use at that time, so they must have been added before that year. Computer Engineering was added around 2009.”

So, the year INFORMATION TECHNOLOGY was added is the only year of computing discipline addition, of the four, that ProQuest editors knew with certainty. Given the five year cycle of review, and from the data itself, it is likely that ISci and CE were also added in 2009, and IS was removed that same year. Also, although the emails with Wadke from 2014 states that INFORMATION SCIENCE was in use in 1984 the data shows that they likely meant INFORMATION SYSTEMS since the first year INFORMATION SCIENCE makes its appearance with any appreciable numbers is in 2009.

Other ProQuest Classifications that could be considered computing disciplines include:

- ELECTRICAL ENGINEERING 0544
- ARTIFICAL INTELLIGENCE 0800
- SYSTEM SCIENCE 0790
- TECHNICAL COMMUNICATION 0643
- INDUSTRIAL ENGINEERING 0546

The above are not included in this research since the focus of this research is on the perception of the terms CS, CE, and IT, which match ACM computing disciplines. ISci was included since it kept the same Classification code as the ACM discipline of ISys.

3.5 Analyses to Be Performed

In order to build a rich understanding of the ProQuest data the following analyses will be performed on the data:

- Counts of theses by computing discipline
- Most prominent co-Classifications by computing discipline
- % of Classifications that co-occur with the computing discipline
- Cluster graphs of each computing discipline
- Frequency of Departments
- Keyword frequency charts
- Most prominent keywords

3.6 Data Analyses

Data was analyzed using SQL language queries in MySQL heavily augmented with Python scripts. Excel, Google Sheets, and Bash were also extensively used.

NodeXL was used to create the network graphs and run the graph metrics and clustering algorithms.

3.6.1 Using the Graphs

Although this research uses graphs of vertices and edges this research does not delve into graph theory, or the many rich uses of graphs. The visual power of graphs is used to show connections between theses subjects and classifications and to perform network clustering. These two functions are not apparent, or easily constructed, using tabular data.

The edges in the graph are undirected edges that illustrate the connections among Classifications and the clusters they form based on their connections. Again, as tables do not easily show connections between data points, graphs and visualizations can bring insights not easily discernible from tabular data alone (LabWrite Resources, 2004).

Directionally edged graphs were considered and not used for this dataset as much of the information from directionality could be inferred from the co-Classification tables. Directional graphs would show many arrows primarily pointing to just a few Classifications. This information is already found in the top five co-Classification tables below. Primary Classifications would point to each Secondary Classification with the sizes of the edges based on the number of times the connection occurs. The directionality of Secondaries is not significant per the ProQuest submission instructions.

As described in Kim (2014) the graphs show edge-thickness, or width, based on the number of times two nodes appear together in a thesis. Almost all graphs will be shown with edge weights 10 and up to be consistent across the computing disciplines and to avoid visual clutter. This means that at least 10 theses shared the two Classifications that are connected. Showing this level of detail should be sufficient to show the general categories of the main Classifications.

Note that the number of connections between a computing discipline and its single most common connecting discipline, which could also be a computing discipline, when there are only two Classifications in a thesis, are not shown.

3.6.2 About the Graphs

The network visualizations are laid out exactly as in Kim (2014), with the Harel-Koren Fast Multiscale algorithm, along with some manual adjustment. And just as in Kim (2014) the X and Y coordinates, as well as the distance between nodes, have no meaning. Additionally:

- All graphs were created using NodeXL.
- Graphs edges are weighted based on the number of theses that shared the Classification.
    - They are all undirected graphs.
- All graphs are clustered using the Wakita-Tsurumi clustering algorithm.

To focus the graphs on the most important content, filtering of nodes and edges was applied. Edges containing the main computing discipline, as done in Kim (2014) with IT, are filtered out to prevent visual clutter. For example, a graph of Classifications that co-appear with CS will not include CS in the graph. Instead, CS will be indicated in the graph's title.

For an edge to appear between two Classifications they must come from a thesis with three or more Classifications. If a thesis has only two Classifications, the only edge is between the main computing discipline and the other Classification. Therefore, at least three Classifications must be present so that the two Classifications that are not the main computing discipline can create an edge. Only graph edges with a minimum edge weight of 10 are shown as in Kim (2014).

The connections between the non-main computing disciplines are the data points of interest in the graphs. Direct connections to the main computing disciplines can be seen in tabular form.

For the following graphs, unless otherwise noted, the vertex size, vertex color, and edge width, edge color, and edge opacity are set as described below:

- ***Vertex size*** represents the *betweenness centrality* of each classification. They have a max pixel size of 40. Logarithmic scaling is an option but not used. Betweenness centrality is calculated based on the entire graph before filtering out the edges with a weight of less than 10.

- ***Vertex color*** represents the clusters produced by the Wakita-Tsurumi algorithm.
- ***Line Width, Color, and Opacity*** represent the edge weight, or the number of theses that share the two Classifications.

Note there are many hidden edges in the majority of the graphs because only edges of weight 10 or greater are shown (except when the number of edges is small).

The graph metrics and clusters are created with all edges taken into account, including single edges. The number of single edges, as well as a select number of other metrics for each graph, are provided. High numbers of single edges should imply a more varied scope of study, or a more interconnected computing discipline.

4 FINDINGS

The findings show relationships among the theses Classifications; the frequency, and variety, of Departments from which theses originate; and the frequencies and varieties of keywords used in the theses.

4.1 ProQuest Overall Numbers

Table 4-1 below shows the total counts of ProQuest theses containing references to each computing discipline from 2009 to 2014, by the position of the computing discipline. IT and ISci show high co-Classifying with only 8% and 6% Solo theses respectively. CS reveals much lower co-Classifying with Solo theses appearing 51% of the time.

For IT, the numbers of Primary versus Secondary theses show a fairly even distribution. From the set of Primary theses, only 15% of IT theses are Solo. CS, on the other hand, from its pool of Primary theses, is 99.8% Solo. ISci is Solo 31% of the time and CE 25% of the time, respectively, when each is Primary. In other words, when Primary, IT has a higher percentage of including another discipline than any of the other three computing disciplines. In contrast, CS has the lowest percentage of including another discipline when it is Primary and thus CS theses appear very strongly single-discipline in nature.

The exceptionally high percentage (99.8%) of CS Primary theses being Solo occurs mostly in the 2009-2014 period. Prior to 2009, CS Primary theses show moderately more co-

Classifications. And, the large drop in co-Classifications appears to coincide with the appearance of CE, ISci, and IT as ProQuest Classification options.

CE shows a low Solo-when-Primary percentage of only 25%, although it is Primary a very high 94% of the time. This indicates that CE, even as Primary, will co-Classify 75% of the time.

ISci as Primary occurs only 18% of the time, which indicates a high co-Classification rate by other disciplines. However, its 31% Solo-when-Primary rate shows ISci is more specific about itself than IT as Primary.

Table 4-1: All Theses by Percentage at Each Position by Discipline (2009-2014)

	Total Number of Theses	Primary	Secondary	Solo	% of Primary That are Solo
CS	21,628	52%	48%	51%	99.8%
CE	5,538	94%	6%	23%	25%
IT	3,748	54%	46%	8%	15%
ISci	2,792	18%	82%	6%	31%

In Figure 4-1 below, looking at the total number of CS theses over time, CS theses are viewed over 30 years (1984-2014) in this instance only, rather than the normal six-year, 2009-2014 time period for this study. ProQuest's Carol Wadke said, in a November, 2014 email to the author, that 1984 is the first reference to CS in their data. She also indicated it was likely already in use by 1984.

CS, as a whole, has seen a steady upward trend. The years from 1992 to 2003 show relatively flat numbers of theses being produced with a dip in 2001 that may be a product of the dot-com boom. The number of total CS theses then shows a dramatic increase from 2001 to 2007.

An interesting feature is the remarkable drop in CS Primary theses and the dramatic rise of CS Secondary theses, especially from 2009 to 2014. The trend appears to begin around 2006, when the number of CS Primary theses actually drops from the previous year despite an overall increase in the number of CS theses.

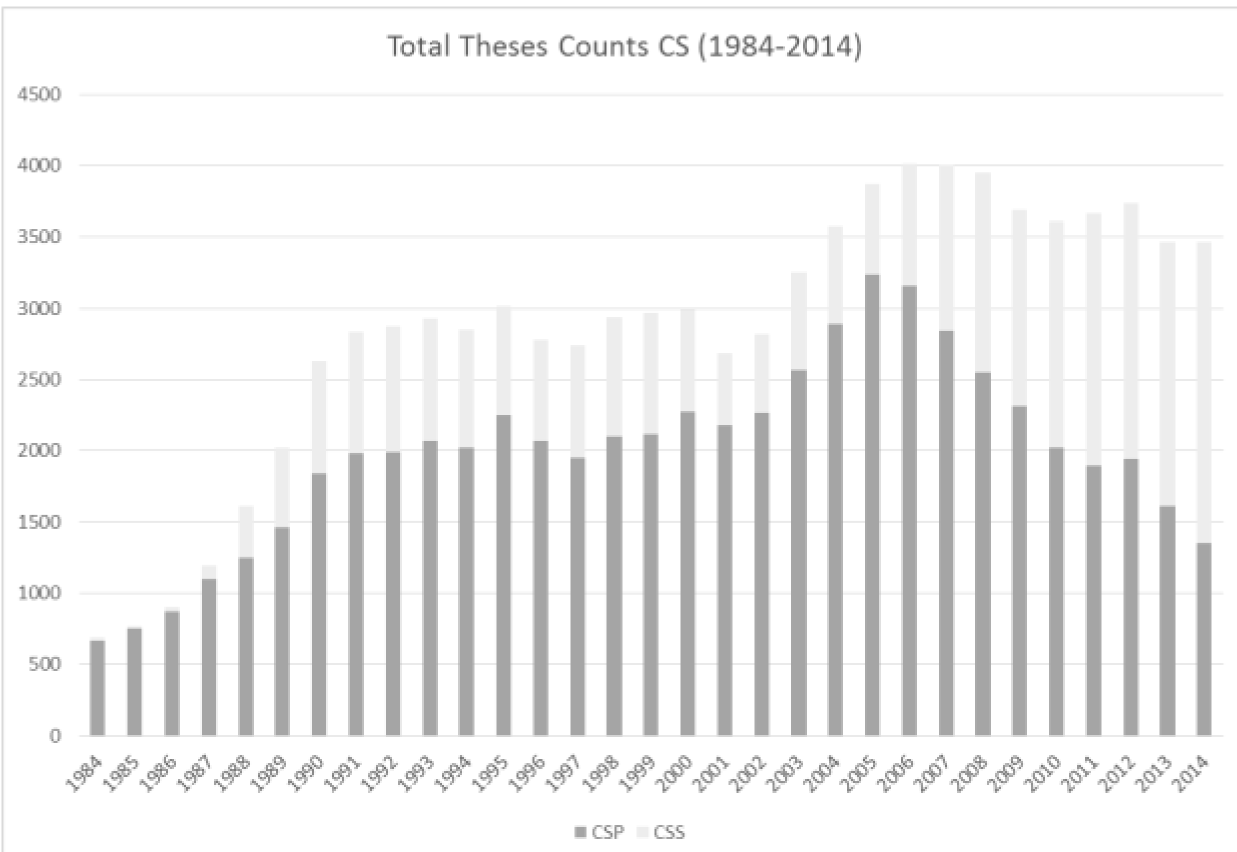


Figure 4-1: CS Thesis Counts by Year, Primary & Secondary (1984-2014)

Coming online in ProQuest around 2009, the number of CE Primary theses dramatically outweighs CE Secondary theses. The overall trend for CE is strongly upwards, with no dips since its introduction in 2009. There appears to be a slight increase in Secondary theses for 2014, however the percentage increase is small at only 0.3%.

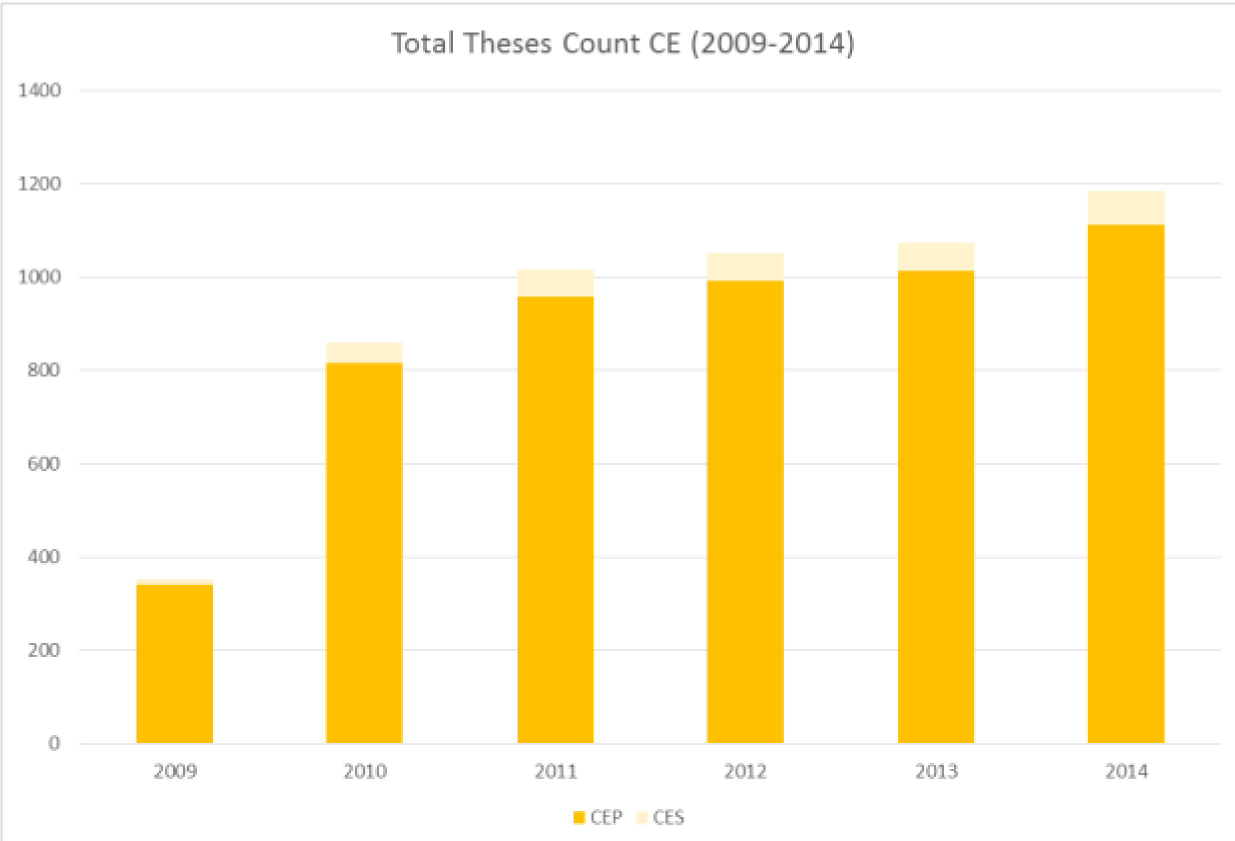


Figure 4-2: CE Thesis Counts by Year, Primary & Secondary (2009-2014)

Information Technology has seen a fairly strong upward trend since its introduction in 2009. The number of IT Primary versus IT Secondary theses has remained fairly even, with the slight edge going to Primary over Secondary.

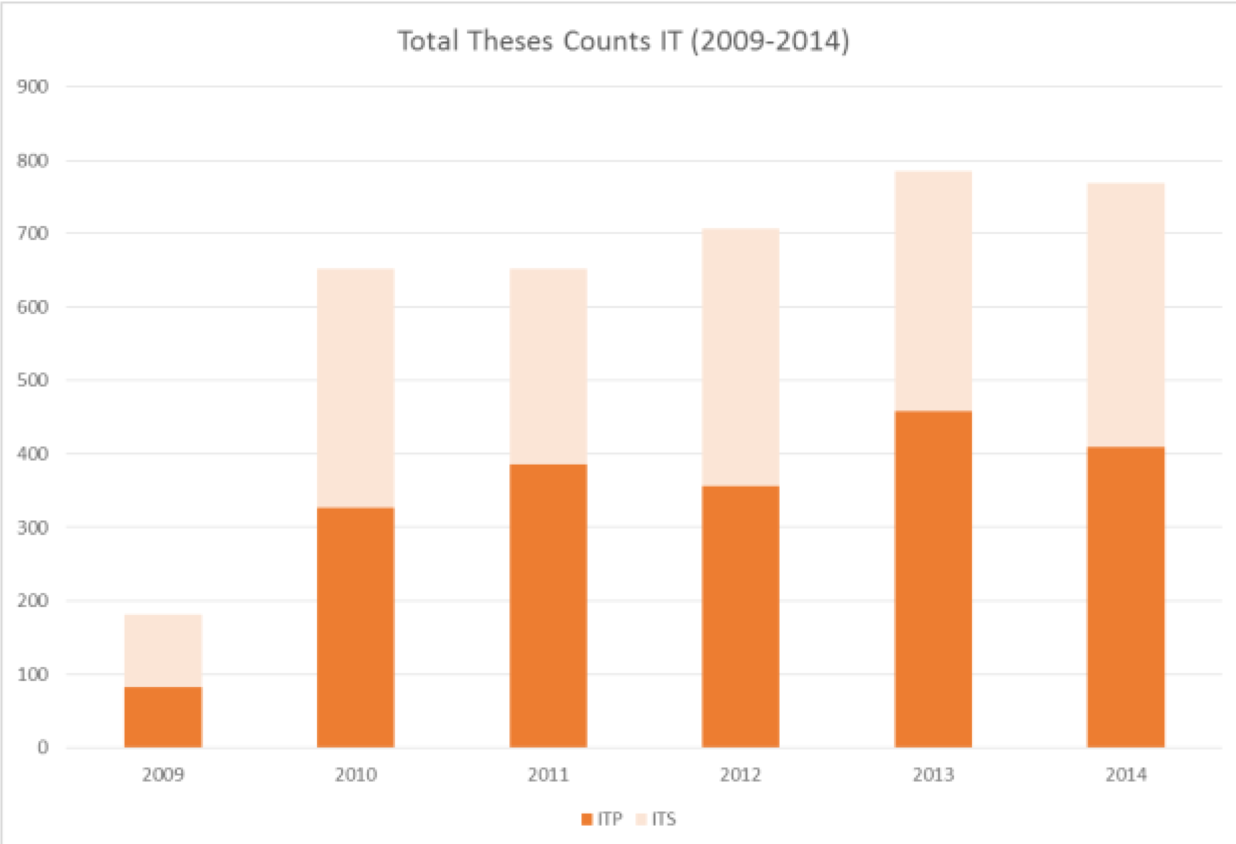


Figure 4-3: IT Thesis Counts by Year, Primary & Secondary (2009-2014)

ISci theses show a steady decline from 2009 till 2013 with an uptick in 2014 in both total numbers as well as in Primary numbers. From Figure 4-3 it doesn't appear that ISci has the same ramp up in numbers beginning in 2009. Going back to 2008 would, however, reveal the same "ramp up" in numbers as IT and CE.

The ratio of Secondary to Primary theses remains high throughout, ranging from a high of 86% in 2011 to a recent low of 74% in 2014.

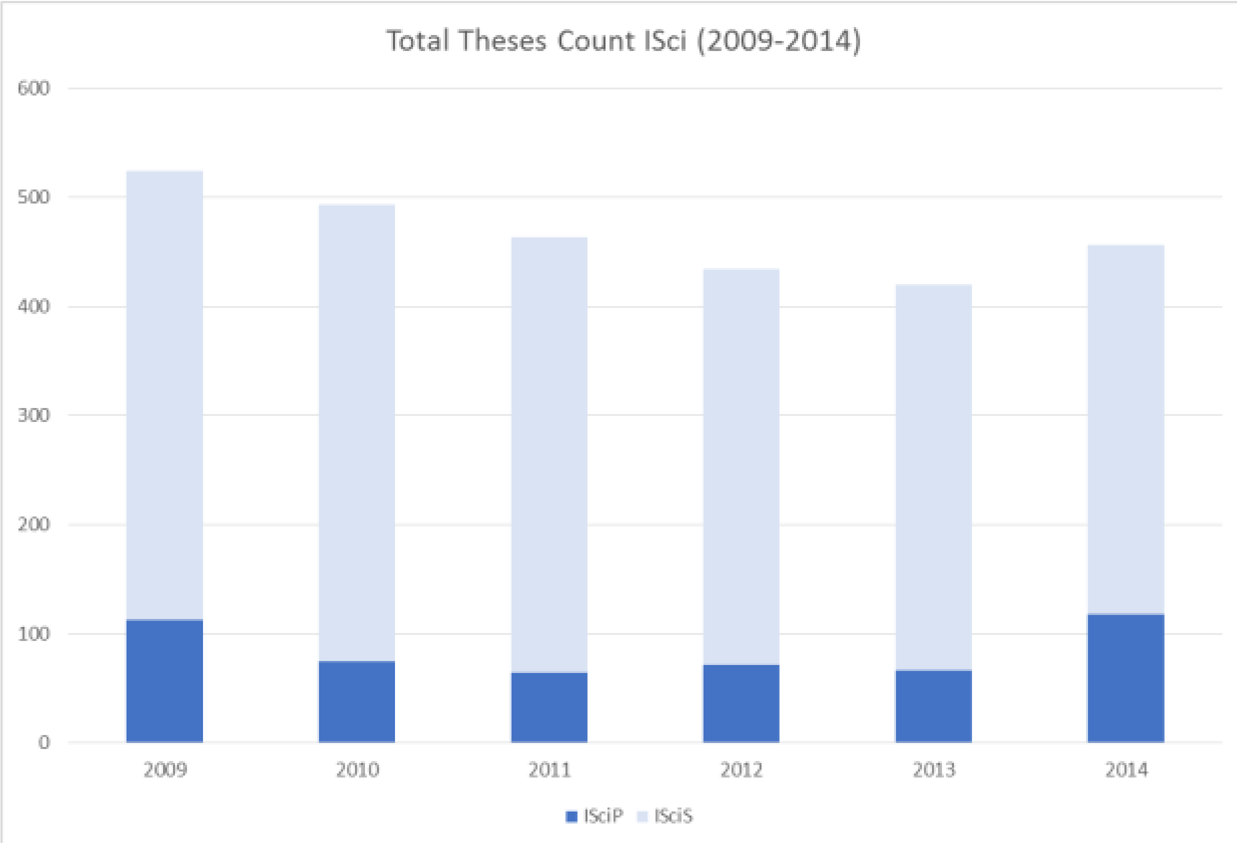


Figure 4-4: ISci Thesis Counts by Year, Primary & Secondary (2009-2014)

As seen below in Figure 4-5, overall, CS still makes up the major share of computing discipline theses in ProQuest with more than all three other disciplines combined. However, the remarkable trend is the overall drop in numbers of CS Primary theses and the corresponding rise in numbers of CE Primary theses.

By 2014 CS Primary and CE Primary theses are almost even, with the difference in the low hundreds versus thousands in previous years. It almost appears as if CS Primary theses are being drained from the CS bar and are filling up the CE bar. Note as well the overall trend for CS is a decrease in theses with increases coming in CE and IT. Again, ISci sees an overall decrease.

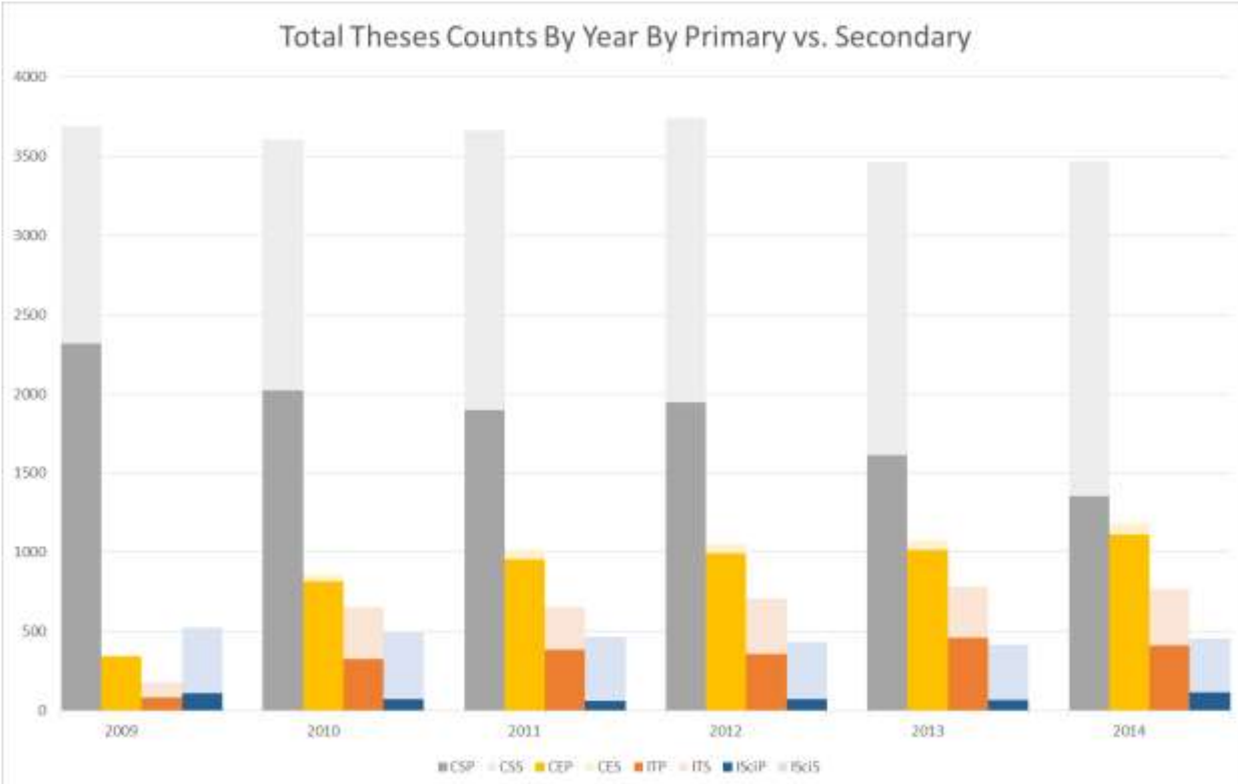


Figure 4-5: Thesis Counts by Year, Primary & Secondary (2009-2014)

Figure 4-6 shows the overall trend for all four computing disciplines combined for the six years from 2009 to 2014 is slightly positive, but there also appears to be a leveling off in the last five years (2010-2014). This would indicate more of a shuffling among disciplines rather than absolute growth in the numbers of computing related theses as CE, IT, and ISci continue to constitute more and more of the total number of computing discipline theses in ProQuest.

Although ISci's absolute numbers trend downwards, it is still part of the three new disciplines that comprise more and more of modern computing research.

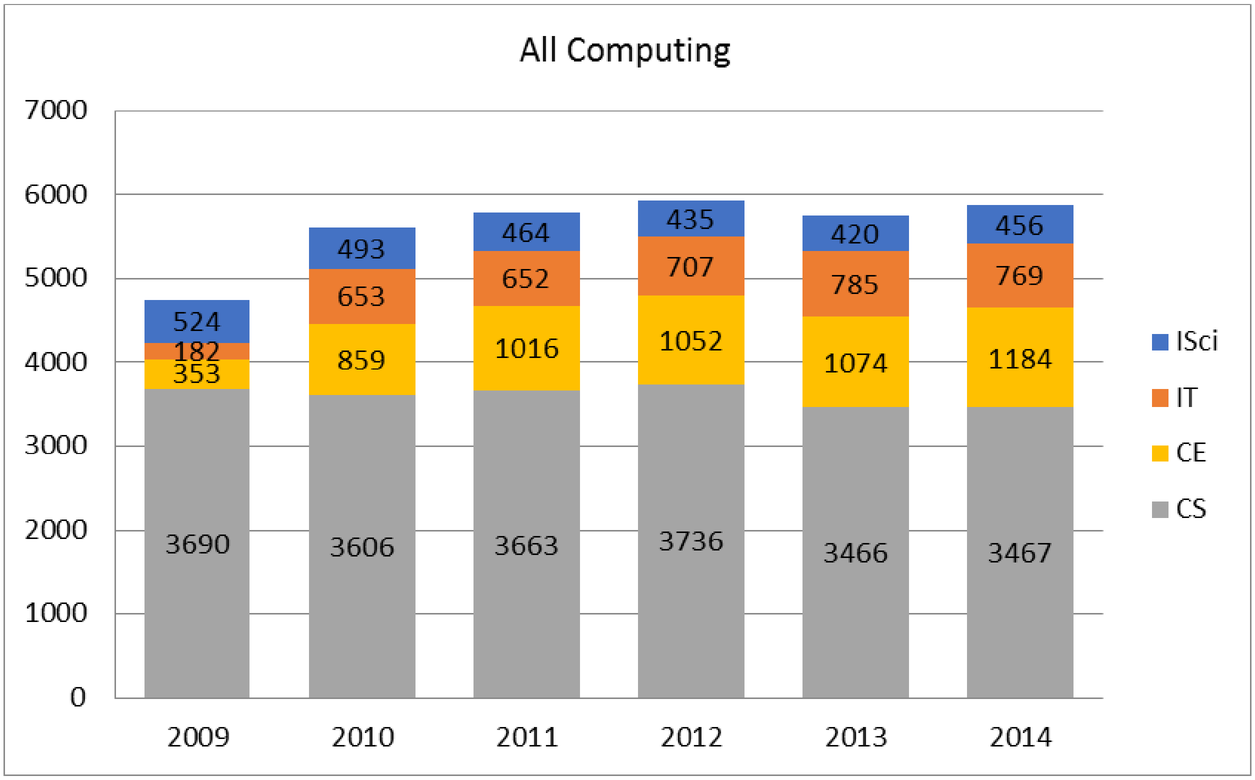


Figure 4-6: Thesis Counts by Year, All Disciplines Combined (2009-2014)

Figure 4-7 shows the Solo theses percentage trends for IT and ISci appear level but CS and CE each show large downward trends, with CS declining from a yearly high of 62% Solo in 2009 to only 39% in 2014, following the decline in Primary classifications. CE declines from a yearly high in 2009 of 43% to a low in 2014 of 15%, even while the number of Primary theses has increased, suggesting that they are more willing to claim secondary Classifications, such as CS, in recent years than in early years.

The percentage of Solo theses for IT and ISci are consistently low at under 10%.

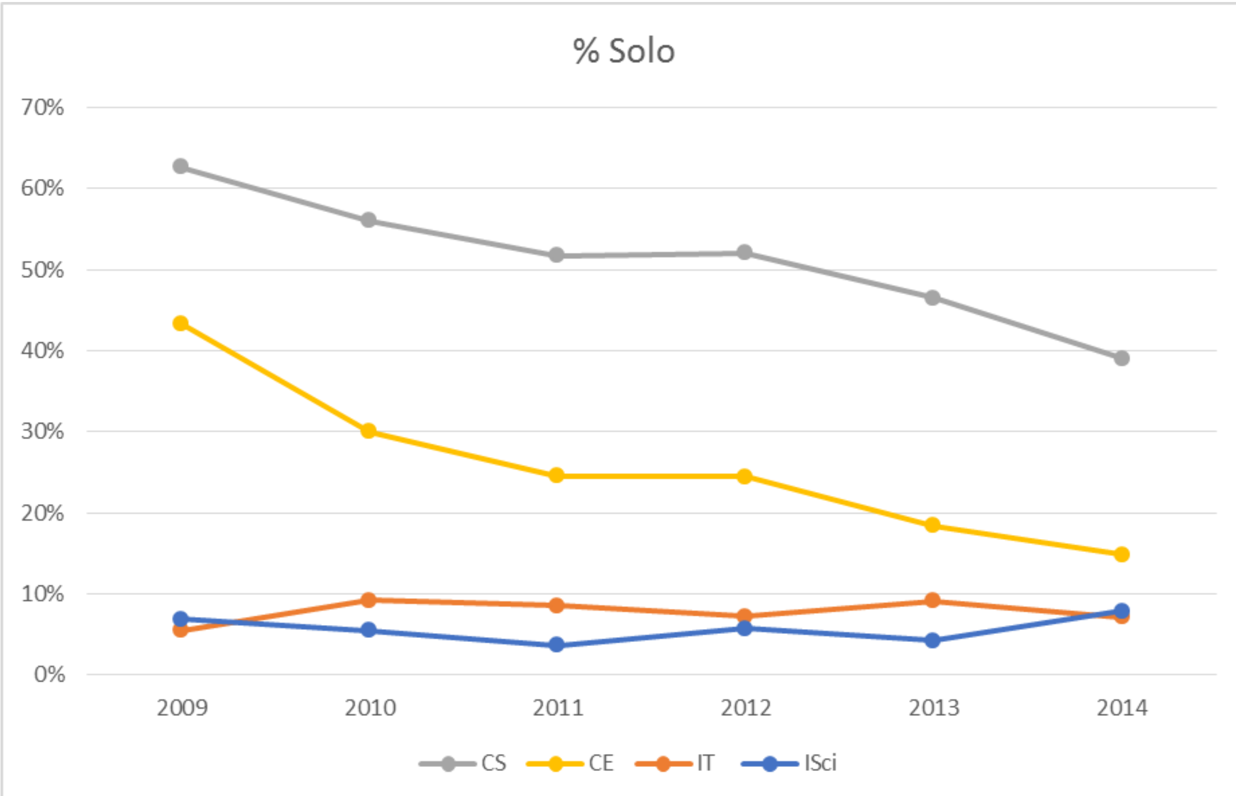


Figure 4-7: Solo Theses Percentage by Year by Discipline (2009-2014)

4.2 Classifications

4.2.1 Overall Classification Comparisons

The data show that the total number of Classifications for any given thesis Classification line is not limited to three (one primary, up to two secondary) per the ProQuest submission instructions. In fact, the max number of Classifications found in any single, computing related thesis for the computing disciplines from 2009 to 2014 is eight. Five theses were found to have eight Classifications. The number of other theses with more than three Classifications are: 10 with seven, 42 with six, 222 with five, and 1,336 with four. The dataset rounds out with: 8,919 with three, 10,285 with two, and 12,887 with one. There are 33,706 total theses in the dataset from 2009 to 2014.

As a sample, the Classification line of one of the five theses with eight classifications follows. Note the inclusion of three (CS, IT, ISci) of the four computing disciplines.

> "0389: Design; 0459: Communication; 0489: Information Technology; 0574: Medical imaging; 0633: Cognitive psychology; 0723: Information science; 0790: Systems science; 0984: Computer science"

Since authors are at liberty to change the Classifications of their works at any time theses with more than three Classifications could be a result of authors that submit with three Classifications initially but add more Classifications after initial submission.

4.2.2 Percentage of Classification Universe

Table 4-2 represents how many of the 411 Classifications found in the 2015-2016 ProQuest submission form are also found with each computing disciplines' theses, by position.

The 411 Classifications represent the universe of Classifications from which thesis authors can choose. The percentages represent which portions of that universe are present in each computing disciplines' theses.

CS only shares 2% of the available Classifications universe when it is Primary, indicating very low levels of co-classifying when Primary as shown earlier. Yet CS shows 75% sharing as secondary, indicating a very high percentage of CS inclusion as a co-Classification.

ISci is also surprisingly low in the amount of the Classification universe it shares when it is Primary, at only 6%. Considering it shares 66% of the Classification universe as Secondary, the 6% for ISci Primary suggests ISci Primary has a highly focused set of co-Classifications.

IT shows modest levels of sharing as Primary at 32% and, although not as high as CS or ISci, shares over half of the Classification universe as a Secondary.

CE is actually the opposite of the other three disciplines. CE Primary sharing is slightly higher than its Secondary counterpart, 28% to 26% respectively, although these are still roughly the same. This means not a lot of other disciplines co-Classify with CE, and CE itself will co-Classify with a smaller set of Classifications.

Note that the Any column in the Table 4-2 will normally be less than the Primary plus the Secondary because Any removes any overlapping Classifications that appear individually in Primary and Secondary. In Table 4-2 below the percentages are based on the 411 available ProQuest Classifications for 2015-2016 (*ProQuest ProQuest 2015-2016 Publishing Agreement*).

Table 4-2: Statistics for CD-related Classifications (2009-2014)

Discipline	% Shared Any	% Shared Primary	% Shared Secondary
CS	75%	2%	75%
CE	44%	28%	26%
IT	63%	32%	55%
ISci	66%	6%	66%

Classification codes are used to account for potential changes in Classification names. If a Classification code is found in the data, but is not found among the 2015-2016 codes, then that Classification is not included in the set of Classifications used to compute the percentages for the computing discipline in Table 4-2. These codes represent Classifications that were available at some time in 2009-2014 but are no longer available in the Classification universe. There were 10 distinct code/name combinations that were present in the data from 2009-2014 but were not found in the 2015-2016 submission form. These are shown below.

Table 4-3: Classification Codes and Names Not Present in 2015-2016 (2009-2014)

Classification Name	Code	Number of Times Present
HEALTH EDUCATION	0350	5
NEUROBIOLOGY	0421	13
PHYSIOLOGY	0433	2
SCHOOL ADMINISTRATION	0461	2
PHARMACY SCIENCES	0491	8
LEGAL STUDIES	0553	6
SOLID STATE PHYSICS	0600	3
ATMOSPHERIC SCIENCES	0608	2
NUCLEAR PHYSICS	0610	2
BIOPHYSICS	0760	9

For the complete dataset there are 10 codes but 14 different code/name combinations that are no longer in the current Classification universe, as shown below in Table 4-4. Note the changing names for Classfication codes 0433, 0491, and 0608.

Table 4-4: Classification Codes and Names Not Present in 2015-2016 (1937-2014)

Classification Name	Code	Number of Times Present
HEALTH EDUCATION	0350	77
NEUROBIOLOGY	0421	13
ANATOMY & PHYSIOLOGY	0433	15
ANIMALS	0433	15
PHYSIOLOGY	0433	4
SCHOOL ADMINISTRATION	0461	2
PHARMACOLOGY	0491	28
PHARMACY SCIENCES	0491	10
LEGAL STUDIES	0553	6
SOLID STATE PHYSICS	0600	3
ATMOSPHERE	0608	70
ATMOSPHERIC SCIENCES	0608	2
NUCLEAR PHYSICS	0610	20
BIOPHYSICS	0760	86

4.2.3 Co-Classifications by Computing Discipline

The following abbreviations and descriptions are necessary to understand the following tables and the rest of the findings and discussion:

- **CSA**: CS as Any
- **CSP**: CS as Primary
- **CSS**: CS as Secondary
  - Note that which of the Secondary positions (e.g., second, third, …) is not accounted for as differences in the Secondary position is given no importance in the ProQuest submission instructions.
- **PCSS**: Primary when CS is a Secondary
  - This is not a count of co-Classifications as the other three columns are, but it is strictly a count of how many times the given Classification appears as the Primary when CS is a Secondary. This number will be different than the CSS column, as the CSS column includes non-CS Classifications at all positions, not just the Primary position.
- **CSO**: CS as Solo, or Only
- The other disciplines will follow the same conventions with the following exceptions:
  - ISci is shortened to ISA, ISP, etc. instead of ISciA, ISciP, etc. in the tables and discussion below; and the 'Primary when ISci is a Secondary' is shortened to PICS.

4.2.4 Co-Classifications for CS

CS appears to be an anomaly because of its stark lack of co-Classifications when it is Primary. There are only six distinct co-Classifications for CSP. These six only occur a total of 19 times over the six year span from 2009-2014. 2006 was the last time there were any appreciable numbers of co-Classifications with CSP.

The following two tables illustrate the top five co-Classifications for CS in 2006 and 2007. Note the drop from 218 instances of co-Classifications in the top five for CSP in 2006 to only 5 instances in the top five for 2007. There are appreciable jumps in two categories: ARTIFICIAL INTELLIGENCE and BIOINFORMATICS for CSS and PCSS from 2006 to 2007, however.

Table 4-5: Top 5 Co-Classifications with CS in 2006

CSP		CSS		PCSS	
ELECTRICAL ENGINEERING	88	ELECTRICAL ENGINEERING	383	ELECTRICAL ENGINEERING	359
ARTIFICIAL INTELLIGENCE	60	MATHEMATICS	72	MATHEMATICS	64
BIOINFORMATICS	33	INFORMATION SYSTEMS	51	INFORMATION SYSTEMS	33
INFORMATION SYSTEMS	21	ARTIFICIAL INTELLIGENCE	46	BIOINFORMATICS	29
BIOMEDICAL RESEARCH	16	BIOINFORMATICS	41	MANAGEMENT	28

Table 4-6: Top 5 Co-Classifications with CS in 2007

CSP		CSS		PCSS	
ACOUSTICS	2	ELECTRICAL ENGINEERING	410	ELECTRICAL ENGINEERING	388
BIOINFORMATICS	2	ARTIFICIAL INTELLIGENCE	161	ARTIFICIAL INTELLIGENCE	113
ARMED FORCES	1	BIOINFORMATICS	108	BIOINFORMATICS	92
ARTIFICIAL INTELLIGENCE	1	MATHEMATICS	75	MATHEMATICS	72
BIOMEDICAL RESEARCH	1	INFORMATION SYSTEMS	51	ROBOTS	40

The top 20 co-Classifications for CS in the 2009-2014 time period are shown in Table 4-7 below.

With CSP, the top five account for 95% of all co-Classifications and naturally the top 20 account for 100% (there are only six distinct co-Classifications and 19 instances, or occurrences, of those co-Classifications for CSP). With CSS, the top five account for 45% of all co-Classifications and the top 20, 73%. With PCSS the top five Classifications account for 48% of all Primaries when CS is a Secondary and the top 20, 76%. This second-to-the-last metric means almost 50% of the theses that place CS as Secondary are accounted for by a set of five Classifications positioning themselves as Primary. These five are COMPUTER ENGINEERING, ELECTRICAL ENGINEERING, INFORMATION TECHNOLOGY, APPLIED MATHEMATICS, and BIOINFORMATICS. ARTIFICIAL INTELLIGENCE is sixth. The fact that CE is number one of the five shows a reason why as the number of CE Primary theses increase, the number of CS Primary theses decrease—given that the absolute numbers of computing discipline theses has been relatively stable for the last five years.

Notable is that CSP and CSS share no Classifications among their top five spots.

PCSS has 269 co-Classifications and shares two with CSP: SYSTEMS SCIENCE (26th PCSS, 6th CSP) and OPERATIONS RESEARCH (31st PCSS, 5th CSP).

Under PCSS, MANAGEMENT and EDUCATIONAL TECHNOLOGY are 16th and 20th respectively. These are notable for their often-strong connection to computing.

Table 4-7: Top 20 Co-Classifications, CS (2009-2014)

CSA		CSP		CSS		PCSS	
COMPUTER ENGINEERING	2,213	ACOUSTICS	10	COMPUTER ENGINEERING	2,213	COMPUTER ENGINEERING	2,109
ELECTRICAL ENGINEERING	2,034	TEXTILE RESEARCH	3	ELECTRICAL ENGINEERING	2,034	ELECTRICAL ENGINEERING	1,196
ARTIFICIAL INTELLIGENCE	1,104	ONCOLOGY	2	ARTIFICIAL INTELLIGENCE	1,104	INFORMATION TECHNOLOGY	655
INFORMATION TECHNOLOGY	860	PHYSIOLOGICAL PSYCHOLOGY	2	INFORMATION TECHNOLOGY	860	APPLIED MATHEMATICS	524
BIOINFORMATICS	840	OPERATIONS RESEARCH	1	BIOINFORMATICS	840	BIOINFORMATICS	510
INFORMATION SCIENCE	677	SYSTEMS SCIENCE	1	INFORMATION SCIENCE	677	ARTIFICIAL INTELLIGENCE	500
APPLIED MATHEMATICS	534		0	APPLIED MATHEMATICS	534	STATISTICS	308
ROBOTICS	435		0	ROBOTICS	435	MATHEMATICS	284
STATISTICS	396		0	STATISTICS	396	ROBOTICS	263
MATHEMATICS	367		0	MATHEMATICS	367	INFORMATION SCIENCE	260
WEB STUDIES	272		0	WEB STUDIES	272	BIOMEDICAL ENGINEERING	207
BIOMEDICAL ENGINEERING	269		0	BIOMEDICAL ENGINEERING	269	WEB STUDIES	162
MECHANICAL ENGINEERING	230		0	MECHANICAL ENGINEERING	230	LINGUISTICS	160
ENGINEERING	204		0	ENGINEERING	204	NEUROSCIENCES	133
EDUCATIONAL TECHNOLOGY	201		0	EDUCATIONAL TECHNOLOGY	201	MECHANICAL ENGINEERING	131
OPERATIONS RESEARCH	194		0	OPERATIONS RESEARCH	193	MANAGEMENT	115
SYSTEMS SCIENCE	173		0	SYSTEMS SCIENCE	172	ENGINEERING	113
COGNITIVE PSYCHOLOGY	170		0	COGNITIVE PSYCHOLOGY	170	GENETICS	92
LINGUISTICS	161		0	LINGUISTICS	161	AEROSPACE ENGINEERING	91
MEDICAL IMAGING	160		0	MEDICAL IMAGING	160	EDUCATIONAL TECHNOLOGY	89

4.2.5 Co-Classifications for CE

The top 20 co-Classifications for CE in the 2009-2014 time period are shown in Table 4-8 below.

When CE is Primary (CEP) the top five account for 79% of all co-Classifications and the top 20 account for 94%. This indicates low variation in co-Classifications for CEP. With CE Secondary (CES) the top five account for 48% of all co-Classifications and the top 20, 75%, indicating a broader range of co-Classifications for CES than CEP.

With 'Primary when CE is Secondary' (PCES) the top five account for 54% and the top 20, 83%. This indicates a slightly more focused set of Classification Primaries that include CE as a Secondary versus all of CE's co-Classifications.

CEP and PCES share no Classifications. None of the 117 distinct co-Classifications found under CEP uses a Classification found among the 57 distinct Classifications under PCES. CEP has 5,632 co-Classification instances and there are 307 PCES instances. The top five PCES fields are APPLIED MATHEMATICS, STATISTICS, MATHEMATICS, NEUROSCIENCES, and COMMUNICATION. This finding shows CE never co-Classifies itself with any of these five, or any of the 307 under PCES, when it is Primary in the dataset.

The top co-Classification for CEP and CES is CS. For CES, since CS has shown that it generally does not co-Classify with other disciplines when it is Primary this means that another discipline, or other disciplines, must be co-Classifying CE and CS together, along with itself. Figure 4-13 below will show these are mainly MATHEMATICS, APPLIED MATHEMATICS, and STATISTICS, as they all have a strong connection to CS in the CES graph.

Table 4-8: Top 20 Co-Classifications, CE (2009-2014)

CEA		CEP		CES		PCES	
COMPUTER SCIENCE	2,212	COMPUTER SCIENCE	2,108	COMPUTER SCIENCE	104	APPLIED MATHEMATICS	60
ELECTRICAL ENGINEERING	1,930	ELECTRICAL ENGINEERING	1,875	APPLIED MATHEMATICS	61	STATISTICS	35
INFORMATION TECHNOLOGY	192	ENGINEERING	184	ELECTRICAL ENGINEERING	55	MATHEMATICS	26
ENGINEERING	184	INFORMATION TECHNOLOGY	178	STATISTICS	40	NEUROSCIENCES	26
BIOMEDICAL ENGINEERING	135	BIOMEDICAL ENGINEERING	121	MATHEMATICS	37	COMMUNICATION	18
ARTIFICIAL INTELLIGENCE	124	ARTIFICIAL INTELLIGENCE	117	NEUROSCIENCES	26	GEOGRAPHIC INFORMATION SCIENCE	10
ROBOTICS	121	ROBOTICS	117	COMMUNICATION	19	MANAGEMENT	10
MECHANICAL ENGINEERING	87	MECHANICAL ENGINEERING	85	BIOMEDICAL ENGINEERING	14	ALTERNATIVE ENERGY	9
APPLIED MATHEMATICS	61	BIOINFORMATICS	53	INFORMATION TECHNOLOGY	14	MUSIC	9
BIOINFORMATICS	61	SYSTEMS SCIENCE	51	GEOGRAPHIC INFORMATION SCIENCE	11	DESIGN	8
SYSTEMS SCIENCE	55	AEROSPACE ENGINEERING	50	MANAGEMENT	10	BEHAVIORAL PSYCHOLOGY	6
INFORMATION SCIENCE	54	INFORMATION SCIENCE	49	ALTERNATIVE ENERGY	9	BUSINESS ADMINISTRATION	5
AEROSPACE ENGINEERING	52	NANOTECHNOLOGY	49	DESIGN	9	LINGUISTICS	5
NANOTECHNOLOGY	51	MEDICAL IMAGING	45	MUSIC	9	FINE ARTS	4
MEDICAL IMAGING	49	ENERGY	43	BEHAVIORAL PSYCHOLOGY	8	OPHTHALMOLOGY	4
ENERGY	48	OPTICS	36	BIOINFORMATICS	8	PHYSICAL THERAPY	4
OPTICS	41	CIVIL ENGINEERING	32	ARTIFICIAL INTELLIGENCE	7	SOCIAL RESEARCH	4
STATISTICS	40	MULTIMEDIA COMMUNICATIONS	31	INDUSTRIAL ENGINEERING	6	SYSTEMATIC	4
CIVIL ENGINEERING	37	INDUSTRIAL ENGINEERING	30	BUSINESS ADMINISTRATION	5	AUDIOLOGY	3
MATHEMATICS	37	MATERIALS SCIENCE	22	CIVIL ENGINEERING	5	ECOLOGY	3

4.2.6 Co-Classifications for IT

The top 20 co-Classifications for IT in the 2009-2014 time period are shown in Table 4-9 below.

When IT is Primary (ITP) the top five account for 54% of all co-Classifications and the top 20 account for 78%. With IT Secondary (ITS) the top five account for 42% of all co-Classifications and the top 20, 69%. With ‘Primary when IT is Secondary’ (PITS) the top five account for 66% and the top 20, 85%.

The Classifications ITP shares with ITS among their top five spots are COMPUTER SCIENCE and INFORMATION SCIENCE. These are in the first and second positions for ITP and second and fifth positions for ITS, respectively. None of the top five PITS Classifications are found in the 131 co-Classifications of ITP. Only three of the 95 Classifications in PITS are found among the 131 in ITP.

Note again that for ITS, IT could be second, third, or any position besides first, and a co-Classification could be first, second, third, fourth, or any position besides IT’s position for that thesis.

Noticeable is the high level of ITP co-Classifying with CS. This also contributes to the increase in CSS theses. ITP co-Classifications with CS increased from 34 in 2009 to a high of 173 in 2013. It dips in 2014 to 130.

MANAGEMENT, as Primary, is by far the largest co-Classifier of IT when IT is a Secondary. There are 617 MANAGEMENT instances under PITS and 705 under ITS. The latter is larger than the former because under ITS, MANAGEMENT may also be a Secondary along with IT with a different Primary. As noted earlier, there is no INFORMATION SYSTEMS

Classification. This high number suggests that many theses in management programs may be using the IT Classification in place of a nonexistent INFORMATION SYSTEM Classification.

4.2.7 **Co-Classifications for ISci**

The top 20 co-Classifications for ISci in the 2009-2014 time period are shown below in Table 4-10.

For ISP the top five account for 90% of all co-Classifications and the top 20 account for 99.3%. With ISS the top five account for 38% of all co-Classifications and the top 20, 63%. With PICS the top five account for 43% and the top 20, 66%.

The only Classification ISP shares with ISS among their top five spots is COMPUTER SCIENCE, which is first for ISP and second for ISS.

The top one ISP co-Classification, COMPUTER SCIENCE, makes up 63% of all ISP co-Classifications. It is five times more frequent than the second most frequent co-Classification, HEALTH CARE MANAGEMENT.

There are only 23 distinct co-Classifications for ISP. None of the 192 PICS Classifications are found among the 23 co-Classifications of ISP.

ISci also contributes to the increasing number of CSS theses as it, with CE and IT, uses CS as the top co-Classification when it is Primary.

Also interesting is that among the Classifications found under the same ProQuest sub-category heading of “COMMUNICATIONS AND INFORMATION SCIENCES,” LIBRARY SCIENCE, COMMUNICATIONS, and WEB STUDIES, all of the latter can be found as a Primary when ISci is Secondary, but, when ISci is Primary ISci never includes any of them as a co-Classification.

Table 4-9: Top 20 Co-Classifications, IT (2009-2014)

ITA		ITP		ITS		PITS	
COMPUTER SCIENCE	861	COMPUTER SCIENCE	656	MANAGEMENT	705	MANAGEMENT	617
MANAGEMENT	705	INFORMATION SCIENCE	339	COMPUTER SCIENCE	205	COMPUTER ENGINEERING	178
INFORMATION SCIENCE	508	HEALTH CARE MANAGEMENT	228	COMPUTER ENGINEERING	192	BUSINESS ADMINISTRATION	173
HEALTH CARE MANAGEMENT	290	EDUCATIONAL TECHNOLOGY	135	BUSINESS ADMINISTRATION	173	COMMUNICATION	104
EDUCATIONAL TECHNOLOGY	225	WEB STUDIES	100	INFORMATION SCIENCE	169	MARKETING	73
WEB STUDIES	197	ELECTRICAL ENGINEERING	83	COMMUNICATION	160	BEHAVIORAL PSYCHOLOGY	40
COMPUTER ENGINEERING	192	NURSING	61	ORGANIZATIONAL BEHAVIOR	116	SOCIAL RESEARCH	32
ORGANIZATIONAL BEHAVIOR	177	ORGANIZATIONAL BEHAVIOR	61	WEB STUDIES	97	HIGHER EDUCATION ADMINISTRATION	31
BUSINESS ADMINISTRATION	173	HIGHER EDUCATION	49	EDUCATIONAL TECHNOLOGY	90	DESIGN	28
COMMUNICATION	160	MULTIMEDIA COMMUNICATIONS	44	MARKETING	78	EDUCATIONAL LEADERSHIP	28
ELECTRICAL ENGINEERING	108	SYSTEMS SCIENCE	42	HEALTH CARE MANAGEMENT	62	LIBRARY SCIENCE	22
MULTIMEDIA COMMUNICATIONS	81	PUBLIC HEALTH	40	BEHAVIORAL PSYCHOLOGY	45	SOCIAL PSYCHOLOGY	22
MARKETING	78	PUBLIC POLICY	40	ORGANIZATION THEORY	43	CULTURAL ANTHROPOLOGY	19
HIGHER EDUCATION	74	ECONOMICS	38	OCCUPATIONAL PSYCHOLOGY	42	ACCOUNTING	17
PUBLIC ADMINISTRATION	72	TECHNICAL COMMUNICATION	38	DESIGN	38	WOMENS STUDIES	17
ECONOMICS	69	PUBLIC ADMINISTRATION	37	MULTIMEDIA COMMUNICATIONS	37	GEOGRAPHIC INFORMATION SCIENCE	16
NURSING	68	ARTIFICIAL INTELLIGENCE	36	HIGHER EDUCATION ADMINISTRATION	35	LINGUISTICS	15
ORGANIZATION THEORY	67	CRIMINOLOGY	33	PUBLIC ADMINISTRATION	35	APPLIED MATHEMATICS	14
SYSTEMS SCIENCE	61	BIOINFORMATICS	32	EDUCATIONAL LEADERSHIP	34	GEOGRAPHY	14
OCCUPATIONAL PSYCHOLOGY	58	INDUSTRIAL ENGINEERING	30	SOCIAL PSYCHOLOGY	32	INSTRUCTIONAL DESIGN	14

Table 4-10: Top 20 Co-Classifications, ISci (2009-2014)

ISA		ISP		ISS		PICS	
COMPUTER SCIENCE	677	COMPUTER SCIENCE	260	INFORMATION TECHNOLOGY	507	INFORMATION TECHNOLOGY	338
INFORMATION TECHNOLOGY	507	HEALTH CARE MANAGEMENT	51	COMPUTER SCIENCE	417	LIBRARY SCIENCE	241
WEB STUDIES	308	ARTIFICIAL INTELLIGENCE	38	WEB STUDIES	308	MANAGEMENT	217
LIBRARY SCIENCE	275	SYSTEMS SCIENCE	13	LIBRARY SCIENCE	275	COMMUNICATION	96
MANAGEMENT	247	OPERATIONS RESEARCH	11	MANAGEMENT	247	WEB STUDIES	88
HEALTH CARE MANAGEMENT	162	HIGHER EDUCATION	10	EDUCATIONAL TECHNOLOGY	161	BUSINESS ADMINISTRATION	50
EDUCATIONAL TECHNOLOGY	161	MUSEUM STUDIES	4	COMMUNICATION	149	COMPUTER ENGINEERING	49
COMMUNICATION	149	CURRICULUM DEVELOPMENT	3	MASS COMMUNICATIONS	127	MARKETING	48
MASS COMMUNICATIONS	127	ENERGY	3	HEALTH CARE MANAGEMENT	111	ELECTRICAL ENGINEERING	41
HIGHER EDUCATION	76	ARCHITECTURE	2	HIGHER EDUCATION	66	EDUCATIONAL TECHNOLOGY	39
ELECTRICAL ENGINEERING	63	ATMOSPHERIC SCIENCES	2	ELECTRICAL ENGINEERING	63	SOCIAL PSYCHOLOGY	34
MULTIMEDIA COMMUNICATIONS	61	BANKING	2	MULTIMEDIA COMMUNICATIONS	61	SOCIAL RESEARCH	33
ORGANIZATIONAL BEHAVIOR	59	FILM STUDIES	2	ORGANIZATIONAL BEHAVIOR	59	GEOGRAPHY	32
PUBLIC HEALTH	57	MILITARY STUDIES	2	PUBLIC HEALTH	57	MASS COMMUNICATIONS	31
ARTIFICIAL INTELLIGENCE	56	RECREATION	2	COMPUTER ENGINEERING	54	PUBLIC HEALTH	30
COMPUTER ENGINEERING	54	REMOTE SENSING	2	BUSINESS ADMINISTRATION	52	BIOINFORMATICS	29
BUSINESS ADMINISTRATION	52	URBAN PLANNING	2	MARKETING	52	MULTIMEDIA COMMUNICATIONS	29
MARKETING	52	EPIDEMIOLOGY	1	BIOINFORMATICS	50	DESIGN	25
BIOINFORMATICS	50	HISPANIC AMERICAN STUDIES	1	SOCIAL PSYCHOLOGY	48	ENGINEERING	25
SOCIAL PSYCHOLOGY	48	ONCOLOGY	1	POLITICAL SCIENCE	47	NURSING	24

4.2.8 Classfication Graphs

Using NodeXL and the Wakita-Tsurumi algorithm, as well as the other metrics and settings as described in Chapter 3, the following graphs show groupings of general areas of study found for each of the computing disciplines. These groupings are like the general headings and sub-headings in the ProQuest Subject Categories guide such as “EDUCATION” or “FINE AND PERFORMING ARTS.”

If the proceeding graphs had a z-axis, pointing out of the plane, every node seen would have a connection to the main computing discipline node “in the sky.” Thus, the tables provide information on what Classifications connect to the computing disciplines but the graphs show how those Classifications that connect to the computing disciplines relate to each other. At a glance each graph conveys what goes with what, and what goes with those other “what’s.” In other words, each graph conveys what areas of study are common with other areas of study (when also common with one of the computing disciplines). And, then goes on to show, even if they do not share the same thesis, with what *other* areas of study they are related (again, when connected with computing).

4.2.9 Classification Graphs for CS

Below are the summary metrics for CS in its three respective positions.

Table 4-11: Summary Graph Metrics for CS

	CSA	**CSP**	**CSS**
Groups	14	1	14
Number of vertices	302	2	302
Distinct edges	1,837	1	1,837
Total edges	6,086	1	6,085
Edges of weight 1	1,170	1	1,170
Distinct edges of weight > 1	667	0	667
Edges of weight > 1	4,916	0	4,915

Higher numbers for the ratios below, for all the computing disciplines, indicates greater diversity in co-Classifications. Another way of putting it, the more single edges are present, the more Classifications are needed to create the single edges.

1170/4916 the ratio of single to >1 edges is 24% for CSA.

There is insufficient data for the ratio of single edges to >1 edges for CSP.

1170/4915 the ratio of single to >1 edges is 24% for CSS.

1837/6086 the ratio of distinct edges to total edges is 30% for CSA

There is insufficient data for the ratio of distinct edges to total edges for CSP.

1837/6086 the ratio of distinct edges to total edges is 30% for CSS

4.2.10 CSA Findings

The CSA graph shows the general landscape of research in CS from within CS and from other disciplines that include CS in their research.

Betweenness centrality measures the number of shortest paths from all nodes to all other nodes that pass through a given node. Large nodes, representing nodes with high betweenness centrality, indicate Classifications that are likely to appear with any of its "bridged" Classifications. "Bridging" here means that the high betweenness centrality node is either the shortest path, or sometimes the only path, between other nodes in the graph and the node connected to the high betweenness node. So, for example, the large EDUCATIONAL TECHNOLOGY node below in Figure 4-8, with only three visible edges, 1) has more edges than are shown because of the minimum 10 edge cutoff, and 2) is large because it is either the shortest, or occasionally the only, path to get to one of its 88 co-Classified edges . Besides the two visible Classifications, SCIENCE EDUCATION and EDUCATION SCIENCE, that appear to have no other connections but to EDUCATIONAL TECHNOLOGY, there are 85 more that are not shown, 51 of which are single edges. Any edges between the 85 nodes and others nodes are also not shown because they do not meet the 10 edge minimum cutoff.

Intuitively, betweenness centrality identifies vertices that are important because they are unique (or one of a few) connectors to other vertices. For example, EDUCATIONAL TECHNOLOGY has a high betweenness centrality because it is one of the few vertices connected to several other vertices (e.g. EDUCATION TEST AND MEASUREMENTS or VOCATIONAL EDUCATION).

Table 4-7 shows that the co-Classifications CS makes with other fields is due almost entirely to CSS.

All three of the other computing disciplines are present and each brings with it connections to a regular set of other Classifications. A few of the main groupings present focus on IT and ISci along with WEB STUDIES, LIBRARY SCIENCE, and MANAGEMENT. Another group seems to focus on the engineering disciplines with COMPUTER and ELECTRICAL ENGINEERING, along with CIVIL, MECHANICAL, AEROSPACE, and BIOMEDICAL ENGINEERING. There is an education group anchored by EDUCATION TECHNOLOGY. There is a biological sciences group clustered around BIOINFORMATICS. A mathematics group is present, clustered around APPLIED MATHEMATICS with STATISTICS and MATHEMATICS. ARTIFICIAL INTELLGENCE is present, connecting many nodes, and bridging studies into diverse areas such as MUSIC, LINGUISTICS, and psychology fields such as COGNITIVE and EXPERIMENTAL PSYCHOLOGY. Three more groups appear to center around geography, operations and industrial processes, and fine arts.

Note that these connections would not be easily seen or noticed, if noticeable at all, without graph and clustering technology.

11 of the 14 Wakita-Tsurumi groups are present in the graph. Those not present due to the weighting requirement are two groups consisting of two nodes each and another with four. The last deals with philosophy and the former two deal with: first, wildlife and second, language and sociology education.

4.2.11 **CSP Findings**

CSP has almost no co-Classifications, as seen in Table 4-7. Moreso, the graph in Figure 4-9 below shows a graph with only two nodes and one edge, the node having an edge weight of one. This shows there is only one thesis with more than two Classifications in the CSP data, and that thesis had exactly three Classifications with CS as the Primary.

4.2.12 **CSS Findings**

Due to CSP having so few co-Classifications, the graph for CSS is nearly identical to CSA. However, the groupings actually change for CSS. IT moves into its own cluster becoming an anchor for MANAGEMENT and BUSINESS ADMINISTRATION along with ARTIFICAL INTELLIGENCE. EDUCATION TECHNOLOGY moves into the same cluster as INFORMATION SCIENCE.

ELECTRICAL ENGINEERING and COMPUTER ENGINEERING naturally keep their strong connection but apparently move into a group unto themselves.

4.2.13 CSA Graph

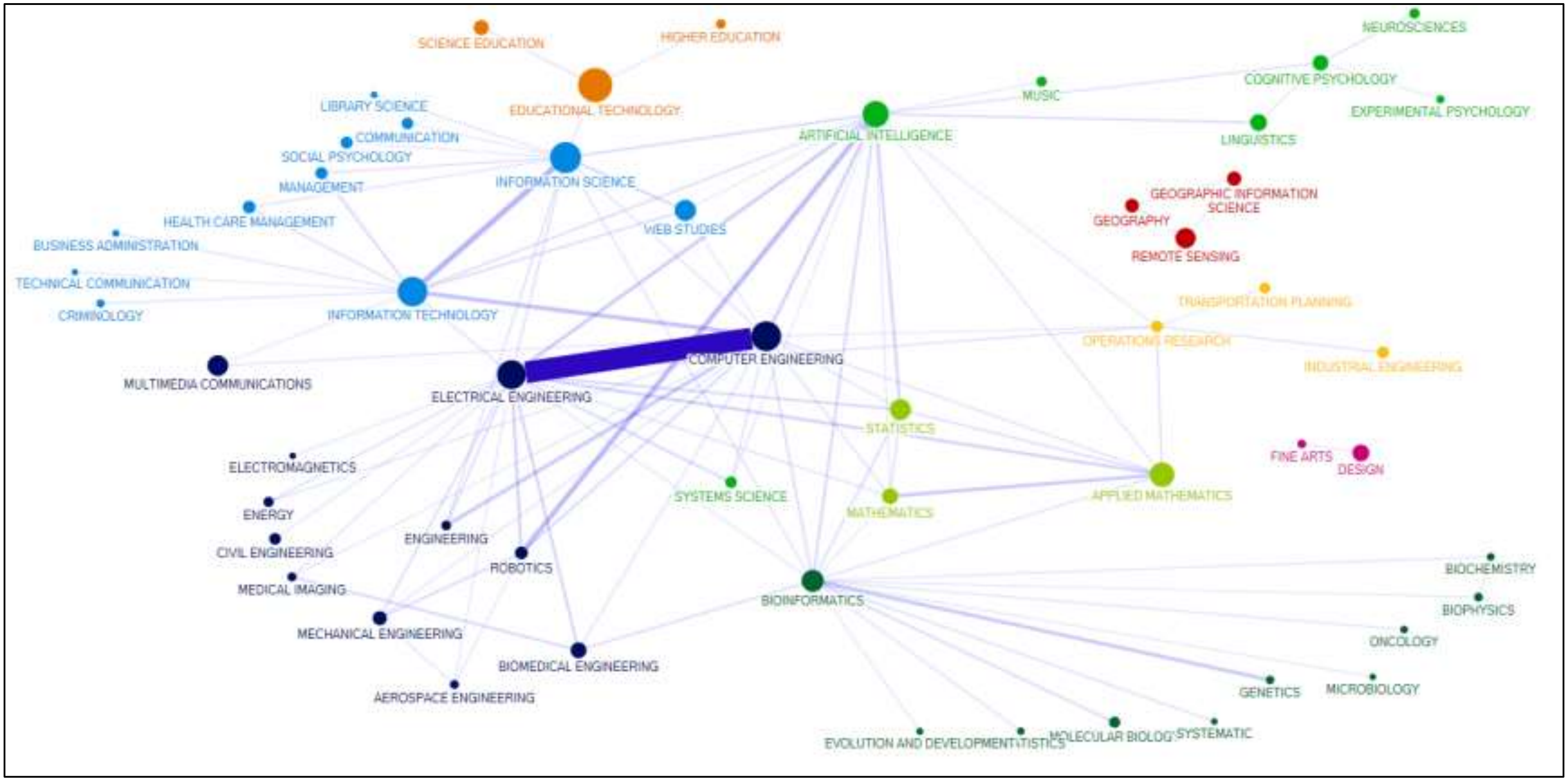


Figure 4-8: CS, Any, Weights >= 10

4.2.14 **CSP Graph**

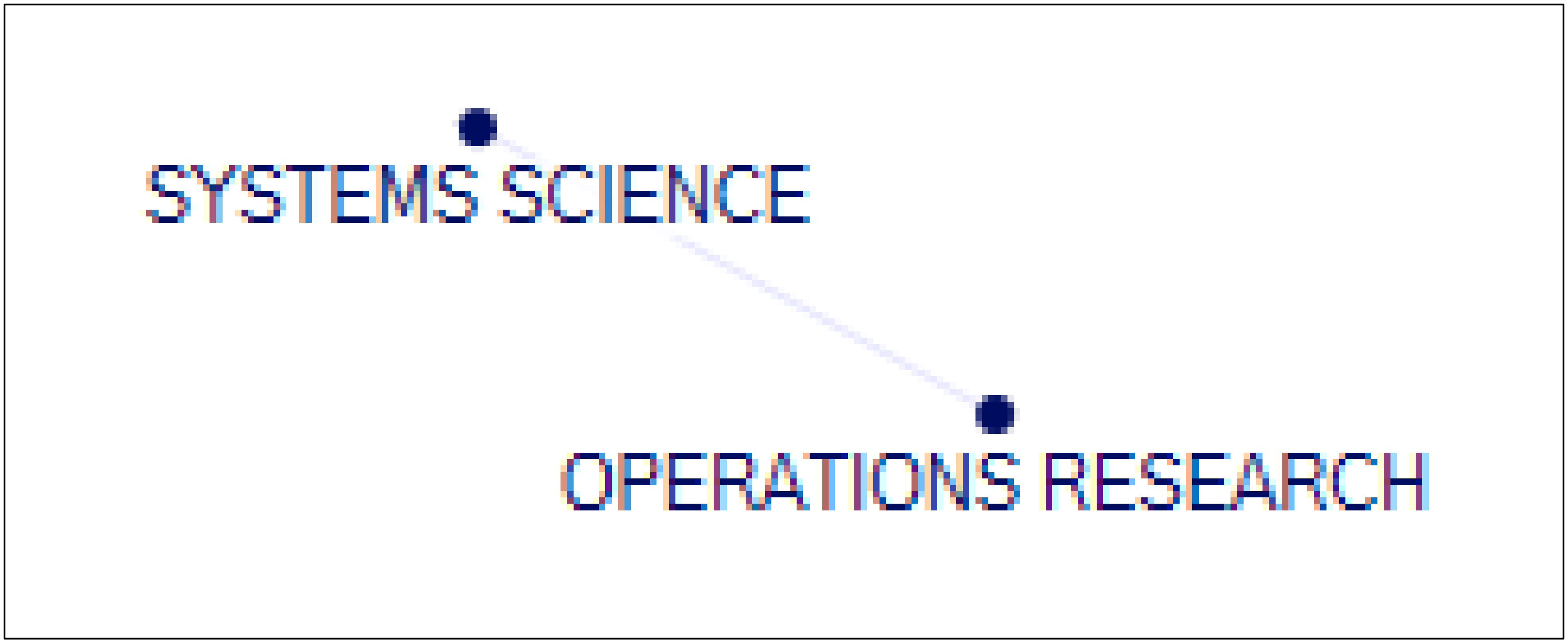


Figure 4-9: CS, Primary, Weights > 0

4.2.15 CSS Graph

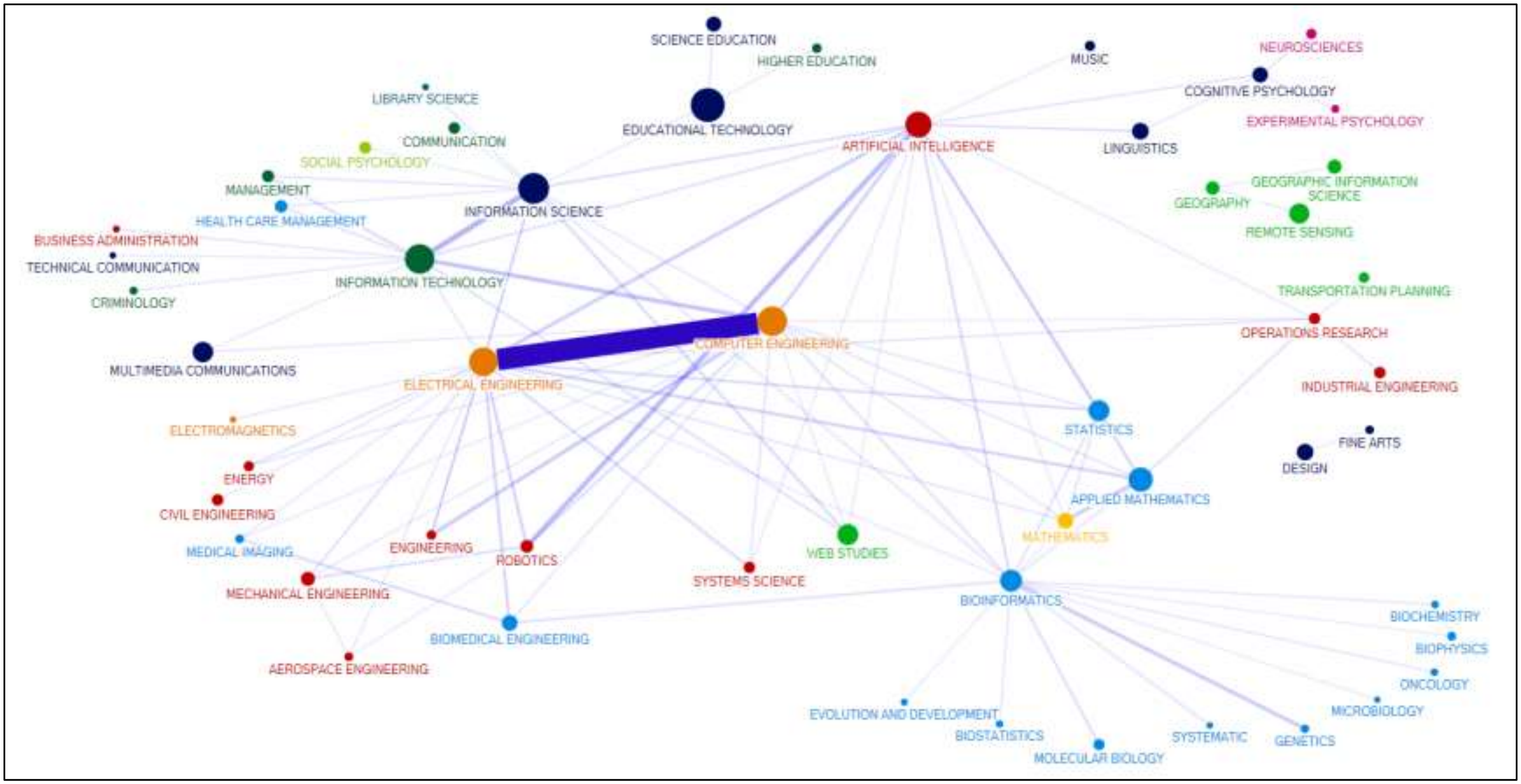


Figure 4-10: CS, Secondary, Weights >= 10

4.2.16 Classification Graphs for CE

Below are the summary metrics for CE in its three respective positions.

Table 4-12: Summary Graph Metrics for CE

	CEA	CEP	CES
Groups	19	11	15
Number of vertices	172	109	101
Distinct edges	494	314	204
Total edges	2,194	1,826	368
Edges of weight 1	324	189	158
Distinct edges of weight > 1	170	125	46
Edges of weight > 1	1,870	1,637	210

324/1870 the ratio of single to >1 edges is 17% for CEA.

189/1637 the ratio of single to >1 edges is 11% for CEP.

158/210 the ratio of single to >1 edges is 75% for CES.

494/2194 the ratio of distinct edges to total edges is 23% for CEA

314/1826 the ratio of distinct edges to total edges is 17% for CEP

204/368 the ratio of distinct edges to total edges is 55% for CES

4.2.17 CEA Findings

The CEA graph in Figure 4-11 displays a very strong relationship between CS and EE, corroborating information in Table 4-8 regarding these two as having the most co-Classifications overall, but more importantly, showing how often they appear together and with other Classifications. Although they are in separate groups the graph shows they are also usually co-Classifications with most of the other major bridge spanners in the graph. The list of other bridge spanners in the CEA graph include: CS, EE, ISci, IT, MECHANICAL ENGINEERING, ARTIFICIAL INTELLIGENCE, and BIOMEDICAL ENGINEERING. These each become the most common co-Classifications with other areas that classify with CE.

4.2.18 **CEP Findings**

The same general description for the CEA graph in Figure 4-11 applies to the CEP graph in Figure 4-12. This is understandable as CE consists predominantly of Primary theses and it is the Primary theses that make the most connections to other Classifications. This is the opposite of CS, where CSA and CSS are more aligned.

CEP has understandably fewer groups than CEA but shares most of the same common bridge nodes and the relationships to those nodes. However, APPLIED MATHEMATICS, MATHEMATICS, and STATISTICS, found in the CEA graph, are not found in the CEP graph.

4.2.19 **CES Findings**

Note the absences from the CEP graph of APPLIED MATHEMATICS, MATHEMATICS, and STATISTICS. These now appear in the CES graph in Figure 4-13. Notable is that CS and EE seem to be less closely associated when CE is a Secondary. These findings indicate that CS will show up in a CES thesis with STATISTICS, MATHEMATICS, or APPLIED MATHEMATICS more often than it will with EE. EE, on the other hand will appear with APPLIED MATHEMATICS and/or STATISICS more often than with CS.

4.2.20 **CEA Graph**

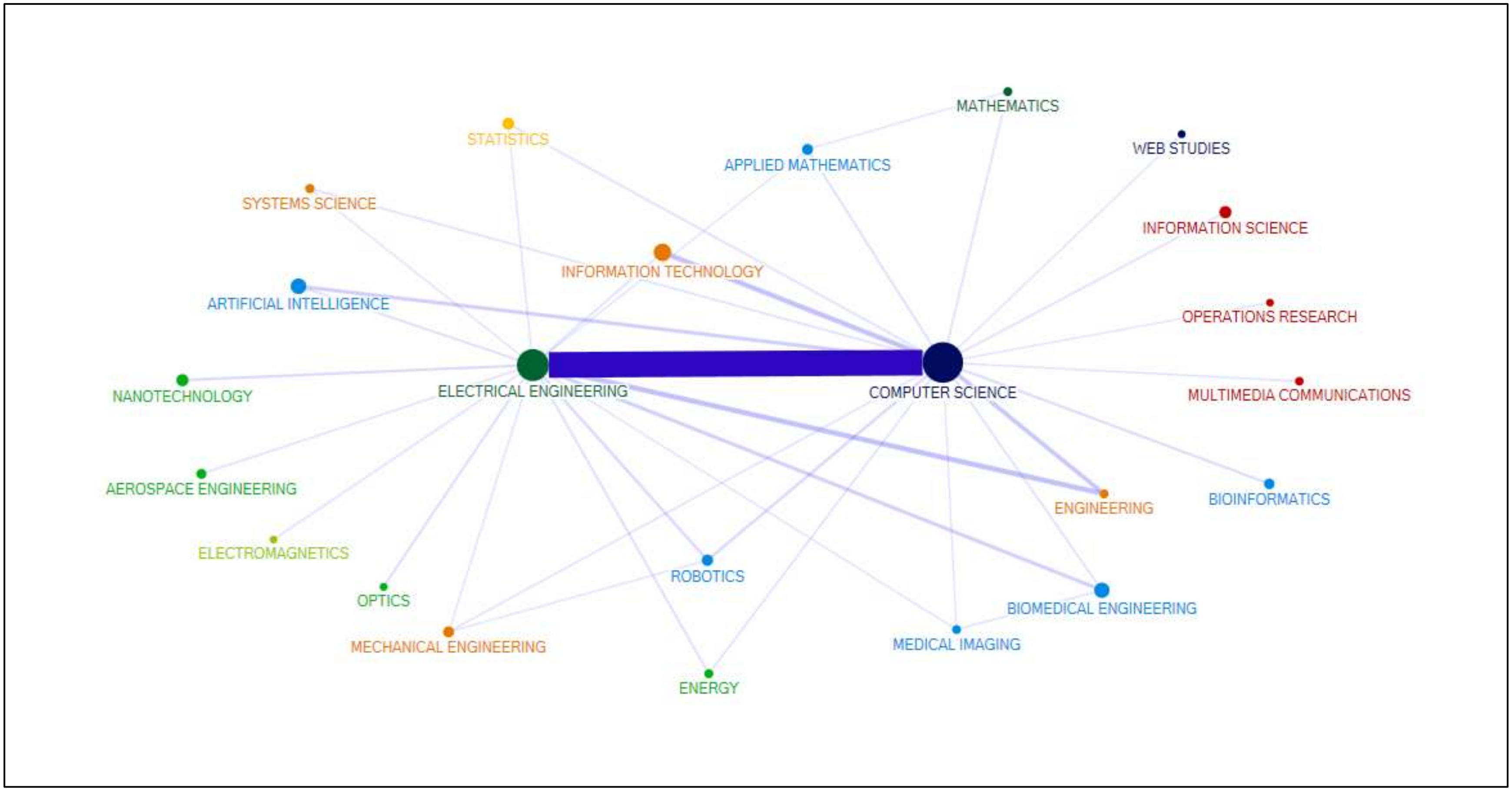


Figure 4-11: CE, Any, Weights >= 10

4.2.21 **CEP Graph**

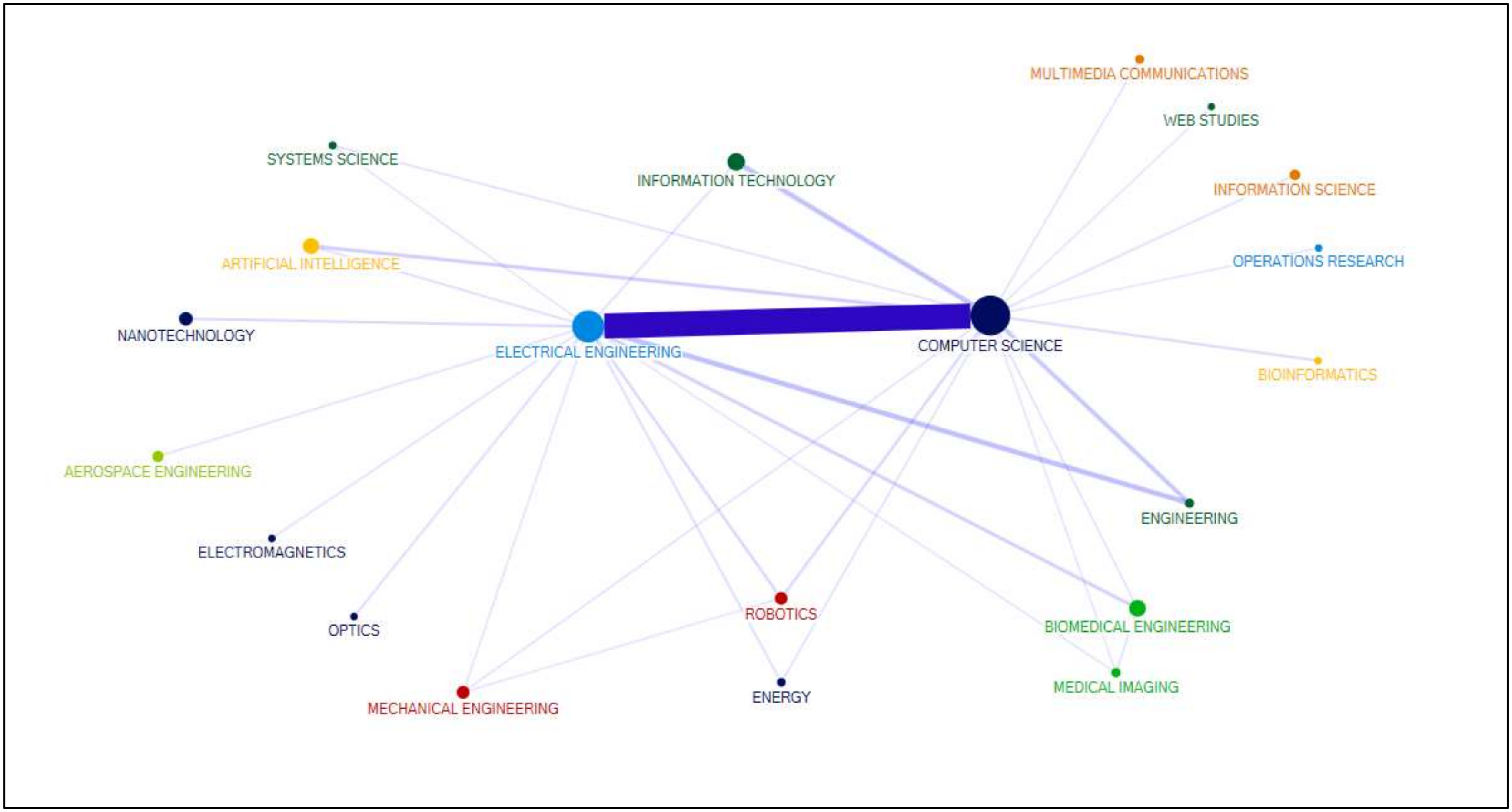


Figure 4-12: CE, Primary, Weights >= 10

4.2.22 CES Graph

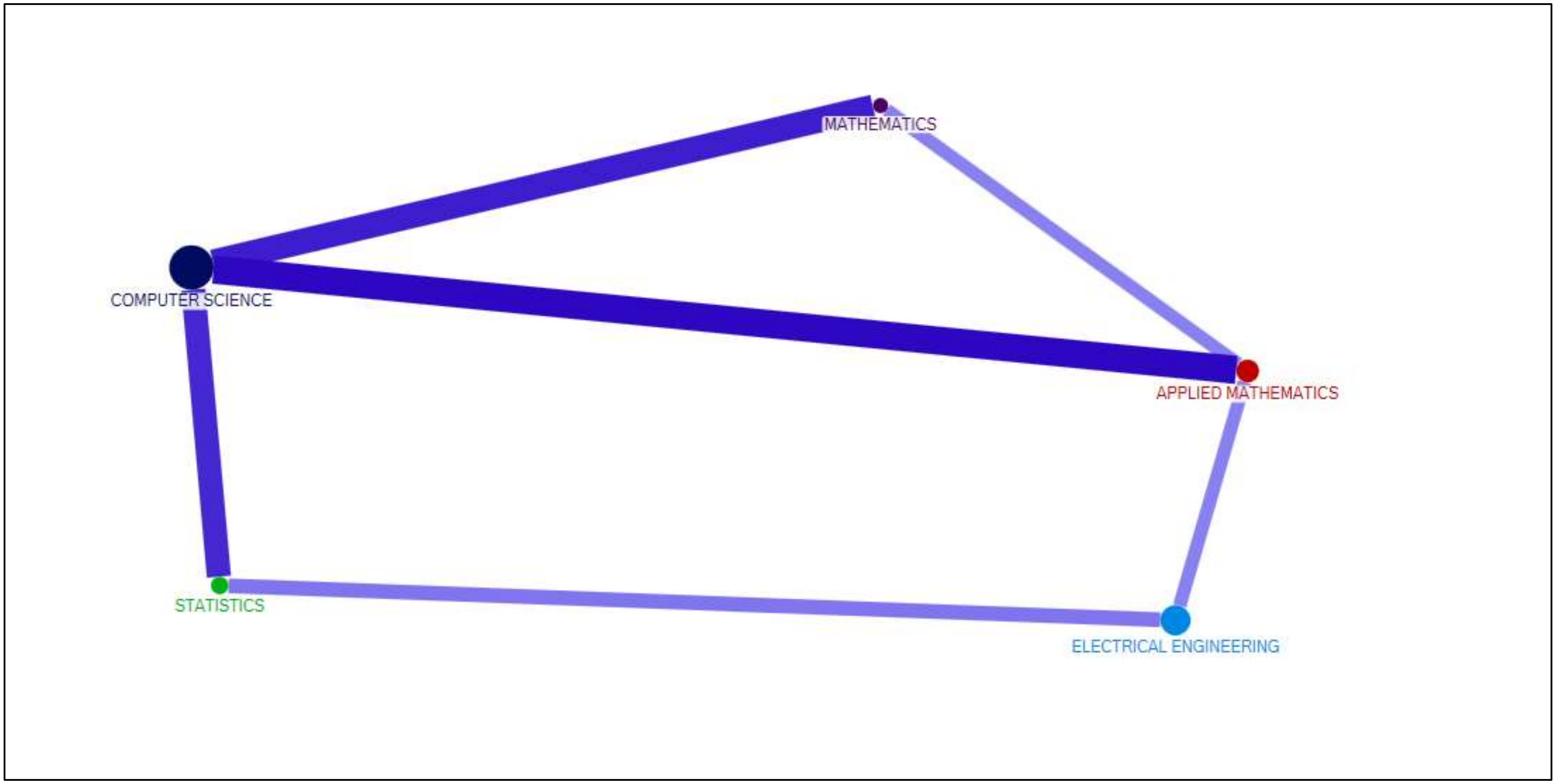


Figure 4-13: CE, Secondary, Weights >= 10

4.2.23 Classification Graphs for IT

Below are the summary metrics for IT in its three respective positions.

Table 4-13: Summary Graph Metrics for IT

	ITA	ITP	ITS
Groups	19	13	16
Number of vertices	250	126	221
Distinct edges	1,345	453	1,026
Total edges	3,503	1,169	2,334
Edges of weight 1	864	296	697
Distinct edges of weight > 1	481	157	329
Edges of weight > 1	2,639	873	1,637

864/2639 the ratio of single to >1 edges is 33% for ITA.

296/873 the ratio of single to >1 edges is 34% for ITP.

697/1637 the ratio of single to >1 edges is 43% for ITS.

1345/3503 the ratio of distinct edges to total edges is 38% for ITA

453/1169 the ratio of distinct edges to total edges is 39% for ITP

1026/2334 the ratio of distinct edges to total edges is 44% for ITS

4.2.24 ITA Findings

The ITA graph in Figure 4-14 exhibits a strong presence of MANAGEMENT and ORGANIZATIONAL BEHAVIOR. CS, ISci, and CE are also heavily connected, but especially CS and ISci. Regarding the lack of INFORMATION SYSTEMS, the strong presence of MANAGEMENT and its association with ORGANIZATIONAL BEHAVIOR shows these two fields, normally associated with ISys according to the ACM description of ISys, now use IT for to classify their graduate work

Education via EDUCATIONAL TECHNOLOGY stands out as a very prominent group of connections.

4.2.25 **ITP Findings**

The ITP graph in Figure 4-15 reveals the elimination of the MANAGEMENT node, while retaining the strong presence of CS and ISci with their dense set of connections. ITP also shows a weakening of the connection between CS and CE, which shows IT as the Primary theses is not as likely to claim CE as part of its Classifications.

Education is still prominent.

4.2.26 **ITS Findings**

The ITS graph in Figure 4-16 appears very different from the ITP graph with MANAGEMENT taking center stage. MANAGEMENT primarily touches ISci, CS, and ORGANIZATIONAL BEHAVIOR but also has strong links to almost every other disciplines shown in the graph.

However, note the lack of connections between MANAGEMENT and CE or ELECTRICAL ENGINEERING.

The heavy connection between CS and CE returns in the ITS graph.

An educational group is still present but is no longer visibly connected to any other group as its previous connection to ISci, as found in ITP, is no longer present in ITS. There may still be edges connecting ISci with the mainstay of the education group, EDUCATIONAL TECHNOLOGY, but the number of edges no longer meets the minimum requirement of 10 instances.

4.2.27 ITA Graph

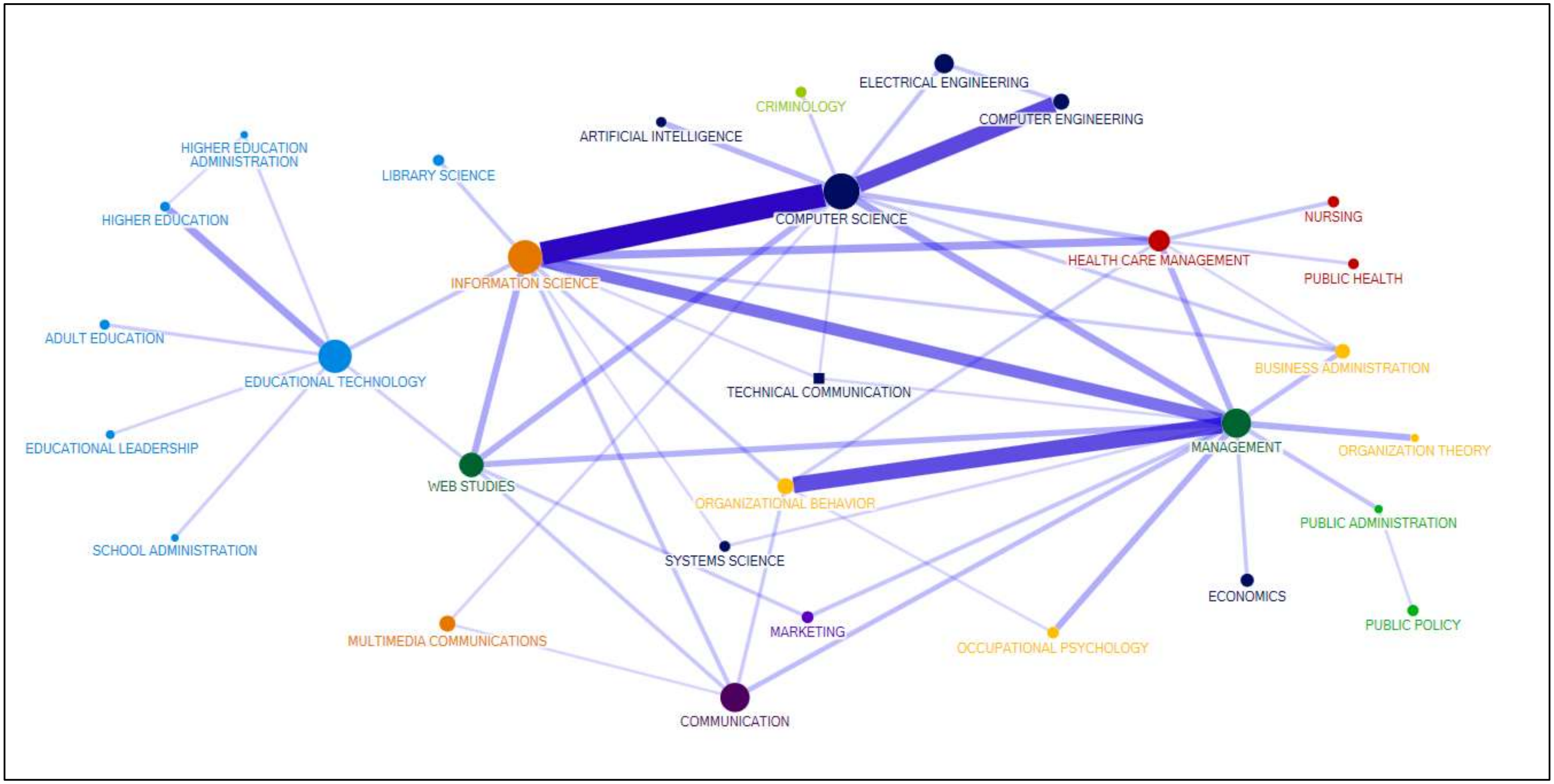


Figure 4-14: IT, Any, Weights >= 10

4.2.28 **ITP Graph**

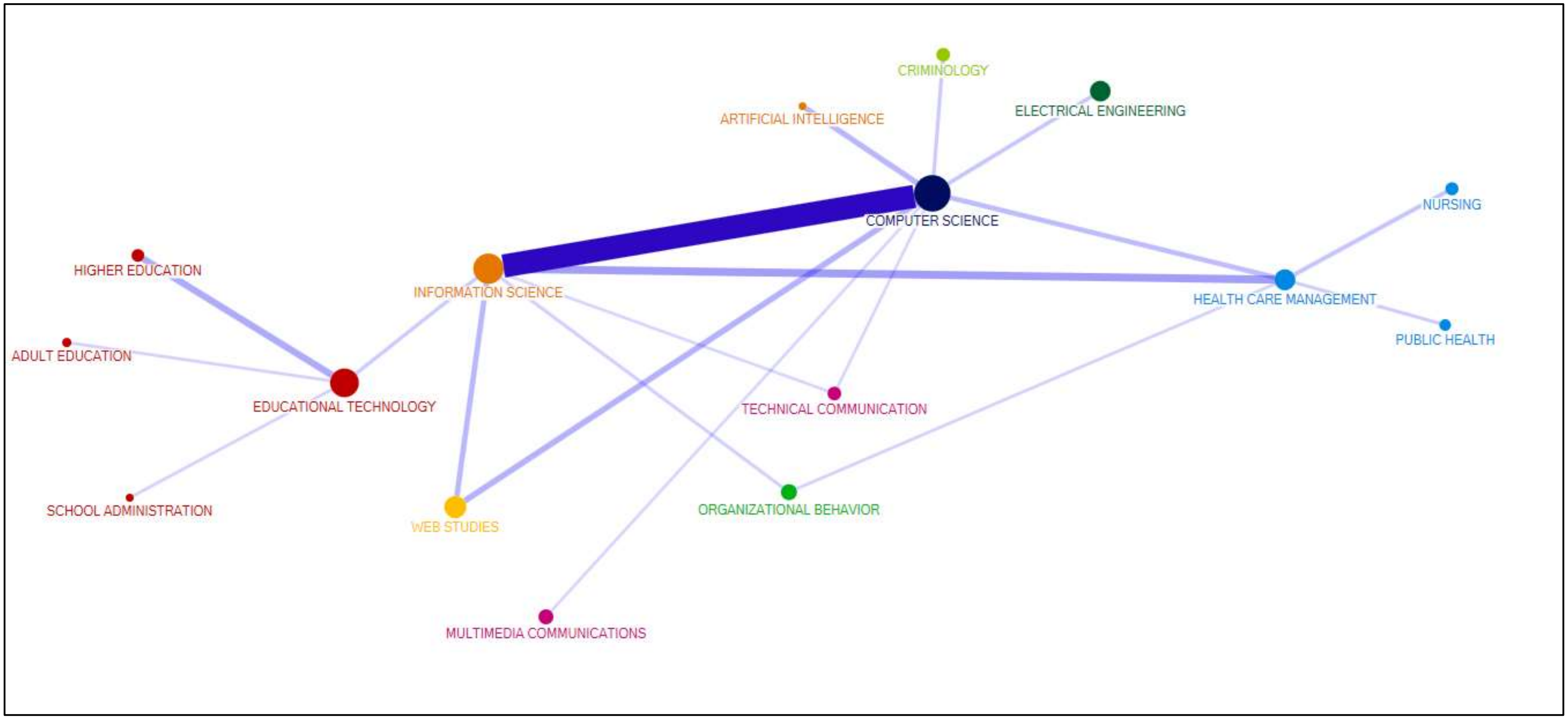


Figure 4-15: IT, Primary, Weights >= 10

4.2.29 ITS Graph

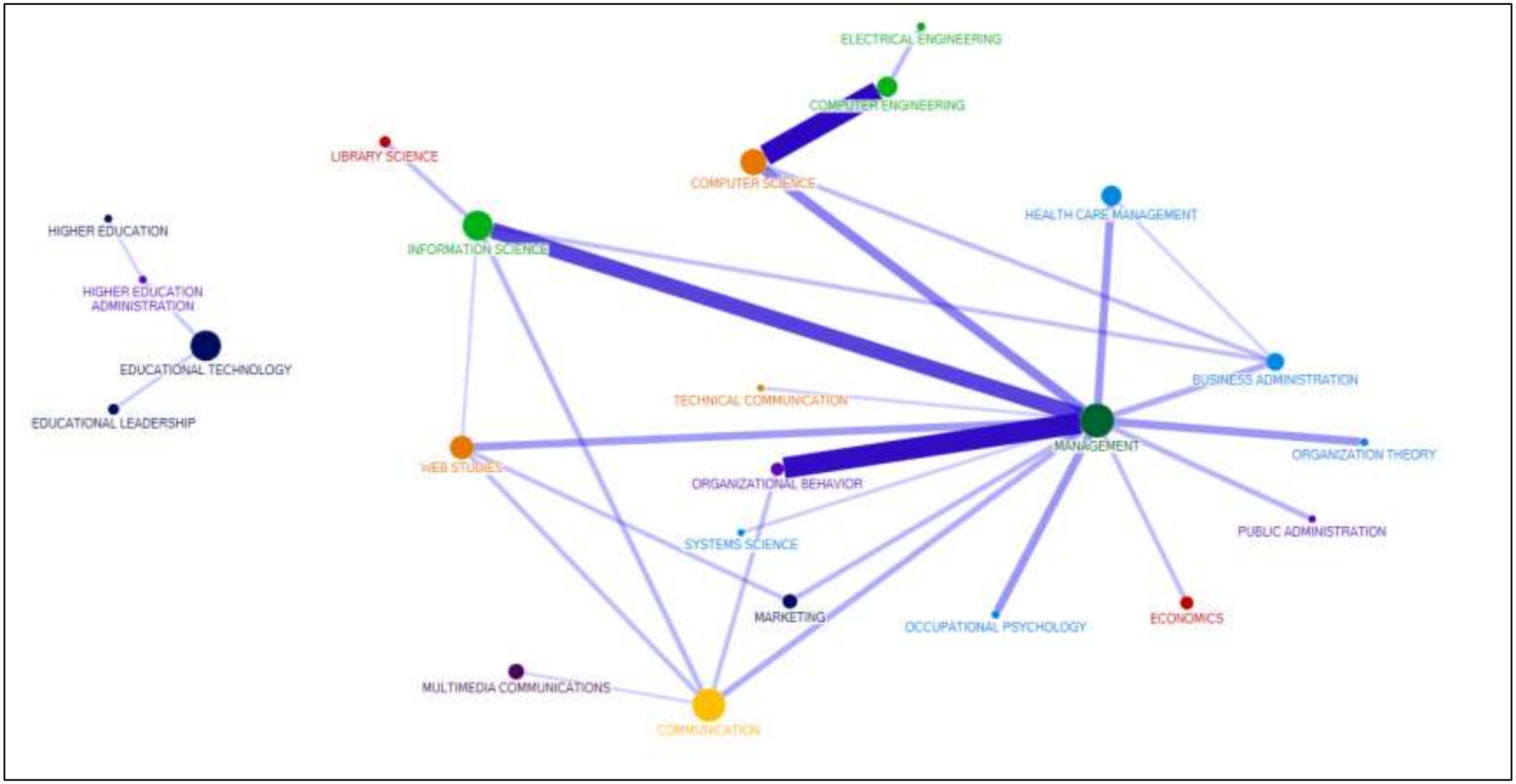


Figure 4-16: IT, Secondary, Weights >= 10

4.2.30 **Classification Graphs for ISci**

Below are the summary metrics for ISci in its three respective positions.

Table 4-14: Summary Graph Metrics for ISci

	ISciA	ISciP	ISciS
Groups	17	4	16
Number of vertices	262	18	262
Distinct edges	1,375	21	1,374
Total edges	3,172	69	3,082
Edges of weight 1	961	12	965
Distinct edges of weight > 1	414	9	409
Edges of weight > 1	2,166	57	2,117

961/2166 the ratio of single to >1 edges is 44% for ISciA.

12/57 the ratio of single to >1 edges is 21% for ISciP.

965/2117 the ratio of single to >1 edges is 46% for ISciS.

1375/3172 the ratio of distinct edges to total edges is 43% for ISciA

21/69 the ratio of distinct edges to total edges is 30% for ISciP

1374/3082 the ratio of distinct edges to total edges is 45% for ISciS

4.2.31 **ISA Findings**

The overall ISA graph is heavily influenced by the high proportion of ISS theses and the narrower focus of ISP theses. However, taken as a whole, the strongest visible connection is between IT and CS. A strong connection between IT and MANAGEMENT is also visible as is the strong presence of WEB STUDIES connecting with IT, CS, LIBRARY SCIENCE, COMMUNICATIONS, and MASS COMMUNICATIONS.

4.2.32 **ISP Findings**

Only 21 distinct edges of various weights exist with ISci as Primary, all of which are shown in Figure 4-18. This contrasts to the 1,375 distinct edges present in ISA and the 1374 in ISS. The overall ProQuest counts from Table 4-1 attest to the low numbers of ISP theses. When it is present, CS and ARTIFICIAL INTELLIGENCE make the strongest connection and thus are likely present *together* in most ISP theses. HEALTH CARE MANAGEMENT's presence is the also relatively strong.

4.2.33 **ISS Findings**

The ISS graph in Figure 4-19 brings to light the major tendencies of co-Classifications of ISci as Secondary. These tendencies are:

- IT and CS
- IT and MANAGEMENT
- IT and LIBRARY SCIENCE
- IT with HEALTH CARE MANAGEMENT.
- LIBRARY SCIENCE and WEB STUDIES
- CS and WEB STUDIES
- WEB STUDIES with each of:
    - COMMUNICATIONS
    - MASS COMMUNICATIONS

4.2.34 ISA Graph

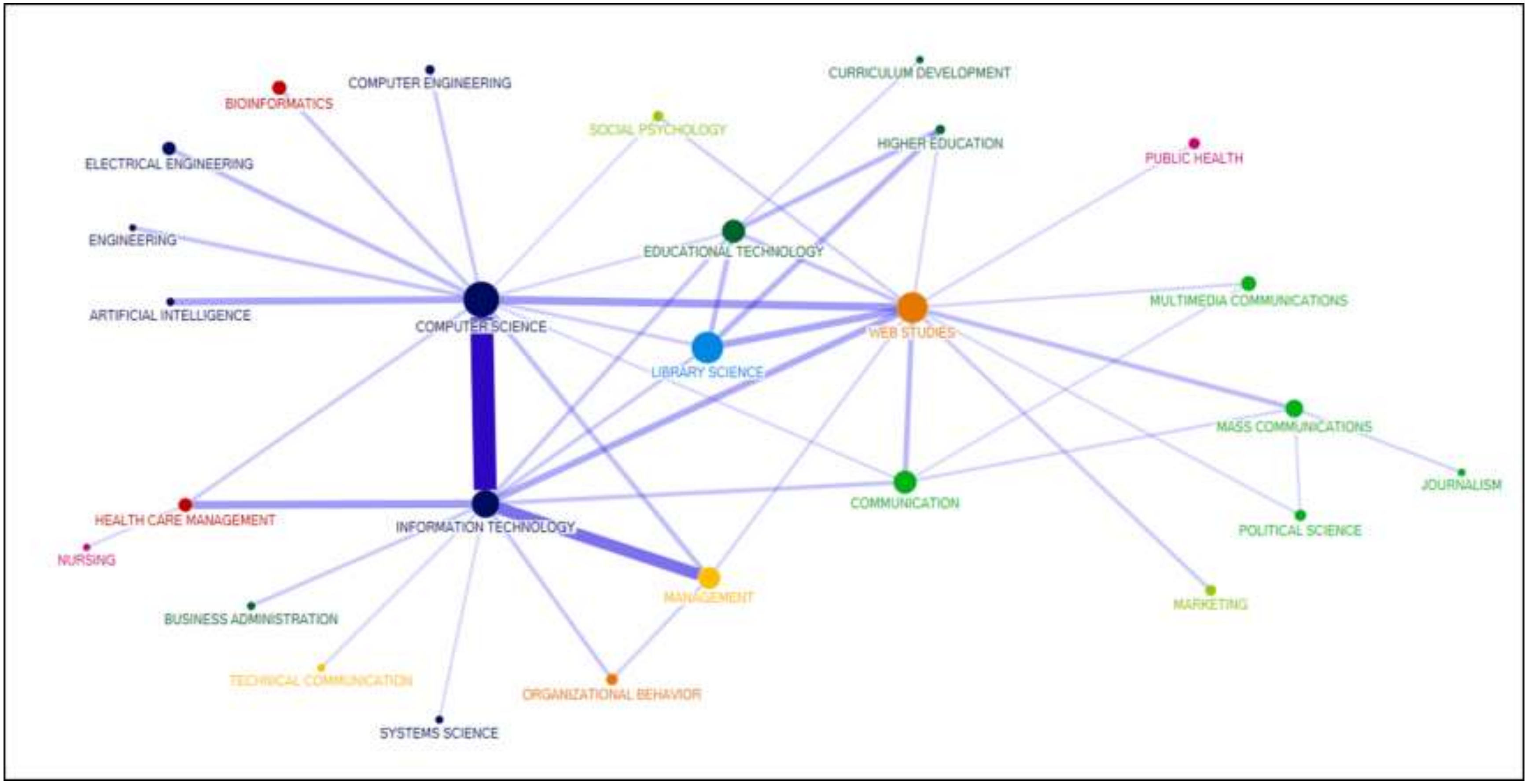


Figure 4-17: ISci, Any, Weights >= 10

4.2.35 **ISP Graph**

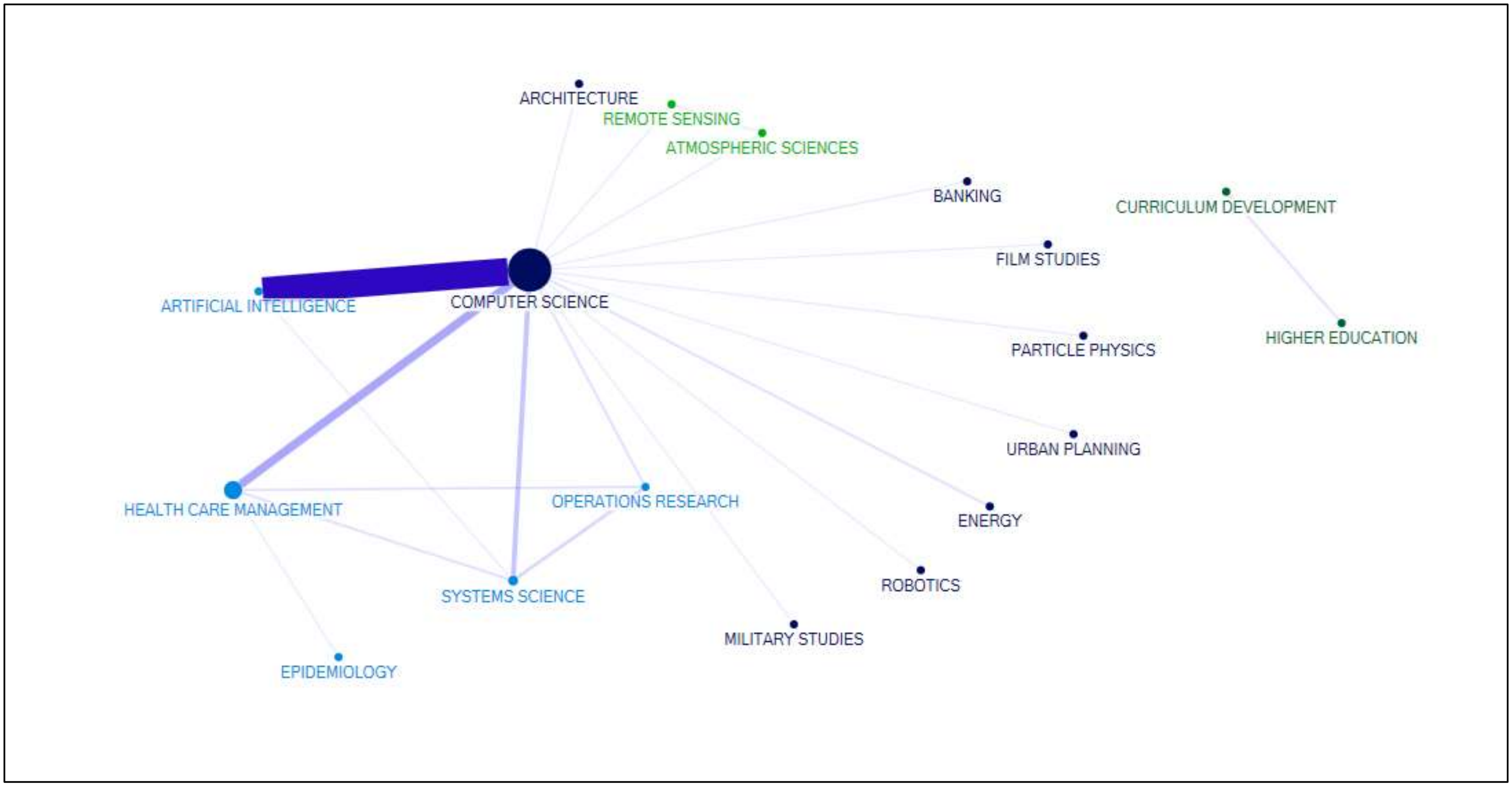


Figure 4-18: ISci, Primary, Weights > 0

4.2.36 **ISS Graph**

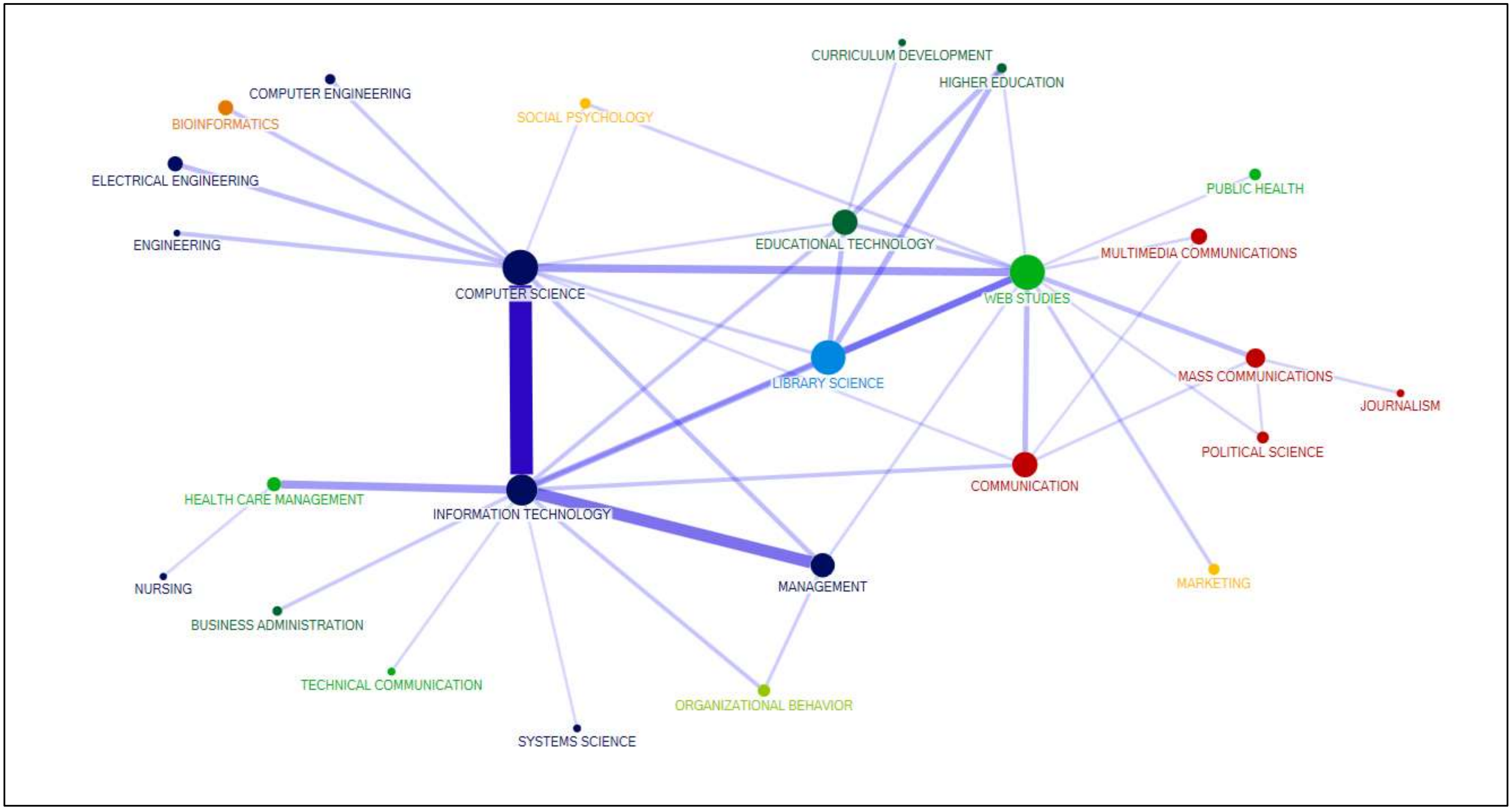


Figure 4-19: ISci, Secondary, Weights >= 10

4.3 Departments

The university departments that awarded the degrees for the theses add context to the origin of computing and computing-related research from within colleges and universities. On rare occasions the Department field is not used correctly by the people submitting theses to the ProQuest database: EG sometimes this field contained an email address (bsheaff@uvic.ca), apparent initials (hkust), and degree titles (Computer Science - Ph.D, Information Technology (PhD)). The first two are all from the University of Victoria (Canada) and Hong Kong University of Science and Technology, respectively.

Departments are included in 64% to 80% of the downloaded ProQuest theses extracts from 2009-2014. Departments are much rarer over the whole dataset, occurring only 21% of the time.

Table 4-15 shows the number of theses with the Department field.

Table 4-15: Theses with Departments (2009-2014)

Classification	Count All	Count with Dept.	% with Dept
CS	21,628	12,963	60%
CE	5,538	3,979	72%
IT	3,748	2,986	80%
ISci	2,792	1,774	64%

Table 4-16 shows the number of distinct departments present for each computing discipline for theses from 2009 to 2014.

CSS has almost 500 more departments than CSP. ITP and ITS have virtually equal numbers. CEP actually has three times more departments than CES. ISS has four times more than ISP. Note that these are not normalized departments but are exact text matches.

- CSP and CSS share 155 Departments.
- CEP and CES share 53 Departments.
- ITP and ITS share 163 Departments.
- ISP and ISS share 85 Departments.

Table 4-16: DISTINCT Departments (2009-2014)

	Any	Primary	Secondary	Solo
CS	770	221	705	219
CE	313	276	91	83
IT	599	382	381	76
ISci	422	107	401	51

The tables on the following pages illustrate the presence and variety of departments which produce computing and computing related theses. They are organized in the same manner as previous tables and show the top 40 departments for each respective position of the computing discipline in the ProQuest Classification field.

A blank row represents the absence of the Department field.

4.3.1 **CS**

CS Primary and CS Solo are almost mirror images. Computer Science is the top Department, not including blanks, across the board, including for CSS.

CS Secondary is actually technology heavy. CSS differs by having a stronger presence of Electrical and Computer Engineering. Leaving the top six, however, shows CSS coming from Departments such as Mechanical Engineering, Mathematics, and Linguistics.

Overall, most CS theses appear to come from computing related departments.

Table 4-17 Top 40 Departments, CS (2009-2014)

Any		Primary		Secondary		Solo	
	8,665		5,168		3,497		5,159
Computer Science	5,804	Computer Science	3,622	Computer Science	2,182	Computer Science	3,618
Computer Science and Engineering	803	Computer Science and Engineering	452	Electrical and Computer Engineering	482	Computer Science and Engineering	451
Electrical and Computer Engineering	533	Computer Sciences	199	Electrical Engineering	406	Computer Sciences	199
Electrical Engineering	416	Computing Science	117	Computer Science and Engineering	351	Computing Science	117
Computer Sciences	274	Computer Science & Engineering	106	Computer Engineering	137	Computer Science & Engineering	106
Computer Science & Engineering	187	Electrical Engineering and Computer Science	79	Mechanical Engineering	103	Electrical Engineering and Computer Science	79
Electrical Engineering and Computer Science	163	Computing	77	Electrical Engineering and Computer Science	84	Computing	76
Computing Science	145	hkust	75	Electrical & Computer Engineering	83	hkust	75
Computer Engineering	144	School of Computing	67	Computer Science & Engineering	81	School of Computing	67
Computing	121	Computer and Information Science	66	Mathematics	80	Computer and Information Science	66
Graduate School - New Brunswick	119	Graduate School - New Brunswick	54	Computer Sciences	75	Graduate School - New Brunswick	54
School of Computing	112	Electrical and Computer Engineering	51	Graduate School - New Brunswick	65	Electrical and Computer Engineering	51
Mechanical Engineering	106	School of Computer Science	49	Information Technology	59	School of Computer Science	49
Computer and Information Science	99	Computer Science 0201	48	Engineering	54	Computer Science 0201	48
Electrical & Computer Engineering	95	Computer Engineering Department	37	Cybersecurity	51	Computer Engineering Department	37
hkust	82	Department of Computing Science	30	Electrical Engineering & Computer Sciences	45	Department of Computing Science	30
Mathematics	82	Electrical Engineering & Computer Sciences	29	School of Computing	45	Electrical Engineering & Computer Sciences	29
Computer Science 0201	74	Information and Computer Science - Ph.D	29	Computing	44	Information and Computer Science - Ph.D	29
Electrical Engineering & Computer Sciences	74	Computer Science - Ph.D	28	Linguistics	43	Computer Science - Ph.D	28
Information Technology	71	Department of Computer and Information Sciences	28	School of Business and Technology	37	Department of Computer and Information Sciences	28
School of Computer Science	71	Electrical Engineering & Computer Science	27	Electrical Engineering & Computer Science	35	Electrical Engineering & Computer Science	27
Engineering	65	Computer Science - M.S	25	Information Systems	34	Computer Science - M.S	25
Cybersecurity	64	Information Technology (PhD)	22	Statistics	34	Information Technology (PhD)	22
Electrical Engineering & Computer Science	62	Computer and Information Sciences	20	Computer and Information Science	33	Computer and Information Sciences	20
Information and Computer Science - Ph.D	61	Department of Computer Science	18	Information and Computer Science - Ph.D	32	Department of Computer Science	18
Computer Science - Ph.D	55	Engineering and Applied Sciences	18	Technology	32	Engineering and Applied Sciences	18

Table 4-17 Continued

Any		Primary		Secondary		Solo	
Information Technology (PhD)	51	Hajim School of Engineering and Applied Sciences	18	Electrical Engineering and Computer Sciences	30	Hajim School of Engineering and Applied Sciences	18
Computer Engineering Department	45	Computer Science (CISC, CISD)	17	Industrial Engineering	30	Computer Science (CISC, CISD)	17
Electrical Engineering and Computer Sciences	45	Department of Computer and Information Science	16	Information Technology (PhD)	29	Department of Computer and Information Science	16
Linguistics	44	Electrical Engineering and Computer Sciences	15	Computer Information Systems (MCIS, DCIS)	28	Electrical Engineering and Computer Sciences	15
Computer Information Systems (MCIS, DCIS)	42	Computer Information Systems (MCIS, DCIS)	14	Computing Science	28	Computer Information Systems (MCIS, DCIS)	14
Computer Science - M.S	42	Cybersecurity	13	Physics	28	Cybersecurity	13
Technology	42	bsheaff@uvic.ca	12	Civil Engineering	27	bsheaff@uvic.ca	12
Department of Computer and Information Sciences	40	Computational Science	12	Computer Science - Ph.D	27	Computational Science	12
School of Business and Technology	37	Computing and Information Sciences	12	Information Systems (DISS)	27	Computing and Information Sciences	12
Computational Science	35	Dept. of Computer Science and Engineering	12	Networking and Systems Administration	27	Dept. of Computer Science and Engineering	12
Information Systems	34	Electrical & Computer Engineering	12	Applied Mathematics	26	Electrical & Computer Engineering	12
Statistics	34	Information Technology	12	Computer Science 0201	26	Information Technology	12
Department of Computing Science	33	College of Arts and Sciences / Department of Computer Science	11	Electrical & Computer Eng	24	College of Arts and Sciences / Department of Computer Science	11

4.3.2 CE

For CE the top five Departments are identical for CEP and CES except for Electrical Engineering and Computer Science switching the 2, 4 and 4, 2 positions in CEP and CES, respectively.

The Graduate School – New Brunswick makes a strong appearance in CEP.

CES shows the presence of Mathematics, Music, Biomedical Engineering and Civil Engineering in its top 11 (not including blanks). Further down CES also includes Business Administration, Cognitive Science, and Computational Analysis and Modeling.

Table 4-18 Top 40 Departments, CE (2009-2014)

Any		Primary		Secondary		Solo	
	1,559		1,477		82		395
Electrical and Computer Engineering	1,031	Electrical and Computer Engineering	993	Electrical and Computer Engineering	38	Electrical and Computer Engineering	294
Electrical Engineering	505	Electrical Engineering	489	Computer Science	22	Computer Engineering	168
Computer Engineering	390	Computer Engineering	370	Computer Engineering	20	Electrical & Computer Engineering	87
Computer Science	377	Computer Science	355	ELECTRICAL ENGINEERING	16	Electrical Engineering	76
Computer Science and Engineering	225	Computer Science and Engineering	217	Computer Science and Engineering	8	Computer Science and Engineering	44
Electrical & Computer Engineering	214	Electrical & Computer Engineering	207	Electrical & Computer Engineering	7	Computer Science & Engineering	18
Engineering	76	Engineering	73	Mathematics	7	Department of Electrical and Computer Engineering	15
Computer Science & Engineering	55	Computer Science & Engineering	52	Music	4	Engineering	15
Graduate School - New Brunswick	53	Graduate School - New Brunswick	50	Biomedical Engineering	3	Computer Engineering Department	14
Electrical Engineering and Computer Science	40	Electrical Engineering and Computer Science	40	Civil Engineering	3	Electrical and Computer	12
Department of Electrical and Computer Engineering	38	Department of Electrical and Computer Engineering	38	Computer Science & Engineering	3	Electrical and Computer Engineering - Ph.D	11
Mechanical Engineering	37	Mechanical Engineering	36	Engineering	3	Electrical Engineering and Computer Science	11
Electrical and Computer Engineering - Ph.D	34	Electrical and Computer Engineering - Ph.D	34	Graduate School - New Brunswick	3	Engineering, Computer	10
Electrical & Computer Eng	24	Electrical & Computer Eng	23	School of Business and Technology	3	Computer Science	9
Electrical and Computer	23	Electrical and Computer	23	School of Mathematical Sciences	3	Electrical and Computer Engineering - M.S	8
Computer Engineering Department	20	Computer Engineering Department	20	Applied Mathematics and Statistics	2	Electrical Engineering & Computer Science	6
Computer Sciences	18	Engineering, Computer	17	Applied Physics	2	Computer Eng	5
Engineering, Computer	18	Computer Sciences	16	Computational Engineering	2	Graduate School - New Brunswick	5
Computing	16	Computing	16	Computer Engineering (Robotics and Control)	2	Electrical and Computer Eng	4
Electrical and Computer Engineering - M.S	16	Electrical Engineering & Computer Science	16	Computer Sciences	2	Electrical and Computer Engineering (Engineering)	4
Electrical Engineering & Computer Science	16	Electrical and Computer Engineering - M.S	15	Industrial Engineering	2	Computer Engineering (College of Engineering)	3
Information Technology	16	Information Technology	15	Physics	2	Computing Science	3
Engineering Science	14	Engineering Science	14	Anthropology	1	Electrical Engineering (Computer Engineering)	3
Electrical and Computer Engineering (Engineering)	12	Electrical Engineering (Computer Engineering)	12	Applied Management and Decision Sciences	1	Engineering and Engineering Technology	3

Table 4-18 Continued

Any		Primary		Secondary		Solo	
Electrical Engineering (Computer Engineering)	12	Electrical and Computer Eng	11	Applied Mathematics	1	Industrial and Systems Engineering	3
Industrial Engineering	12	Electrical and Computer Engineering (Engineering)	11	Bioengineering	1	Color Science	2
Technology	12	School of Electrical and Computer Engineering	11	Business Administration	1	Computer Engineering and Computer Science	2
Civil Engineering	11	Science informatique et génie électrique / Electrical Engineering and Computer Science	11	Chemical and Biological Engineering	1	Dept. of Electronics and Elec. Comm. Engineering	2
Computer Science 0201	11	Technology	11	Civil and Environmental Engineering	1	Electrical & Computer Eng	2
Electrical and Computer Eng	11	Computer Eng	10	Civil Engineering(Structural Engineering)	1	Electrical Engineering(VLSI Design)	2
Electrical Engineering & Computer Sciences	11	Computer Science 0201	10	Cognitive Science	1	Mechanical Engineering	2
School of Electrical and Computer Engineering	11	Electrical Engineering & Computer Sciences	10	Communication Sciences and Disorders	1	Reliability Engineering Centre	2
Science informatique et génie électrique / Electrical Engineering and Computer Science	11	Engineering : Electrical & Computer	10	Computational Analysis and Modeling	1	Science informatique et gÃ©nie Ã©lectrique / Electrical Engineering and Computer Science	2
Computer Eng	10	Industrial Engineering	10	Computational Science and Engineering	1	Aerospace Engineering	1
Engineering : Electrical & Computer	10	Aerospace Engineering	9	Computer and Information Science	1	Art and Design	1
Aerospace Engineering	9	Computer Engineering and Computer Science	9	Computer and Information Sciences and Engineering	1	bsheaff@uvic.ca	1
Computer Engineering and Computer Science	9	School of Computing	9	Computer Science - Ph.D	1	Chemical Engineering	1
School of Computing	9	Civil Engineering	8	Computer Science 0201	1	Civil and Environmental Engineering	1
Biomedical Engineering	8	Electrical Eng	8	Computing and Information Sciences	1	College of Engineering and Applied Science - Security	1

4.3.3 **IT**

IT has the expected variety of Departments coinciding with previous findings on co-Classifications above. IT Primary has a strong showing of technology related programs in its top six, while IT Secondary is predominantly Business and Management in its top six.

Table 4-19: Top 40 Departments, IT (2009-2014)

Any		Primary		Secondary		Solo	
	762		371		391	School of Business and Technology	40
School of Business and Technology	229	Computer Science	165	School of Business and Technology	121		38
Computer Science	188	School of Business and Technology	108	Applied Management and Decision Sciences	69	Networking and Systems Administration	15
Applied Management and Decision Sciences	99	Information Technology	59	Management	45	Information Technology	14
Information Technology	84	Information Systems (DISS)	52	School of Business	45	Information Systems (DISS)	13
Information Systems	76	Cybersecurity	49	Business	34	Information Technology Leadership (formerly Computer Information Systems)	12
Information Systems (DISS)	72	Information Systems	45	Business Administration	32	Cybersecurity	11
School of Business	63	Networking and Systems Administration	36	Information Systems	31	Applied Management and Decision Sciences	10
Cybersecurity	61	Applied Management and Decision Sciences	30	Communication	28	Computer and Information Technology	9
Management	57	Electrical Engineering	28	Information Technology	25	Information Systems	9
Business	46	Technology	28	Computer Science	23	Technology	9
Business Administration	42	Computer and Information Technology	27	Electrical and Computer Engineering	20	Computer Science	7
Networking and Systems Administration	42	Nursing	24	Information Systems (DISS)	20	Interactive Arts and Technology	6
Electrical and Computer Engineering	40	Electrical and Computer Engineering	20	Doctor of Management Program	19	Management Information Systems	6
Technology	39	Information Technology Leadership (formerly Computer Information Systems)	20	Technology Management	19	Business Administration	5
Electrical Engineering	38	Management Information Systems	20	Education	18	Information Technology (PhD)	5
Management Information Systems	38	Information Technology (PhD)	19	Management Information Systems	18	Business	4

Table 4-19 Continued

Any		Primary		Secondary		Solo	
Computer and Information Technology	33	School of Business	18	Information Systems and Communications	17	Computer Information Systems (MCIS, DCIS)	4
Information Technology (PhD)	31	Health Informatics	16	Business and Technology	16	Computer Security and Information Assurance	4
Communication	29	Informatics	15	School of Education	13	Information Studies	4
Nursing	29	Information and Computer Science - M.S	15	Cybersecurity	12	School of Business	4
Business and Technology	27	Computer Information Systems (MCIS, DCIS)	14	Information Technology (PhD)	12	School of Information Technology	4
Education	27	Computer Information Systems	13	Computer Information Systems	11	Business and Technology	3
Information Technology Leadership (formerly Computer Information Systems)	27	Engineering Mgt and Systems Engineering	13	Geography	11	Engineering Mgt and Systems Engineering	3
Information Systems and Communications	25	Business	12	Organizational Leadership	11	College of Business Adminstration	2
Computer Information Systems	24	Electrical Engineering and Computer Science	12	Technology	11	Computing and Information Sciences	2
Technology Management	24	Health Informatics (formerly Medical Informatics)	12	Computer Engineering	10	Department of Management	2
Doctor of Management Program	21	Health Information Management	12	Electrical Engineering	10	Doctor of Management Program	2
School of Education	20	Industrial Engineering	12	School of Business and Technology Management	10	Electrical Engineering and Computer Science	2
Engineering Mgt and Systems Engineering	19	Management	12	Management Science and Systems	8	Informatics	2
Health Informatics	19	Mechanical Engineering	12	Sociology	8	Information Sciences and Technology	2
Computer Information Systems (MCIS, DCIS)	18	School of Public Service Leadership	12	The School of Information Studies	8	Information Systems & Technology Management	2
Informatics	18	Business and Technology	11	Anthropology	7	Management	2
The School of Information Studies	18	Informatics-Information Science	11	Communication and Leadership	7	School of Business and Entrepreneurship	2
Computer Science and Engineering	17	Biomedical Informatics	10	Computer Science and Engineering	7	School of Information Systems and Technology	2
Health Informatics (formerly Medical Informatics)	17	Business Administration	10	English	7	Technology and Information Management	2
Information and Computer Science - M.S	17	Computer Science and Engineering	10	Information Technology Leadership (formerly Computer Information Systems)	7	Technology Leadership and Innovation	2

Table 4-19 Continued

Any		Primary		Secondary		Solo	
Biomedical Informatics	16	Engineering	10	Logistics, Operations, and Management Information Systems	7	Applied Biostatistics and Epidemiology	1
Engineering	16	Information and Computer Science - Ph.D	10	Biomedical Informatics	6	Business Administration (DBA)	1
Health Information Management	15	School of Information Systems and Technology	10	Computer and Information Technology	6	Business and Management: Decision & Information Technologies	1

4.3.4 **ISci**

Computer Science takes the top department for both ISci Primary and Secondary. In fact, CS is by far the most common Department, by a large margin for both ISP and ISS, although especially so for ISP. Beyond the top six ISci Solo and ISci Primary remain mostly technical, while ISci Secondary includes Departments such as English, Psychology, and Nursing.

Table 4-20 Top 40 Departments, ISci (2009-2014)

Any		Primary		Secondary		Solo	
	1,018		176		842		66
Computer Science	169	Computer Science	83	Computer Science	86	Information Science	8
Information Systems	51	Health Information Management	12	Information Systems	40	Information Systems	5
Information Science	43	Information Science	11	Information Studies	34	HKUST	4
Information Systems (DISS)	43	Information Systems	11	Information Systems (DISS)	34	Information & Library Science	4
Information Studies	40	Informatics	9	Information Science	32	Information Studies	4
Informatics	30	Information Systems (DISS)	9	Communication	29	Library and Information Studies	4
Communication	29	Information and Computer Science - Ph.D	7	School of Business and Technology	25	Information Sciences and Technology	3
Informatics-Information Science	28	Information & Library Science	6	The School of Information Studies	25	Information Systems (DISS)	3
School of Business	28	Information Studies	6	Informatics-Information Science	24	Information Technology	3
The School of Information Studies	28	Computer Sciences	5	School of Business	24	Library & Information Services	3
Graduate School - New Brunswick	26	HKUST	5	Applied Management and Decision Sciences	22	Management Information Systems	3
Management Information Systems	26	Library and Information Studies	5	Graduate School - New Brunswick	22	Business Administration	2

Table 4-20 Continued

Any		Primary		Secondary		Solo	
School of Business and Technology	26	Management Information Systems	5	Informatics	21	Communication, Culture & Technology	2
Applied Management and Decision Sciences	25	Networking and Systems Administration	5	Management Information Systems	21	Cybersecurity	2
Information and Computer Science - Ph.D	24	Computer and Information Science	4	Electrical and Computer Engineering	20	Informatics	2
Electrical and Computer Engineering	23	Computer Information Systems (MCIS, DCIS)	4	Electrical Engineering	20	Informatics-Information Science	2
Electrical Engineering	21	Computer Science and Engineering	4	Business	19	Information & Library Science: Information Science	2
Health Information Management	21	Graduate School - New Brunswick	4	Geography	17	Information School	2
Information Technology	20	Informatics-Information Science	4	Information and Computer Science - Ph.D	17	Information Systems and Decision Sciences	2
Library and Information Studies	20	Information Technology	4	Information Systems and Communications	17	School of Business	2
Business	19	School of Business	4	Information Technology	16	Technology	2
Information & Library Science	19	Applied Management and Decision Sciences	3	Library and Information Science	16	Business and Management: Decision & Information Technologies	1
Information Systems and Communications	18	Biomedical Informatics	3	Library and Information Studies	15	Computer Information Systems (MCIS, DCIS)	1
Business Administration	17	bsheaff@uvic.ca	3	Business Administration	14	Computer Science (CISC, CISD)	1
Geography	17	Business Administration	3	Information & Library Science	13	Computer Science and Information Systems	1
Library and Information Science	16	Computer Science & Engineering	3	Information School	12	Computing Science	1
Information School	14	Computer Science (CISC, CISD)	3	Information Sciences and Technology	11	Electrical & Computer Engineering	1
Information Sciences and Technology	14	Computer Science and Information Systems	3	Management	11	Electrical and Computer Engineering	1
Cybersecurity	13	Cybersecurity	3	Cybersecurity	10	Graduate School - New Brunswick	1
Biomedical Informatics	12	Electrical and Computer Engineering	3	Industrial Engineering	10	Graduate School - Newark	1
Computer Information Systems (MCIS, DCIS)	12	Information Sciences and Technology	3	Information and Computer Science - M.S	10	Human Centered Computing	1
Computer Sciences	12	Information Systems and Decision Sciences	3	Biomedical Informatics	9	Industrial Technology	1
Library & Information Services	12	Library & Information Services	3	English	9	Information and Computer Science - Ph.D	1
Industrial Engineering	11	The School of Information Studies	3	Health Information Management	9	Information Management and Systems	1
Information and Computer Science - M.S	11	Applied Science	2	Information Management & Systems	9	Information Resources and Library Science	1

Table 4-20 Continued

Any		Primary		Secondary		Solo	
Management	11	Communication, Culture & Technology	2	Information Technology (PhD)	9	Information Science (DISC)	1
Computer and Information Science	10	Computing Science	2	Library & Information Services	9	Information Science and Technology	1
Engineering Mgt and Systems Engineering	10	Electrical Engineering & Computer Science	2	Nursing	9	Information Studies (College of Computing and Informatics)	1
Information Science (DISC)	10	Engineering Mgt and Systems Engineering	2	Psychology	9	Information Systems and Communications	1

4.4 Keywords

As described in Chapter 3, *Identifier / keyword* is a free-form field in the ProQuest thesis submission form that allows for the addition of any terms that the author thinks will provide greater visibility to search engines. A May 5, 2016 email from Carol Wadke revealed that keywords are the most commonly used search field.

The table below shows the counts of distinct keywords found for each computing discipline from 2009-2014, as well as the number of keywords that are unique to the computing discipline in this time period. Unique means the keyword is *only* associated with the computing discipline for the given position criteria among ALL keywords from all other computing disciplines.

Table 4-21: DISTINCT Keyword Counts Based on CD Position (2009-2014)

	Any	Primary	Secondary	Solo	Any Unique	Primary Unique	Secondary Unique	Solo Unique
CS	39,804	19,453	24,913	19,400	27,434	15,286	13,345	15,241
CE	13,617	12,786	1,270	3,125	5,212	4,837	390	1,668
IT	10,438	6,046	5,588	993	4,573	2,047	2,658	437
ISci	8,291	1,727	7,130	599	3,233	402	2,863	261

All keywords used in findings and discussion will be lower-cased and italicized. The reason is for easier differentiation of the keyword itself being discussed.

Note these are not normalized keywords except on case. Whatever the author input is what gets counted. So, if *java script* and *javascript* are present, they are counted as two different keywords.

4.4.1 Overview of Findings

The most common keyword associated with all computing disciplines is *applied sciences*. It is found 32,139 times out of 33,706 total theses in the dataset from 2009-2014, or 95% of the time. It is found 22,710 times (67%) as the first keyword listed.

The table below shows the top twenty keywords across all disciplines from 2009-2014. After *applied sciences*, *communication and the arts* is by far the second most common—which may be somewhat surprising. The sixth position down is the first non-generic, non-overly-broad term, *machine learning*.

Note the strong presence of *education*, *psychology*, *data mining*, and *security*.

The letter "A" was a surprisingly common keyword, occurring 176 times. It turns out 162 authors input keywords as comma separated word splits of their titles, or of some other statement. Examples include:

- Applied sciences, A, GRASP, BASED, MOTION, PLANNING, ALGORITHM, INTELLIGENT, CHARACTER, ANIMATION
- Communication and the arts, Applied sciences, A, BASED, INFORMATION, KNOWLEDGE, LANGUAGES, NATURAL, RETRIEVAL, STORAGE, SYSTEM

Table 4-22: Top 20 Keywords, All Disciplines (2009-2014)

2009-2014 (keyword)	
applied sciences	32,140
communication and the arts	5,244
social sciences	3,107
biological sciences	1,495
health and environmental sciences	1,436
pure sciences	1,339
machine learning	1,193
education	1,170
psychology	715
data mining	630
security	626
computer vision	546
cloud computing	450
wireless sensor networks	433
wireless networks	348
information technology	346
earth sciences	344
language	333
literature and linguistics	308
privacy	307

There was also often found an odd/unique keyword that appeared to be an identification string. It follows the regular expression pattern:

“\(UMI\)AAIU[0-9]{6}”

Research using Google showed that the code above breaks down to:

- (UMI): University Microfilm Inc.

- AAI: Authentication and Authorization Infrastructure
- [0-9]: The thesis/dissertation number

4.4.2 Keyword Frequency Charts

Below are keyword frequency charts that graph the frequencies of the most commonly used keywords. All disciplines were seen to have a long tail and a very fast, exponential increase in frequency of popular keywords. So, from the keywords above in Table 4-23, keywords such as *applied sciences* and *communication and the arts* would be at the far left, and many of the top 40 keywords, in Tables 4-30 – 4-33, would live on the left as well.

The stepped appearance is due to the sharp jump of moving from hundreds of keywords being equally popular at, for example, three occurences, and then another long stretch of the next most popular keywords that share four as their number of occurrences. Near the origin on the left it can be seen that as the popularity of words increases, the number of words with that level of popularity decreases, and the steps get shorter.

The long tail represents hundreds of words that share a popularity level as measured by the number of occurrences. They are separate words.

4.4.3 Keyword Frequency: Any

Figure 4-20 below shows the frequencies of keywords when all theses all taken into account, regardless of the computing Classification's position in the Classification line.

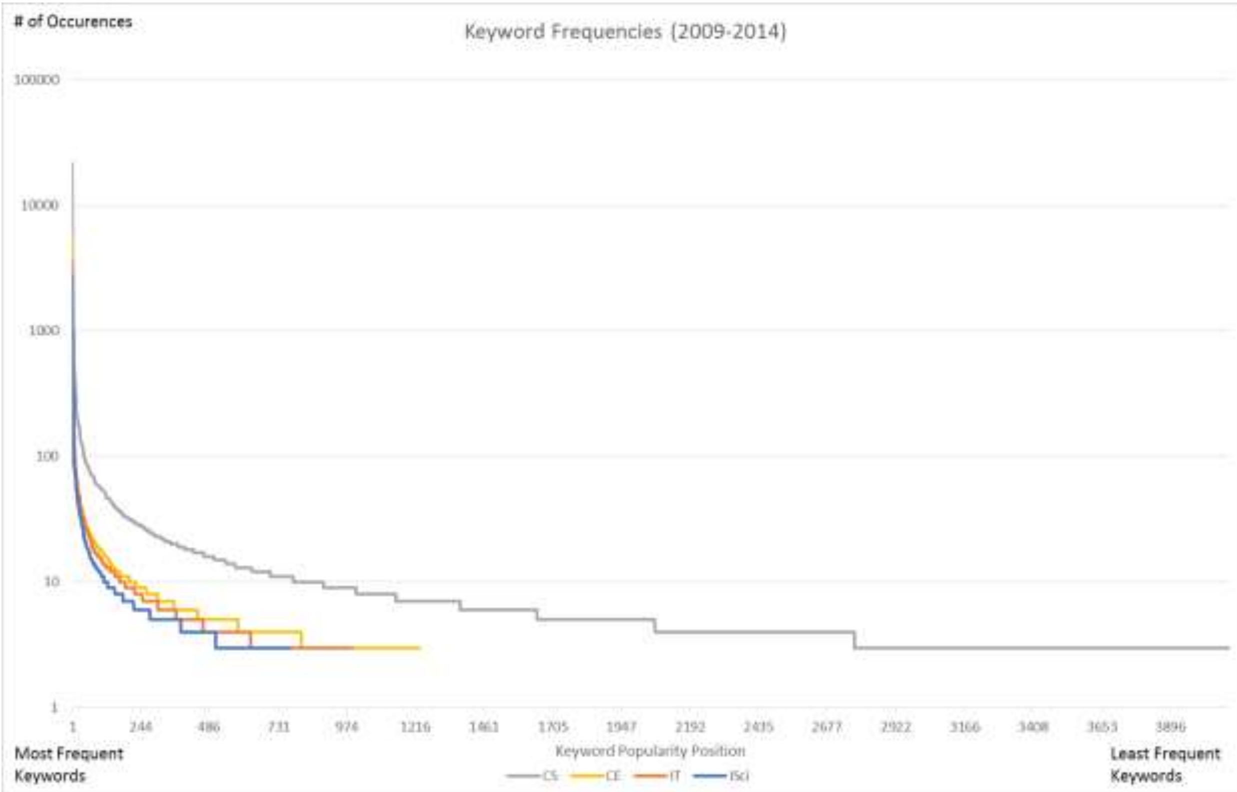


Figure 4-20: Keyword Frequencies, Classification in Any Position, Freq. > 2

The table below shows the number of frequency 1 and 2 (occurring only once or twice) keywords that are not mapped in the above chart. Were these to be included, then due to CS, the tail would stretch to 35,700 to the right instead of 4,104, obfuscating any trending information on the left.

Table 4-23: Keyword Frequencies Not Shown, Classification in Any Position

Discipline	CS	CE	IT	ISci
Freq. 1	32,221	11,198	8,498	6,798
Freq. 2	3,479	1,186	948	719
Total not shown	35,700	12,384	9,446	7,517

4.4.4 Keyword Frequency: Primary

ISci's range of keywords is narrow when Primary.

CEP uses a relatively large number of popular keywords, which appears to mirror CS.

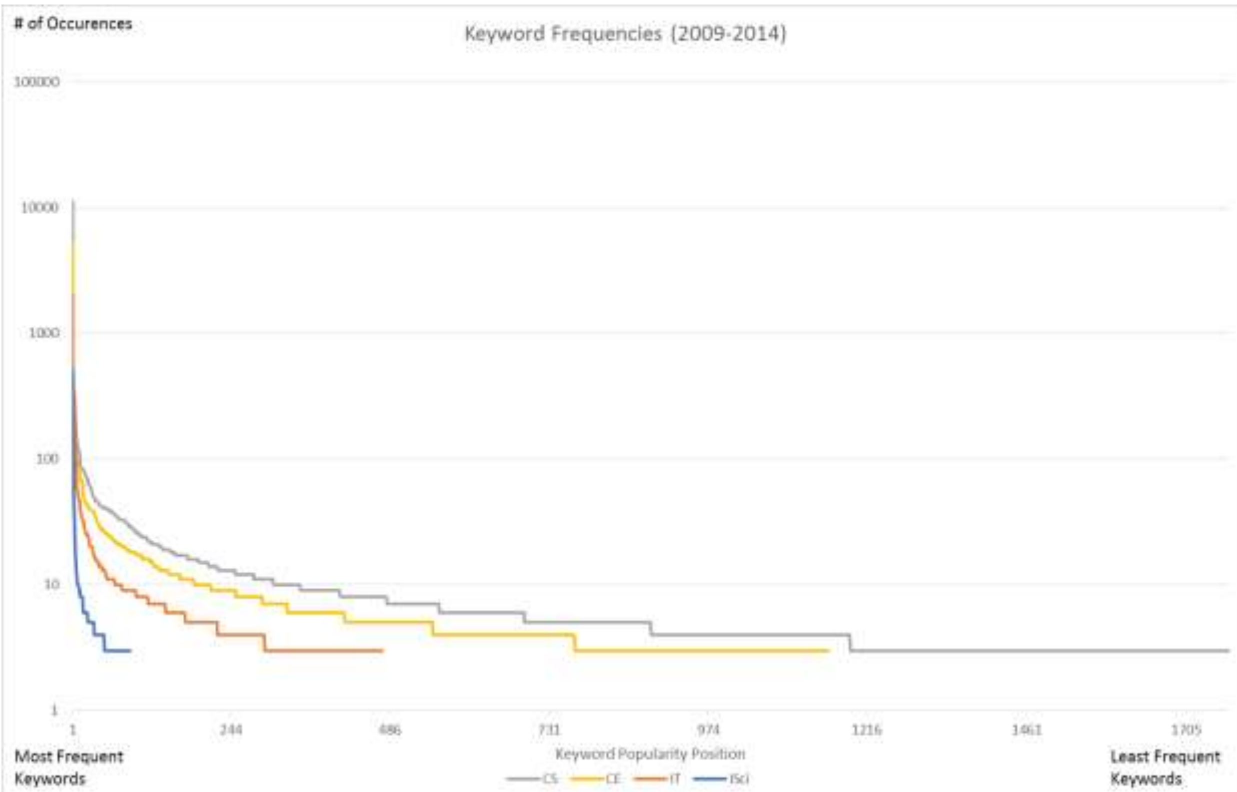


Figure 4-21: Keyword Frequencies, Classification in Primary Position, Freq. > 2

Table 4-24: Keyword Frequencies Not Shown, Classification in Primary Position

Discipline	CS	CE	IT	ISci
Freq. 1	16,066	10,511	5,100	1,537
Freq. 2	1,616	1,117	471	101
Total not shown	17,682	11,628	5,571	1,638

4.4.5 Keyword Frequency: Secondary

The chart below shows CES using very few keywords.

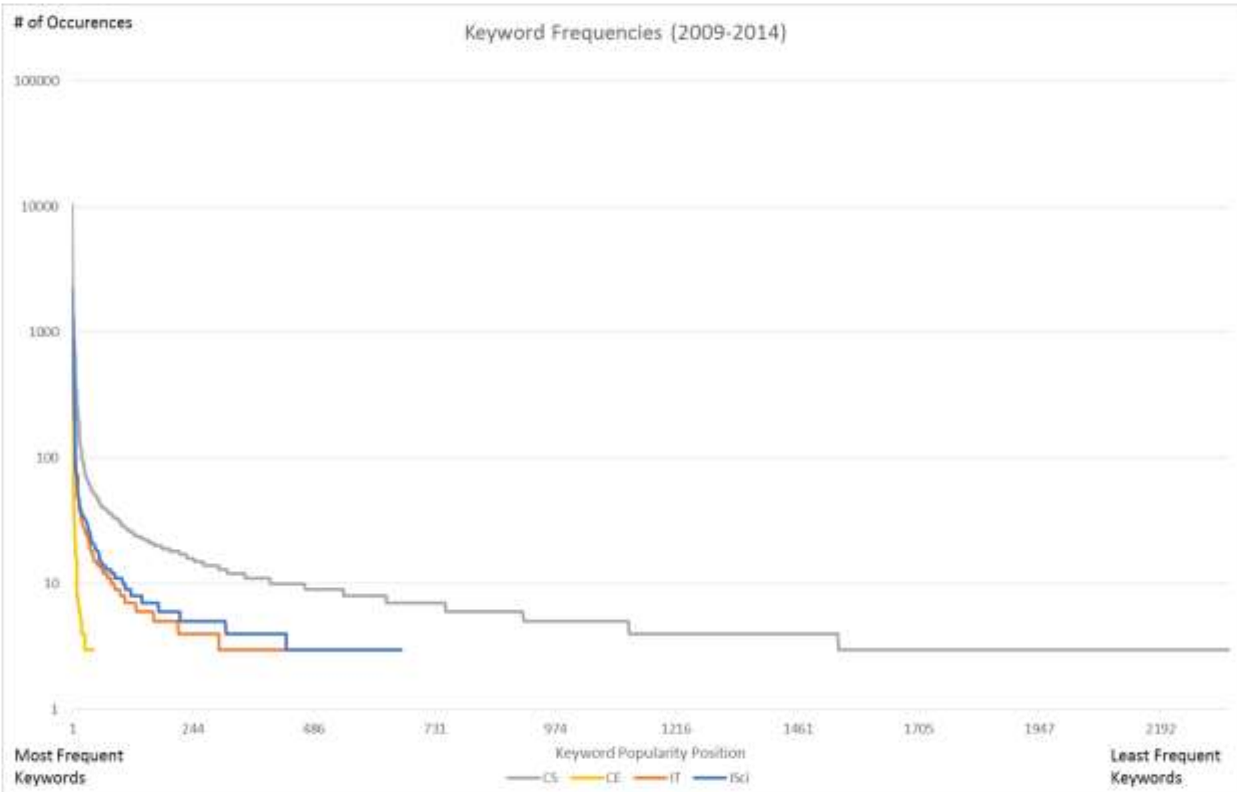


Figure 4-22: Keyword Frequencies, Classification in Secondary Position, Freq. > 2

Table 4-25: Keyword Frequencies Not Shown, Classification in Secondary Position

Discipline	CS	CE	IT	ISci
Freq. 1	20,327	1,171	4,611	5,873
Freq. 2	2,257	56	494	594
Total not shown	22,584	1,227	5,105	6,467

4.4.6 Keyword Frequency: Solo

Note the exceedingly high numbers of keywords for CS Solo and the exceedingly low numbers of keywords for ISci and IT Solo.

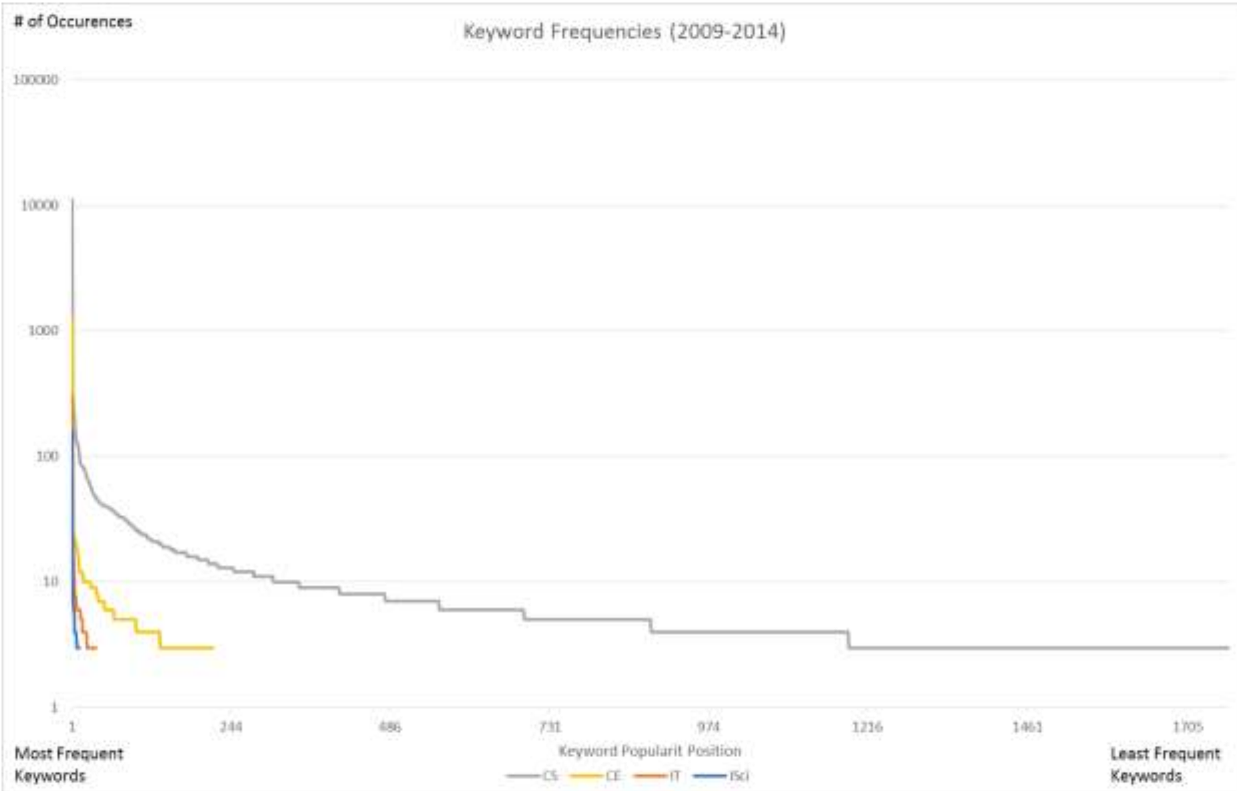


Figure 4-23: Keyword Frequencies, Classification as Solo, Freq. > 2

Table 4-26: Keyword Frequencies Not Shown, Classification as Solo

Discipline	CS	CE	IT	ISci
Freq. 1	16,021	2,664	886	552
Freq. 2	1,611	245	70	35
Total not shown	17,632	2,909	956	587

4.4.7 Keyword Frequency: Primary Unique

The Primary Unique and Solo Unique charts below illustrate the keywords that are found in Primary and Solo theses, but are not found in any of the other three computing disciplines. The keywords do repeat within the discipline across positional charts

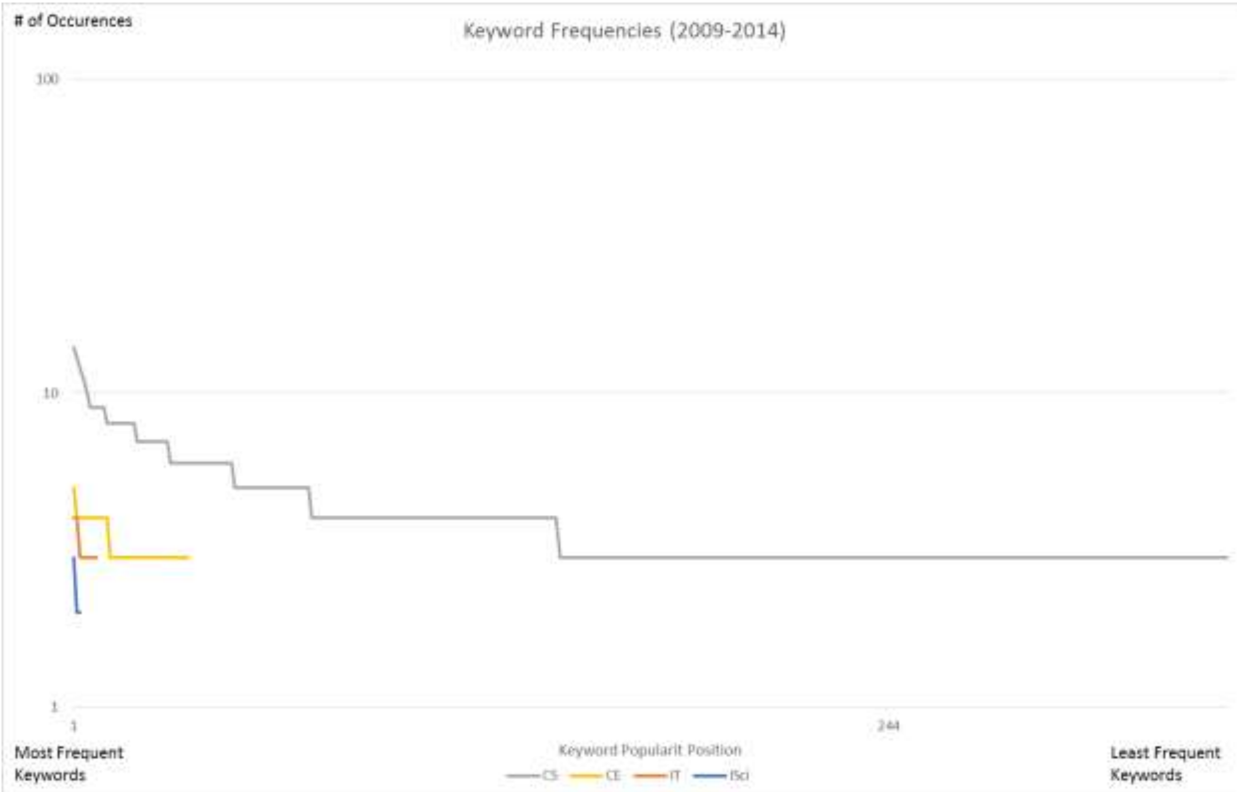


Figure 4-24: Keyword Frequencies, Classification in Primary Position, Unique Keywords, Freq. > 2

Table 4-27: Keyword Frequencies Not Shown, Classification in Primary Position Unique Keywords

Discipline	CS	CE	IT	ISci
Freq. 1	14,075	4,635	1987	399
Freq. 2	866	167	52	2
Total not shown	14,941	4,802	2039	401

4.4.8 Keyword Frequency: Primary Solo Unique

These are all keywords from the universe of all keywords present in the dataset from 2009-2014 that only exist in a thesis that is Primary and Solo for each CD. The tail, especially for CS and even CE is long, but note how few there are for IT and ISci. The top 40 terms can be seen below in Tables 4-29 thru 4-32.

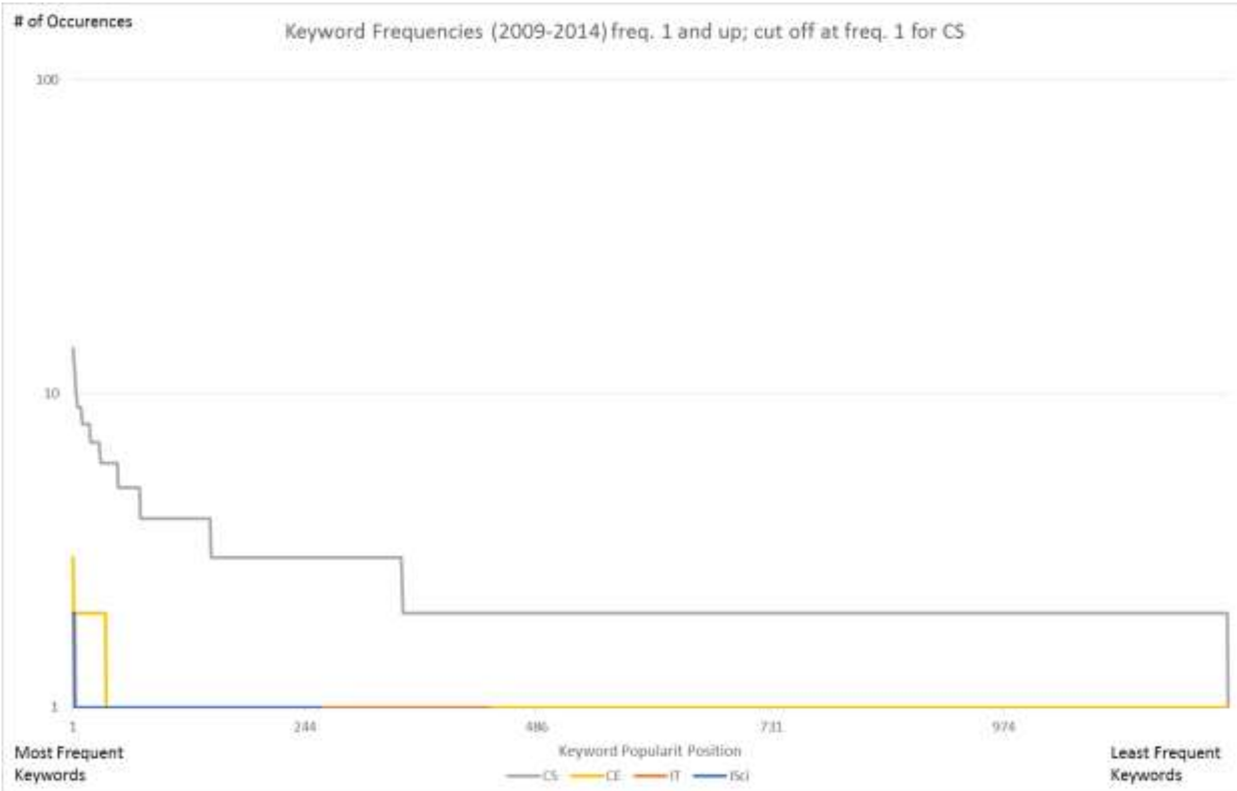


Figure 4-25: Keyword Frequencies, Classification as Solo, Unique Keywords, up to First CS Freq. of. 1—Popularity Position 1,209

Table 4-28: Keyword Frequencies Not Shown for Classification as Solo, Unique Keywords

Discipline	CS	CE	IT	ISci
Freq. 1	14,032	459	0	0
Freq. 2	0	0	0	0
Total not shown	14,032	459	0	0

4.4.9 **Top 40 Keywords by Discipline**

Keywords for each discipline as Secondary will be the same as the Primary when Secondary, since the latter is simply a measure of the Primary Classification whenever the computing discipline is not the Primary. Thus, for example, every CSS thesis is also a PCSS thesis—because each CSS *will* have a Primary that is not CS. Keyword counts are measures against theses not Classifications.

4.4.10 **Top 40 Keywords for CS**

Figure 4-29 shows the top 40 keywords for CS.

Keywords for CSS and PCSS will be the same, since every thesis that isn't CSP is going to be CSS, and PCSS is just a count of the Primary Classifications in CSS, which in turn is just a count of the number of CSS theses.

The keyword applied sciences is present in every one of the 21,628 CS theses from 2009 to 2014 and is the most used keyword across all four computing disciplines. In the complete dataset, 95,992 CS theses out of 96,075 use the keyword applied sciences.

CS contains 39,804 distinct total keywords or phrases.

Out of 24,913 distinct keywords in CSS, only 4,561 match the 19,453 in CSP for a 23% match rate, and CSS matches 4,550 of the 19,400 in CSO (CS Solo, or Only) for a 23% match rate.

4.4.11 **Top 40 Keywords for CE**

Figure 4-30 shows the top 40 keywords for CE in its various Classification positions, as well as the unique-to-CE keywords for each position.

CE contains 13,617 distinct total keywords or phrases.

Out of 1,270 distinct keywords in CES, only 439 match the 12,786 in CSP for a 3% match rate, and CES keywords match 208 of the 3,125 in CEO (CE Solo, or Only) for a 7% match rate.

Table 4-29: Top 40 Keywords CS, (2009-2014)

Any		Primary		Secondary		Solo		P Uniq		S Uniq		Solo Uniq	
applied sciences	21,628	applied sciences	11,151	applied sciences	10,477	applied sciences	11,134	Javascript	14	comparative genomics	18	javascript	14
communication and the arts	1,323	machine learning	305	communication and the arts	1323	machine learning	304	uncertain data	13	systems biology	17	uncertain data	13
biological sciences	1,234	security	225	biological sciences	1234	security	225	heuristic search	12	gene regulatory networks	14	heuristic search	12
pure sciences	1,073	data mining	224	pure sciences	1063	data mining	224	secure computation	11	heuristic search	11	secure computation	11
machine learning	963	computer vision	196	social sciences	701	computer vision	196	local search	10	structure learning	11	local search	10
social sciences	702	wireless sensor networks	161	machine learning	658	wireless sensor networks	161	delay-tolerant networks	9	combinatorics	11	delay-tolerant networks	9
health and environmental sciences	554	cloud computing	135	health and environmental sciences	552	cloud computing	135	constraint satisfaction	9	conditional random fields	11	conditional random fields	9
data mining	468	wireless networks	129	education	363	wireless networks	129	large-scale	9	rna-seq	11	large-scale	9
computer vision	447	distributed systems	128	psychology	337	distributed systems	128	conditional random fields	9	biomarkers	10	constraint satisfaction	9
security	388	privacy	118	computer vision	251	privacy	118	global illumination	9	population genetics	9	global illumination	9
education	368	software engineering	114	data mining	244	software engineering	114	type inference	8	next generation sequencing	9	online algorithms	8
psychology	339	social networks	104	earth sciences	233	social networks	104	Internet routing	8	matrix completion	9	character animation	8
wireless sensor networks	288	computer graphics	88	language	201	computer graphics	88	speculation	8	random graphs	9	type inference	8
cloud computing	257	natural language processing	86	literature and linguistics	190	natural language processing	86	programming language	8	computer-aided diagnosis	8	Internet routing	8
earth sciences	233	clustering	85	security	163	clustering	85	service discovery	8	bayesian statistics	8	programming language	8
wireless networks	227	cryptography	84	wireless sensor networks	127	cryptography	84	online algorithms	8	cognitive modeling	8	service discovery	8
natural language processing	208	computer security	81	cloud computing	122	computer security	81	character animation	8	mixed integer programming	8	speculation	8
language	206	visualization	81	natural language processing	122	visualization	81	lower bounds	8	single nucleotide polymorphisms	8	cryptographic protocols	8
literature and linguistics	190	information retrieval	81	optimization	108	information retrieval	81	cryptographic protocols	8	computational	8	lower bounds	8
privacy	189	scheduling	78	wireless networks	98	scheduling	78	markov decision processes	7	multigrid	8	type theory	7

Table 4-29 Continued

Any		Primary		Secondary		Solo		P Uniq		S Uniq		Solo Uniq	
clustering	178	programming languages	73	information retrieval	96	programming languages	73	texture synthesis	7	phylogenetic trees	7	markov decision processes	7
social networks	177	network security	72	artificial intelligence	94	network security	72	program verification	7	transcription factors	7	texture synthesis	7
information retrieval	177	sensor networks	69	clustering	93	optimization	69	sparql	7	bayesian inference	7	relational learning	7
optimization	177	optimization	69	image processing	85	sensor networks	68	relational learning	7	robot control	7	sparql	7
distributed systems	175	peer-to-peer	63	visualization	78	peer-to-peer	63	clone detection	7	branch and bound	6	tiling	7
software engineering	164	routing	63	social networks	73	routing	63	nearest neighbor search	7	microrna	6	nearest neighbor search	7
visualization	159	algorithms	62	classification	73	algorithms	62	hash functions	7	film	6	clone detection	7
cryptography	147	parallel computing	60	privacy	71	parallel computing	60	type theory	7	virtual screening	6	hash functions	7
scheduling	139	virtualization	58	pattern recognition	70	virtualization	58	tiling	7	brain imaging	6	program verification	7
artificial intelligence	135	intrusion detection	57	neural networks	66	intrusion detection	57	determinism	6	protein-protein interaction	6	real-time rendering	6
sensor networks	135	operating systems	53	algorithms	66	operating systems	53	real-time rendering	6	phonology	6	randomness	6
computer graphics	131	databases	52	sensor networks	66	databases	52	multipath	6	protein function	6	information flow control	6
algorithms	128	distributed computing	50	cryptography	63	distributed computing	50	code clones	6	micrornas	6	determinism	6
image processing	125	human-computer interaction	49	parallel computing	62	human-computer interaction	49	constraint programming	6	microarray data	6	language design	6
computer security	123	semantic web	49	human-computer interaction	62	semantic web	49	language design	6	genome-wide association studies	6	formal specification	6
parallel computing	122	fault tolerance	49	scheduling	61	fault tolerance	49	theoretical computer science	6	structural biology	5	reasoning	6
network security	122	energy efficiency	46	energy efficiency	60	energy efficiency	46	reasoning	6	protein design	5	generic programming	6
classification	112	object recognition	46	simulation	57	object recognition	46	information flow control	6	branch-and-bound	5	constraint programming	6
human-computer interaction	111	compilers	46	bioinformatics	55	compilers	46	cross-site scripting	6	gene expression data	5	data replication	6
energy efficiency	106	game theory	45	philosophy	54	game theory	45	data replication	6	deconvolution	5	theoretical computer science	6

Table 4-30: Top 40 Keywords CE, (2009-2014)

Any		Primary		Secondary		Solo		P Uniq		S Uniq		Solo Uniq	
applied sciences	5,538	applied sciences	5,231	applied sciences	307	applied sciences	1291	silicon photonics	5	terahertz	2	area	3
pure sciences	188	machine learning	124	pure sciences	89	computer architecture	25	gnu radio	4	control-display gain	1	application specific hardware	2
communication and the arts	154	wireless sensor networks	116	biological sciences	49	energy efficiency	23	hardware trojans	4	correlated noise	1	softcore	2
machine learning	142	communication and the arts	106	communication and the arts	48	fpga	23	aodv	4	progression prediction	1	branch misprediction	2
wireless sensor networks	119	pure sciences	99	social sciences	27	wireless sensor networks	22	routability	4	linear-invariants	1	hybrid computing	2
biological sciences	116	wireless networks	98	health and environmental sciences	21	wireless networks	20	power amplifiers	4	numerical modelling	1	general purpose graphical processing units	2
security	103	security	97	machine learning	18	machine learning	20	reversible logic	4	elastic laminae	1	microarchitectures	2
health and environmental sciences	103	computer vision	87	earth sciences	16	cloud computing	18	adc	4	geometry independence	1	aodv	2
wireless networks	99	health and environmental sciences	82	psychology	15	embedded systems	17	on-chip network	4	cnc interpolator	1	circuit watermarking	2
computer vision	89	computer architecture	80	language	8	security	16	dc-dc converters	4	decentralized multi-armed bandit	1	cmp	2
cloud computing	83	cloud computing	80	signal processing	8	fpgas	15	seu	4	musical instrument	1	value prediction	2
computer architecture	81	fpga	77	literature and linguistics	7	performance	12	transient analysis	3	line-commutated converters	1	high performance architectures	2
fpga	78	energy efficiency	74	education	7	routing	12	reconfigurable logic	3	instruments	1	transient faults	2
energy efficiency	76	biological sciences	67	cryptography	6	reliability	12	null convention logic	3	circuit simulators	1	distributed real-time scheduling	2
optimization	69	embedded systems	67	data mining	6	gpu	12	resistive ram	3	phoneme recognition	1	finite state machine watermarking	2
embedded systems	68	optimization	64	security	6	parallel computing	12	hybrid computing	3	partitioning around medoids	1	workload variation	2
social sciences	66	scheduling	52	philosophy	5	scheduling	11	distribution system	3	spectral multidomain penalty	1	tradeoffs	2
data mining	55	data mining	49	optimization	5	graphics processing units	11	quantum dots	3	clave	1	runtime system	2
scheduling	52	fpgas	49	religion and theology	5	chip multiprocessors	10	nanowire	3	event related potential	1	memory encryption	2

Table 4-30 Continued

Any		Primary		Secondary		Solo		P Uniq		S Uniq		Solo Uniq	
image processing	51	image processing	47	software development	4	network security	10	high speed	3	finite geometries	1	storage class memory	2
routing	49	routing	45	image processing	4	compilers	10	partial	3	psychotropic medication	1	path delay	2
fpgas	49	gpu	44	software engineering	4	vlsi	10	model predictive control	3	multilevel design optimization	1	gpu architectures	2
parallel computing	47	parallel computing	44	routing	4	computer vision	10	hardware accelerators	3	temperature distribution	1	performance optimizations	2
gpu	46	virtualization	41	brain-computer interface	4	image processing	10	radiation hardening	3	korean pop culture	1	photonic	2
cryptography	44	network security	41	wireless	4	parallel programming	10	transient faults	3	academic assessment	1	power analysis attacks	2
network security	44	performance	41	performance evaluation	3	fault tolerance	10	superscalar processors	3	semiclassical tunneling model	1	reconfigurability	2
reliability	43	quality of service	40	parallel computing	3	network coding	10	cooperative control	3	visualization design	1	radiation hardening	2
virtualization	43	reliability	40	wireless sensor networks	3	optimization	10	area	3	long term investment	1	maxwell's equations	2
performance	42	resource allocation	40	network security	3	network-on-chip	10	dynamic reconfiguration	3	processed recordings	1	on-chip network	2
resource allocation	41	fault tolerance	40	renewable energy	3	virtualization	9	nanoelectronics	3	techno-economic optimization	1	minority game	2
fault tolerance	41	social sciences	39	semantics	3	gpgpu	9	delay faults	3	complex neural computation	1	feature reduction	2
quality of service	40	cryptography	38	database	3	quality of service	9	through-silicon vias	3	periodic	1	quantum cost	2
earth sciences	40	vlsi	38	networks	3	reconfigurable computing	9	microcontroller	3	sparse model learning	1	simultaneous multi-threading	2
cognitive radio	38	cognitive radio	37	sensor networks	3	system-on-chip	9	power analysis attacks	3	computer audition	1	hybrid noc	2
vlsi	38	low power	37	networking	3	high performance computing	9	power supply noise	3	rock fracturing	1	power supply noise	2
sensor networks	37	network-on-chip	35	topology	3	multicore	9	vector processing	2	convergence rate	1	biomolecular simulations	1
low power	37	sensor networks	34	tracking	3	cache	9	supercapacitors	2	airblast sprayers	1	cvt	1
network-on-chip	35	clustering	32	distributed systems	3	cryptography	8	multimedia networking	2	dual control problem	1	pixel truncation	1
clustering	34	multicore	31	simulation	3	memory	8	multiuser transmission	2	line noise	1	developers' expertise	1
wireless	34	chip multiprocessors	31	neural networks	3	mobile computing	8	embedded system security	2	bioeconomics	1	msg	1

Table 4-31: Top 40 Keywords IT, (2009-2014)

Any		Primary		Secondary		Solo		P Uniq		S Uniq	
applied sciences	3,748	applied sciences	2,019	applied sciences	1,729	applied sciences	304	nursing documentation	4	firm performance	7
social sciences	1,489	communication and the arts	496	social sciences	1,169	information security	20	health information systems	4	organizational agility	4
communication and the arts	975	health and environmental sciences	340	communication and the arts	479	cybersecurity	12	patient engagement	3	coping	4
health and environmental sciences	454	social sciences	320	education	211	security	9	data breaches	3	patents	4
education	421	education	210	information technology	193	information technology	8	health care organizations	3	it outsourcing	4
information technology	262	information security	71	psychology	124	software development	8	health it	3	banking industry	4
psychology	179	information technology	69	health and environmental sciences	114	cloud computing	7	pediatrics	3	chief information officers	4
information security	113	security	60	technology	71	project management	6	quality of care	3	disruptive technology	4
technology	110	psychology	55	information systems	62	technology acceptance model	6	nursing education	2	workforce	4
information systems	94	electronic health records	49	project management	59	biometrics	6	security effectiveness	2	business process outsourcing	4
security	92	cloud computing	47	knowledge management	49	data mining	6	it outsourcing	2	cio	4
cloud computing	85	cybersecurity	46	leadership	48	information systems	6	intensive care unit	2	competitive advantage	4
project management	74	technology	39	social media	47	governance	6	lean	2	it investments	4
knowledge management	70	data mining	38	information security	42	technology acceptance	5	theory u	2	senior management	3
social media	67	health information technology	34	cloud computing	38	text mining	5	icann	2	legislation	3
leadership	64	biological sciences	33	technology adoption	37	technology	5	electronic personal health record	2	it personnel	3
Internet	62	machine learning	32	innovation	37	virtualization	4	information systems success	2	transformational	3
biological sciences	62	information systems	32	Internet	36	risk management	4	qualitative methods	2	labor	3
electronic health records	62	privacy	29	security	32	information	4	earned value management	2	boundary spanning	3
cybersecurity	59	information retrieval	28	language	31	knowledge management	4	asian indian	2	business-it alignment	3

Table 4-31 Continued

Any		Primary		Secondary		Solo		P Uniq		S Uniq	
technology adoption	57	Internet	26	communication	30	enterprise resource planning	4	electronic documentation	2	it project management	3
data mining	53	electronic medical records	25	biological sciences	29	leadership	4	self-confidence	2	social software	3
technology acceptance	51	technology acceptance	25	literature and linguistics	29	software	4	it diffusion	2	civic engagement	3
software development	49	usability	25	earth sciences	28	human-computer interaction	3	ttf	2	it organizations	3
information and communication technologies	49	software development	24	social networks	27	forensics	3	ehr adoption	2	it investment	3
privacy	48	information and communication technologies	24	technology acceptance	26	mobile technology	3	digital natives	2	middle managers	3
innovation	48	knowledge management	21	management	26	malware	3	student participation	2	employee retention	3
health information technology	45	technology adoption	20	software development	25	e-commerce	3	cameroon	2	agency theory	3
language	42	social media	20	information and communication technologies	25	network security	3	long-term care	2	it managers	3
social networks	41	natural language processing	20	virtual teams	24	social networks	3	redcap	2	mergers and acquisitions	3
literature and linguistics	40	collaboration	20	trust	24	business intelligence	3	organizational commitment	2	organizational commitment	3
usability	39	healthcare	19	philosophy	23	user experience	3	preceptor	2	it professionals	3
trust	37	technology acceptance model	17	religion and theology	23	instant messaging	3	skills	2	computer-assisted instruction	3
communication	36	electronic health record	17	technology management	22	visual analytics	3	patient care delivery	2	work-life balance	3
machine learning	35	leadership	16	enterprise resource planning	20	virtual teams	3	capability approach	2	bring your own device	3
collaboration	35	information	16	privacy	19	information retrieval	3	identity fraud	2	organizational	3
human-computer interaction	34	human-computer interaction	16	outsourcing	19	innovation	3	preceptorship	2	policy implementation	3
earth sciences	34	semantic web	16	human-computer interaction	18	organizational change	2	physician resistance	2	dynamic capabilities	3
electronic medical records	34	software	15	project success	18	access control	2	information warfare	2	health care organizations	3
virtual teams	33	simulation	15	decision-making	18	security policy	2	delphi study	2	videoconferencing	3

Table 4-32: Top 40 Keywords ISci, (2009-2014)

Any		Primary		Secondary		Solo		P Uniq		S Uniq	
communication and the arts	2,792	communication and the arts	506	communication and the arts	2,286	communication and the arts	159	information behavior	3	information behavior	30
applied sciences	1,226	applied sciences	283	applied sciences	943	information retrieval	7	museum studies	2	academic libraries	20
social sciences	850	health and environmental sciences	52	social sciences	841	information technology	6	common ground	2	public libraries	7
education	354	information retrieval	33	education	343	data mining	5	corporate security	1	723	7
health and environmental sciences	325	machine learning	18	health and environmental sciences	273	technology acceptance	4	imax	1	information organization	6
psychology	164	data mining	14	psychology	163	social media	4	dust aerosols	1	turkish national police	6
social media	87	security	11	biological sciences	83	security	4	social tagging systems	1	information behaviors	5
biological sciences	83	education	11	social media	77	human-computer interaction	3	audience design	1	knowledge organization	5
Internet	77	human-computer interaction	10	language	75	usability	3	geo-local systems	1	blogging	5
language	75	social media	10	Internet	74	information	3	hospital visits	1	health disparities	5
information retrieval	71	social sciences	9	literature and linguistics	71	privacy	3	hispanic	1	information policy	5
literature and linguistics	71	clustering	9	knowledge management	54	information systems	3	arkose	1	hofstede	4
knowledge management	58	information technology	8	pure sciences	48	risk assessment	2	jacobs	1	hurricane katrina	4
information technology	55	text mining	8	information technology	47	entropy	2	instituto dois irmaos	1	library instruction	4
data mining	54	electronic medical records	8	online	44	metadata	2	information source	1	purchasing behavior	4
machine learning	53	privacy	8	data mining	40	risk management	2	networking security	1	social tags	4
pure sciences	49	natural language processing	6	social networks	40	data management	2	germane document identification	1	teens	4
online	47	information visualization	6	information retrieval	38	machine learning	2	message content	1	science studies	3
human-computer interaction	47	search	6	human-computer interaction	37	information behavior	2	conceptual network	1	school libraries	3
social networks	46	information	6	machine learning	35	methodology	2	landscape contracting	1	kuwait	3
security	43	semantic web	6	information systems	35	healthcare	2	psychological empowerment	1	section 508	3

Table 4-32 Continued

Any		Primary		Secondary		Solo		P Uniq		S Uniq	
privacy	43	social networks	6	privacy	35	network security	2	time pressure	1	information ethics	3
information systems	40	artificial intelligence	6	technology	34	cscw	2	data registry	1	international law	3
earth sciences	37	health informatics	5	earth sciences	34	social networks	2	automatic categorization	1	institutionalization	3
information	36	classification	5	religion and theology	33	Internet	2	body language	1	online search	3
collaboration	36	electronic health records	5	philosophy	33	hci	2	brain fibers	1	health behavior	3
information security	36	computer vision	5	security	32	data quality	2	consumer motivation	1	human information behavior	3
technology	34	wireless sensor networks	5	collaboration	32	information literacy	2	visual narrative	1	millennial generation	3
religion and theology	33	ontology	5	information security	31	text mining	2	online consumer reviews	1	scientometrics	3
information behavior	33	technology acceptance	5	information	30	user experience	2	hubbert peak	1	undergraduates	3
philosophy	33	information systems	5	information behavior	30	online	2	indigenous knowledge	1	information infrastructure	3
natural language processing	32	information security	5	information literacy	29	supply chains	2	scientific data repositories	1	information seeking behavior	3
technology acceptance	31	information extraction	5	facebook	26	social informatics	2	mediated self-reflection	1	digital archive	3
information literacy	31	risk assessment	4	technology acceptance	26	information management	2	network community detection	1	knowledge-sharing	3
information seeking	29	metadata	4	natural language processing	26	social network analysis	2	information systems design	1	self-presentation	3
electronic health records	28	information science	4	information seeking	26	data visualization	2	dynamic decision-making	1	community archives	3
facebook	27	user experience	4	communication	24	context	2	laboratory test	1	pennsylvania	3
electronic medical records	26	cloud computing	4	electronic health records	23	learning	2	human crafters	1	consumer health information	3
cloud computing	25	usability	4	social network analysis	22	telecommunications	2	graduate student	1	police	3
social network analysis	24	network security	4	technology adoption	21	software development	2	medical humanities	1	records management	3

4.4.12 **Top 40 Keywords for IT**

Table 4-31 above shows the top 40 keywords for IT.

IT contains 10,438 distinct total keywords or phrases.

Out of 5,588 distinct keywords in ITS, 1,196 match the 6,046 in ITP for a 20% match rate, and ITS keywords match 363 of the 993 in ITO (IT Solo, or Only) for a 37% match rate.

4.4.13 **Top 40 Keywords for ISci**

Table 4-32 above shows the top 40 keywords for ISci.

ISci contains 8,292 distinct total keywords or phrases.

Out of 7,131 distinct keywords in ISci, 566 match the 1,727 in ISP for a 33% match rate, and ISS keywords match 262 of the 599 in ISO (ISci Solo, or Only) for a 44% match rate.

5 DISCUSSION

5.1 ProQuest Overall Numbers

5.1.1 Overall Thesis Numbers and Trends for CS

Figure 4-1 shows relatively flat numbers for total CS theses, with a small dip in CS interest, based on numbers of theses, roughly corresponding to the dot-com bust in 2000-2001. From 2001 onwards the number of CS theses shows a strong increasing trend for six to seven years before leveling off again.

The slight decline in interest in CS graduate programs in ’96 and ’97 was likely due to students choosing to go to work rather than complete an advanced degree. It likely meant they chose to quite or postpone their graduate work as the 2-3 year lead time for master’s students and 7-8 year (O’Shaughnessy, 2012; InfoBrief, 2006) lead time for doctoral students would suggest a lag in any attempts to follow the business cycle. Positive movements in the business cycle should appear more quickly as it is easier to stop while in a program than to start or finish one. So, the dip in 2000-2001 is likely not a result of people leaving school for work, but the lack of potential graduates entering graduate programs during the heady days of ’94 through ’99.

These economic cycles also explain the large rise in theses in the ensuing years from 2003-2004 to 2006 -2007, which is about the amount of time required to finish a master’s or doctoral degree, respectively. This increase in theses in from 2001-2007 speaks to normal behavior of people going back to school, or staying in school, during a recession (Parker, 2015).

Also note from Figure 4-1 the remarkable drop in Primary CS theses and the dramatic rise of Secondary CS theses, especially in the 2009-2014 timebox. The trend appears to begin around 2006, when the number of Primary CS theses actually drops from the previous year, despite an overall increase in the number of CS theses. This is likely due to the addition of other computing disciplines into ProQuest. These other disciplines include Information Science, Computer Engineering, and Information Technology. Thus, Computer Science is still actively being included, but now it is moving to a secondary position for those who use the primary classification of CE, IT, or ISci. This should show in large numbers of CS Secondary appearing under ISci, CE, and IT. And this is in fact the case. In 2006 and 2007 there are no CE co-Classifications for CS but in 2009-2014 CE is the top co-Classification and the top Primary when CS is Secondary. IT also makes the expected strong appearance being the third most frequent Primary for CS when CS is Secondary.

CS may appear to be evenly divided between CS Primary and CS Secondary. But, a look at the Solo column in Table 4-1 shows that CS is pre-dominantly Solo if it is Primary.

Why CS is so seemingly insular is unknown. An explanation could simply be that CS, as a discipline, feels there is more than enough within its own area of research that CS authors feel no need to include other disciplines, at least in recent years. It could also be a reaction to “resume inflation” that other disciplines may be subject to, although this is speculation. Regardless of the reason, and since authors’ intents are not extractable, the data shows that CS Primary is almost entirely NON-interdisciplinary. So, if there is a co-Classification for CS, almost 100% of the time it is another discipline including CS rather than the other way around.

5.1.2 Overall Thesis Numbers and Trends for CE

In 2007 the first theses for CE began to appear and they don't appear in significant numbers till 2009. It is reasonable to believe that CE was added in 2009 based on these numbers and the information from ProQuest Specialist Carol Wadke in an email from October 13, 2014 stating that, "Computer Engineering was added around 2009."

CE has seen steady growth, never seeing a year-over-year decline. The growth has come almost entirely from CE Primary theses. This shows a willingness by CE to include other disciplines even though not many others claim, or are able to claim, CE.

5.1.3 Overall Thesis Numbers and Trends for IT

IT overall numbers have seen a slightly positive trend in the five years from 2010-2014. 2009 understandably has lower numbers as IT was introduced that year.

IT has shown a remarkably even balance of Primary vs. Secondary theses and its Solo theses percentage has also held steady over the total six year period in this study at roughly 7-9%.

The lack of solo theses for IT suggests it is highly interdisciplinary. IT's low overall total Solo theses percentage of 8% means that, by-and-large, even if IT is the Primary, it will have other Classifications 85% of the time.

The larger co-Classification rate shows IT is more broadly used, or used with a wider spectrum of other Classifications, when it is Secondary versus when it is Primary.

5.1.4 **Overall Thesis Numbers and Trends for ISci**

ISci theses show a steady decline from 2009 till 2013 and then in 2014 show an uptick in total numbers as well as Primary numbers. Whether this uptick is simply part of a normal cycle or indicates a strong upswing is unknown.

ISci self-identifies most with CS, based on CS being the top co-Classification with ISci Primary. However, looking at PICS from Table 4-10 two things stand out. First, IT is ISci's most common Primary when ISci is Secondary. However, the next four are LIBRARY SCIENCE, MANAGEMENT, COMMUNICATIONS, and WEB SCIENCE. Every one of the four, except MANAGEMENT, is in the same sub-heading group suggested by the ProQuest Subject Categories Guide or the publishing agreement (*ProQuest 2015-2016 Publishing Agreement*). Whether this is behavior *due* to the ProQuest guide or authors really believe this is how their thesis should be classified is unknown. If it's the former it shows the possible influence that ProQuest category suggestions can have over the classification of theses. So, when ProQuest places IT under the sub-heading Interdisciplinary, what hint does that give to authors, especially non-IT practitioners? It would appear to welcome the inclusion of IT in various studies that include some sort of computing. A future dive into the abstracts would shed more light on this question.

5.1.5 **Overall**

Computing theses seem to have leveled off in the 2009-2014 time period.

Again, from Figure 4-7, it is seen that the Solo theses percentage trends for IT and ISci appear level but CS and CE each show large downward trends, with CS declining from a yearly high of 62% Solo in 2009 to only 39% in 2014. CE declines from a yearly high in 2009 of 43% to a low in 2014 of 15%. This trend may be expected for CS given its downward trend in

Primary theses. But for CE, it is surprising because CE theses are still primarily CE, and the overall number of CE theses is increasing. So why is the number of CE Solo theses declining? It is because co-Classifications for CE Primary theses are increasing.

The addition of the three additional computing options for ProQuest likely explains the slight decrease in overall ProQuest CS numbers as well as the large drop in CS Primary numbers.

5.2 Classifications

5.2.1 Overall Classification Comparisons

Does the liberal application of CS, IT, and ISci indicate a general lack of understanding of what research in the computing disciplines entails? It would appear so, especially on the part of the authors of those theses that contained eight classifications, three of which included all three computing disciplines CS, IT, and ISci.

Table 4-2 shows the percentage of the ProQuest Classification/Subject Category universe each computing discipline uses. A higher percentage is indicative of a more interdisciplinary field. The difference in percentage from Primary to Secondary is potentially very important as it shows the differences between potential perceptions when a computing discipline is placed in the Primary versus Secondary spot. As Primary the majority of the computing disciplines are more focused, sharing fewer, sometimes many fewer, Classifications. But, when Secondary the number of shared Classifications grows tremendously, except interestingly, for CE—which has more co-Classifications as Primary. And, it seems likely that this is the case for CE because CS, IS, and IT can all simply connote “computing” to some extent. But, CE, with “engineering” may denote a more rigourous field that authors may be disinclined to include lest their works lose credibility. Of course, as the term “engineering” is being diluted by simply attaching the word “engineer” to all sorts of disciplines, there may be a future trend where CE becomes much like the rest of the computing disciplines—focused as the Primary but included by a broad spectrum of other disciplines when Secondary.

CS, as it has so few non-Solo Primary theses, will naturally share very few of the universe of ProQuest Classifications. Yet, maybe more surprising is how much its shared universe grows when it is Secondary. ISci also shares this characteristic. Again, is this a result of “resume inflation” on the part of other disciplines that desire co-Classification with CS and ISci? IT’s relatively modest jump in percentage shared from Primary to Secondary may simply indicate that IT itself realizes it is interdisciplinary.

Is the fact that CE shares less of the Classification universe as Secondary may indicate an understanding by the *public* of the term engineering, at least for now (Wilson, 2010). Yet, it is ironic to consider “resume inflation” and term dilution considering the creation of the software “engineering” field.

That the CEP and CES numbers are so close may simply indicate there is no significant difference between CE Primary’s and CE Secondary’s use of the Classification universe. However, dissecting the 44% use of the Classification universe for CE under Any indicates that the overlap of shared Classifications between CE Primary and CE Secondary is only 10%, leaving 16% and 14% unique Classifications, respectively, from the Classification universe in each of CE Primary and CE Secondary. In other words, they choose a separate set of co-Classifications more often than not.

5.2.2 Co-Classification Tables

Table 5-1 below summarizes the percentages that the top 5 and top 20 co-Classifications, or Classifications, make up of all the co-Classifications, or Classifications, for Primary and

Secondary, as well as the Primary when computing is Secondary. Lower numbers indicate a broader range of co-Classifications, and Classifications, since the top 5 and top 20 cover less of total number of co-Classifications, and Classifications, found under each respective column.

Notice that in the Secondary position the percentages are generally similar with a low of 38% for ISS and a high of 48% for CES for the top 5, and a low of 63% and a high of 75% for the top 20. In the Primary position the percentages differ from a top 5 low of 54% for ITP to 95% for CSP and a top 20 low of 78% for ITP and 100% for CSP. For the Primary when Secondary Classifitions the top 5 has a low of 48% for PICS and a high of 66% for PITS. This indicates a larger set of Primary Classifications associating with PICS versus the smaller set for PITS. And for the top 20, a low of 66% for PICS and 85% for PITS, which means that 85% of all Primaries co-Classifying IT are made up of only 20 unique Classifications. This indicates that IT, although interdisciplinary, has a narrower set of likely co-Classifications than all of the other three computing disciplines in this study.

The last grouping for "Primaries when Secondary" shows IT having the most focused top 5 group of Primary Classifications versus the other disciplines. Surprisingly, ISci is the broadest. The latter may be due to a lack of understanding of what INFORMATION SCIENCE entails or ISci provides any easy way to add a computing discipline to a theses. CS demonstrates similar numbers. The lower percentage covered by the top 5 and top 20 represents a broader range of Classifications that claim the computing discipline.

Table 5-1: Percentages of All Co-Classifications and Classifications Covered by Top 5 & 20

	CSP	CEP	ITP	ISP	CSS	CES	ITS	ISS	PCSS	PCES	PITS	PICS
Top 5	95%	79%	54%	90%	45%	48%	42%	38%	48%	54%	66%	43%
Top 20	100%	97%	78%	99.3%	73%	75%	69%	63%	76%	83%	85%	66%

5.2.3 CS Co-Classification Discussion

It seems intuitive to expect to see CE and ELECTRICAL ENGINEERING (EE) share the top two spots in PCSS. As shown from the overall theses numbers in Figure 4-5 there is a steady rise in CE thesis numbers as well as in total CEP numbers coinciding with both a decline in CSP numbers and an increase in CSS numbers. Figure 4-7 corroborates the increase in CSS numbers as partially due to the steady decline in CE Solo numbers, indicating a steady increase in CEP co-classifying with CS. The top two co-Classifications for CE, from 2009 to 2014, are CS and EE. CS overtakes EE in 2012 as the top co-Classification for CE. And, the number of CS co-classified CEP theses jumps from 88 in 2009 to 567 in 2014, and the rise is steady, year-over-year. Combine these data points with Figure 4-6, showing the leveling of number of computing theses in general, and intra-computing discipline movement of theses is likely.

It is almost moot to point out that in Table 4-7 the PCSS column shares little with the co-Classifications in the CSP column with so few in the CSP column. However, both OPERATIONS RESEARCH and SYSTEMS SCIENCE are present, showing some relationship for CS as Primary with these two. The rest of the 268 Classifications in the PCSS column are not found among the co-Classifications in the PCS column. What does this mean? In this case probably not much as there are so few PCS co-Classifications. However, it does say again that CS-focused research is NOT inclined to include ANY of the fields under the PCSS column.

Also, as can be seen, the high percentage of co-classifications with CSA is not indicative of a highly interdisciplinary field as further investigation shows how different CS Primary is from CS Secondary.

5.2.4 **CE Co-Classification Discussion**

As discussed under CS and shown in Table 4-7 and Table 4-8, CE is the primary co-Classifier of CS theses.

The fact that none of the 57 Classifications found under PCES are found among the 117 in CEP is more telling than finding only two matches from 270 among only six for CS. The story this begins to tell is that CEP doesn't appear to agree with PCES regarding what constitutes CE research. A mitigating factor for this discrepancy is that CE is highly biased towards CEP and CES is not the norm. However, the types of Classifications that seek co-Classification with CE is somewhat surprising. The smaller sample size may be a factor but although the top five appear to potentially have reasonable connections to CE, moving further down the PCES list are Classifications such as ASIAN STUDIES and WILDLIFE MANAGEMENT, which are not shown as they are outside of the top 20.

The ASIAN STUDIES thesis' connection to CE appears tenuous as the ACM description of CE involves hardware, software, and communications systems while this thesis involves computer simulated faces to measure reactions. Again, authors are not beholden to any rules on how to classify their theses.

The WILDLIFE MANAGEMENT thesis appears to have a closer connection to CE as the ACM description of CE deals with hardware, software, and communications. This begs the question, "What layer of abstraction differentiates IT and CE?"

5.2.5 **IT Co-Classification Discussion**

In Table 4-9, out of the 95 Classifications under PITS and the 131 co-Classfications used by ITP, the overlap is only three. And the three are MEDICINE (at the 24th position under ITP), HISTORY (88th position under ITP), and SCHOOL ADMINISTRATION (36th under ITP).

This result, along with the results for CE, ISci (below), and CS shows the potential divide between the perceptions of the computing disciplines between *practitioners* and the *public*. Further research is needed to understand why the differences appear so stark but it is interesting to see how MANAGEMENT is never used as a co-Classification for ITP even though it is by far the largest user of IT as a co-Classification under PITS. In fact, the fact that CE is also never used by ITP may indicate a greater understanding of what CE entails on the part of the IT *practitioner*.

Point being, if academia has the *public* perception of IT, then IT and the computing disciplines have a ways to go to educate the public and other disciplines about what research in each computing discipline entails.

These results do not indicate that ITP is always created by a practitioner. Looking at departments later will show that this is likely not the case.

5.2.6 **ISci Co-Classification Discussion**

The fact that there are only 23 distinct co-Classifications for ISP says something about how specific an area of research ISci appears to be at least to ISci practitioners. Maybe more importantly it shows how different the perception ISci may be outside the field of ISci.

The top co-Classification of ISP, COMPUTER SCIENCE, makes up 63% of all ISP co-Classifications. This means ISP authors associate themselves more with CS than any other discipline.

Because ISP only has 23 distinct co-Classifications it may not be as surprising that none of the 192 Classifications under PICS are shared between the two. However, it *is* still surprising as it further illustrates the potential perception gap between practitioners and the public.

There are 268 co-Classifications with ISS. This may simply be due to the fact that there are so many more ISS theses than ISP theses. But, looking at the percentage covered by the top 5 and top 20 shows ISS appears to simply include a broader range of Classifications versus, for example, IT, which has comparable numbers of ITP and ITS but covers a larger percentage of its co-Classification instances with its top 5 and 20. Thus, the lower coverage by ISS of its co-Classifications shows that the practitioner viewpoint tends to be more more focused than the public viewpoint.

It should be noted that inclusion in academia does not automatically remove someone from the public realm and place them as practitioner across all academia. That would be as misguided an assumption as saying that becoming an IT student makes them a practitioner of dance or drawing. Therefore, that other academic areas may have looser interpretations of the four computing disciplines, coming from a public view, is not surprising.

5.2.7 Classification Graphs

The classification graphs from Figure 4-8 to Figure 4-19 show how *other* fields, connected to a computing discipline, are related to each other, not to the subject computing discipline. The graphs show connected fields that are more likely to appear together. Isolated nodes that are connected to high betweenness centrality nodes will tend to appear with those high betweenness centrality nodes.

The findings for each set of graphs show the general agreement the graphs have with the tables of co-Classifications previously discussed.

For each of the *ratio of single to >1 edges* and *ratio of distinct edges to total edges* in the Chapter 4 graph findings, higher ratios indicate more co-Classifications are present.

As stated above, if there are not at least three Classifications associated with a thesis the Classification attached to the computing discipline will not appear in the graph. This is by design as it is the connections between Classifications that is being investigated. It is desired to see, "What shows up with what?" A large presence in the graphs shows how often theses in a given computing discipline have more than just two Classifications.

5.2.8 CS Graphs Discussion

The reason for the CSS graph being as different as it is from the CSP graph is unknown, since the difference should be a single edge. It could be a result of the greedy optimization algorithm used by Wakita-Tsurumi and the inclusion of the single edge setting the greedy path differently. It is hard to imagine the single edge CSP brings would alter the groups so much, but that appears to be the case.

5.2.9 CE Graphs Discussion

The CE graphs from Figure 4-11 through Figure 4-13 are expected given the make up of CEA, CEP, and CES data from previous findings and discussions. It does prove interesting to see which nodes in CES will likely appear together and not seeing EE and CS together for CES is a surprise.

Because only connection weights of 10 or greater are shown only the strongest connections are shown. Thus, when CS is shown with EE with as heavy a connecting line as it has it shows the commonality of the two in connection with CE and each other.

Looking to the far right for CEA and CEP shows the frequent inclusion of WEB STUDIES, ISci, OPERATIONS RESEARCH, MULTIMEDIA COMMUNICATIONS, and BIOINFORMATICS in connection with CS in the CEA and CEP graphs.

CES shows a rather different picture with CS and EE no longer directly connected but rather CS aligned with MATHEMATICS, APPLIED MATHEMATICS, and STATISTICS and CE aligned with APPLIED MATHEMATICS and STATISTICS.

5.2.10 **IT Graphs Discussion**

An expected feature in the ITA graph given the fact that INFORMATION SYSTEMS no longer exists in ProQuest is the strong presence of MANAGEMENT and its association with ORGANIZATIONAL BEHAVIOR. These two disciplines, normally associated with Information Systems, according to the ACM description of Information Systems, use IT to classify their graduate work in ProQuest. And again, this is unfortunate given that Information Systems is a well recognized computing discipline. Not having INFORMATION SYSTEMS as a Classification potentially obfuscates the research in other computing disciplines, especially IT.

The strong CS, ISci, and CE connections, especially between CS and ISci, indicate that the presence of IT will often mean a heavy presence in the above three. The weakening of the connection between CS and CE in the ITP graph shows that IT does not often claim CE as a co-Classification. However, the strong return of the connections between CS and CE in the ITS graph indicates if a thesis includes IT, and it has CE, it will also likely have CS. And therefore, either CS or CE will often claim IT as a Secondary. And since it was shown CS rarely includes any other discipline when it is Primary, in PITS, it is likely CE claims IT.

The disconnecting of the education group in the ITS graph indicates that any connections that may exist between the education group and the other visible groups such as INFORMATION SCIENCE, WEB STUDIES, and MANAGEMENT are now slighter.

Given the distinctness and specificity of technology and education it would not be surprising to see a top-level discipline devoted to technology and education. There exist programs today that focus on this very topic. For example, there exist now programs such as:

- Technology and Engineering Education (BYU College of Engineering and Technology, *Technology and Engineering Education*)
- Instructional Psychology & Technology (BYU School of Education, *Instructional Psychology & Technology*)
- Master's Degree: Learning, Media and Technology Concentration (UMassAmherst, *Master's Degree: Learning, Media and Technology Concentration*)

At the same time, are some of these disciplines better suited to remain with education colleges? Which brings about the question of why Information Systems became an ACM recognized discipline when it's simply the application of technology in the business and organizational space? It may simply be because that's where the money is.

5.2.11 **ISci Graph Discussion**

The heavy connection between CS and ARTIFICIAL INTELLIGENCE in the ISP graph shows they are likely present *together* in most ISP theses.

A strong connection from MANAGEMENT to IT doesn't mean a strong connection from MANAGEMENT to HEALTH CARE MANAGEMENT unless they're directly connected. Their only relationship comes from sharing IT (and ISci), or a weaker relationship not shown due to the 10 edge weight cut-off.

Is ISci used as a label and not a discipline by non-ISci versus ISci? Can the same argument be made for IT?

A key point would be whether IT, CS, or CE are also treated as “classifiers” and not “disciplines.” The data shows that this is not the case with CE, generally. As shown above, there is a distinct set of classifications that shares names with actual disciplines that co-classify with CE (EE, CS, MATHEMATICS, STATISTICS, etc.). Also seen is that CS, when primary, tends to treat itself like a discipline and treats the other classifications as disciplines as well; while, on the other hand, other disciplines appear more inclined to take a looser interpretation of CS and include their work with CS, or include CS with their work.

Before Information Technology appeared in 2009 that Information Systems had a large number of both primarily and secondarily classified theses. Information Science may best be looked upon as a label. But, from the earlier numbers it can be shown that as authors begin to understand what Information Science entails its numbers may increase. Time will tell if the uptick from 2014 continues in the upcoming years.

HEALTH CARE MANAGEMENT is understandably present as information management in the health care industry is an important topic recently and has been for at least the last two decades. (Gross and Cassidy, 1992; Hosny, 2016) And this brings up another point along the lines of ISys and technology in education. Does ISys with its organizational focus become the blanket ACM discipline for technology in healthcare—ISys covering its use, IT implementing the infrastructure, CS/SE building the software, and CE building the hardware ?

5.3 Departments

5.3.1 CS Department Discussion

Departments add depth and context to the tables and figures in this research. They shed greater light on both the Classifications and keywords used, as well as broaden the understanding of which fields use computing or claim it. A high number of distinct departments implies a large variety of departments that classify with the respective computing discipline—although many of those may still be technology-related departments by various names as shown in Table 2-1.

Table 4-18 shows that Computer Science is, not surprisingly, the main department for CS Primary theses. But surprisingly, the main department for CS Secondary theses is also Computer Science. For CS Secondary, however, there are also many EE and CE theses, which hints at the commonality of these three fields, and which is confirmed by the graph in Figure 4-10. This not surprising given Table 4-7 and Table 4-8, showing CE is one of the top departments co-classifying with CS, and showing that CE very often co-classifices with CS.

The differences happen at the tail end of CS Secondary departments. CS Primary has 221 distinct Departments claiming CS. CS Secondary, on the other hand, has 705, as shown in Table 4-17. This implies there may well be a misunderstanding as to what should constitute CS graduate research. Some of the Departments in the long tail of Departments self-classifying as CS research include: Packaging, World Cultures, Dance, and Visual Arts. And many of these are single instances, where the Department only occurs once in the test data. There are 436 single instances for CS from 2009 to 2014.

It is fairly easy to imagine how computing could be used in any of the above: from package optimization, world cultures using compute, and variations of digital art.

A glance at the actual abstracts for the above four theses reveals the Packaging thesis involves RFID with SCIENCE EDUCATION as the Primary Classification. The World Cultures thesis involves archeology and 3D technologies with ARCHAEOLOGY as the Primary. The Dance thesis discusses Dance ontologies with DANCE as the Primary. And finally, the Visual Arts Department thesis has FINE ARTS as a Primary and involves using video to capture reactions to art exhibits and incorporate those reactions back again as part of the exhibit.

Some example theses pulled from Computer Science Departments where CS is the Primary include topics such as "Rate adaption (RA) … for energy efficiency in 802.11n MIMO NICs," and "text based similarity metrics that characterize the relation between semantic web graphs…."

Some samples from Computer Science Departments where CS is the Secondary include topics such as Direct-To-Consumer genetic tests and their interpretation with BIOINFORMATICS as the Primary (which frankly seems to have very little to do with CS), and synthetic biology with BIOLOGY as the Primary.

These four theses from non-technology departments and those from Computer Science departments would seem to expose the differences in research *in* CS versus research that *uses* some portion of CS or something that is currently connected to CS, like ontology. Theses coming from Computer Science departments that were not CSP appear more similar to those coming from non-CS and non-Technology departments: the connection to CS is 'technology' or some form of compute, but it does not seem to *be* CS but rather a *use* of CS at best.

And this *research IN*, versus *research USING* may be the common plight of all academic computing research. The difficulty would be determining when a *USE* can be called *IN*, especially considering computing, by necessity, becomes systems of abstraction. Figure 2-4 shows how one layer *uses* the work of another layer and creates for other layers.

The long tail of computing departments may potentially speak to the absence of a problem. If the occurrence of random co-Classifications with CS happens over such a broad range of departments, usually happening only once, it may simply indicate noise.

Incorporation of computing by other fields is happening, and even by Humanities and Arts majors, as has been shown. (Georgia Tech, *Digital Media at Georgia Tech*) Whether this is a good or bad thing, whether it leads to more work, and a greater variety of work that extends the uses of computing and leads to more innovation, or becomes discipline dilution (Wilson, 2010; LinkedIn, *Marketing Engineer Jobs*; Automattic, *Happiness Engineer*) and leads to a glut of technology practitioners—or a large number of inadequately trained ones—is yet to be seen.

5.3.2 CE Department Discussion

The Departments for CE may not be surprising given how many more CEP theses there are than CES theses. They all come from similar technical fields. However, in the eighth and ninth positions for CES Mathematics and Music appear. Do Mathematics and Music really do CE research? Apparently they do with audio signal processing, creation of music via computers, and classifical division algorithms.

5.3.3 IT Department Discussion

Arguably, the most telling of departmental differences comes from IT. The top five departments for ITP and ITS (not including the absence of a department) share only one department: School of Business and Technology. The fact that this is second for ITP may well be a reflection of the lack of an INFORMATION SYSTEMS classification in ProQuest. And, ISci, despite sharing the same classification code, does not appear to have the same focus as ISys. The latter is already shown to be more business and organizationally oriented (Association for

Computing Machinery, *Information Systems*) and ProQuest may have decided that IT is where they would prefer authors to classify business and organizational theses with technology aspects. In turn, ISci becomes a place to classify, as the ProQuest related classifications indicate, a more social, and possibly less technical, body of work.

The top five departments for ITS contain either the word 'business' or 'management.' Despite one overlap, the top five for ITP starts with Computer Science and ends with Cybersecurity.

There may be something to other disciplines classifying their own thesis as primarily a computing discipline thesis, and then adding their own classification afterwards. So, looking at all primary position theses and the departments from which they come, if the field is available, when the department field is present the department can vary from *Business* to *Urban & Public Affairs*.

There are 75 total distinct listed Departments for IT as Solo from 2009-2014.

The smaller number is not surprising considering there are only 304 IT Solo theses from 2009-2014, while there are 2019 ITP and 3748 ITA. The point being the differences between ITP, ITS, and ITA are likely less pronounced because of the Primary placement by many non-IT, and non-computing disciplines of IT into the main position while submitting theses.

Of the 75 Departments for IT Solo, 58 (77%) appear to be from some sort of computing related department, 11 of the 17 that are not computing related are business related, 4 are for graduate studies, and the last two are Civil Engineering and Biostatistics.

The criteria for determining whether a department is computing related is whether it contains words such as:

- Technology/ies
- Information/Informatics
- Systems
- Computer science
- Cyber
- Computing
- Electronic/Electrical
- Digital
- Industrial engineering
- Interactive
- Telecommunications

Of the 381 listed departments for IT PRIMARY, 175 (46%) are computing related, 204 are not, and 2 (Imaging Science and Library Science) are questionable.

Department names can be inconsistent, with *Communication, Culture and Technology* showing up as *Communication, Culture & Technology*.

The biggest difference then in keywords may then come from the Solo column versus any of the other columns, even Primary, as there is enough dilution from other disciplines/classifications that keywords more specific to the computing discipline are masked.

5.3.4 **ISci Department Discussion**

ISci may be a confused discipline, at least in terms of ProQuest Classification. Looking at the Departments, there appear to be those that seek to classify it as ProQuest desires, under a

LIBRARY SCIENCE/COMMUNICATIONS umbrella, while there are others that appear to look upon it as more of a hard science, more likely to include it as a part of CS.

ProQuest aligns ISci with four communications classifications: COMMUNICATION, JOURNALISM, MASS COMMUNICATION, and TECHNICAL COMMUNICATION, along with LIBRARY SCIENCE and WEB STUDIES. The last, WEB STUDIES, appears to be high-level perspective rather than the nitty-gritty of protocols and packets.

The Department data shows that the fifth most popular Department for ISS is Communications. Yet, most authors continue to associate ISci as a technical term with the top Departments for ISP or ISS starting with Computer Science and also including Information Systems.

5.4 Keywords

Why *applied sciences* is present as a keyword in almost every thesis across all four disciplines is unknown. Also, many of the keywords are regular, much like the Classifications. It almost appears that authors are initially choosing from a set of keywords, rather than openly choosing from an open corpus of available words. Of course, there are common and accepted terms to describe some research areas, like ‘machine learning,’ which is not a Classification choice.

The top 20 keywords in Table 4-23 may be the best indicator of what is “hot” in computing research in the six year span from 2009 to 2014. That *machine learning* and other AI related fields such as *data mining* and *computer vision* are present should not be a surprise. Nor is the presence of *security* and *privacy* and *cloud computing*. Potentially surprising is the presence of *psychology*. Along with *machine learning language* and *literature and linguistics* are commonly found. *Wireless sensor networks* and the emergence of the Internet of Things is also not a surprise. Given that much of this research started 2-8 years before its publication date shows either research leading industry and public perception or the existence of these fields well before the time period of this research. It could also be authors adding keywords to their works after-the-fact.

Of note is the presence of *information technology* as a keyword. Of the 354 theses with the keyword *information technology* in the 2009-2014 time period, 52 have the word “adoption” in the Abstract, with information technology adoption being a common theme. 46 of the 52 have IT as a Classification. 16 have IT as the Primary. It would seem that the keyword is usually used with the Classification as well.

5.4.1 **Keyword Frequency Charts**

Figure 4-22, with CE as Secondary using very few keywords, implies a narrow range of topics are covered by theses that are not principally CE. The same pattern that was shown with CE co-Classifications is apparent in its keywords as well.

What is generally seen in the frequency charts is that most disciplines across all positions have a long tail of various keywords and a large concentration of popular keywords. This makes sense as authors are likely to broadly classify their works and then add finely tuned words that will likely be in the long tail. Thus, the generic words, like *applied sciences*, will be at the far left with more esoteric words and phrases, such as *zeta-image* and *xenobiotic elimination*, in the long tail.

5.4.2 **Top 40 Keyword Tables**

The keywords associated with a thesis are likely to give, more than the Classification fields, Classification field connections, or Departments, the most popular topics for research in each computing discipline, for the time period 2009-2014.

5.4.3 **Top 40 CS Keywords Discussion**

For CS it is understandable that the Primary and Solo columns will be nearly identical. *machine learning*, *data mining*, and *computer vision* are all related. *security*, is generally considered a very "hot" topic currently as witnessed by the number of computer security related programs coming online. And *wireless sensor networks*, as they are related to Big Data and the Internet of things, is also very popular. These top five (not including *applied sciences* in any of these discussion for purposes of ranking) should resonate with anyone working in the computing field at the time of this research. Rounding out the top 14 shows other highly common CS topics: *cloud computing*, *wireless networks*, *distributed systems* (and who hasn't heard of Hadoop today?), *privacy*, *software engineering*, *social networks* (is anyone not on Facebook?), *computer graphics*, *natural language processing* (which is connected to *machine learning*), and *clustering* (which can also be considered part of machine learning).

5.4.4 **Top 40 CE Keywords Discussion**

Not surprisingly CE shares many of its top keywords with CS but somewhat surprising is the presence of *communication and the arts* even with CEP. Moving to CE Solo shows nothing but computing related keywords. The keywords under CE Solo show what is currently popular in the field of CE, such as: *computer architecture*, *energy efficiency*, *fpga*, *wireless sensor* *networks,* and *machine learning*. Because of the high presence of co-Classifications under CEP some generalness is found in that column while CE Solo appears very specific.

5.4.5 Top 40 IT Keywords Discussion

The keywords present under IT show far more generality under IT Secondary than under IT Primary, which still has some generality. And IT Solo is the most specific of them all, which is not surprising. The keywords found under IT Solo should paint the picture of what IT graduate research currently consists of, at least for the practitioner. As Table 4-32 shows it is currently focused on security.

And again, the presence of general keywords under ITP can be due to non-IT departments' presence in IT Primary as shown in the findings under Departments.

5.4.6 Top 40 ISci Keywords Discussion

ISci doesn't seem to have sharp differences in any of its three columns of Primary, Secondary, or Solo. And, this turns out to be accurate as ISci has the highest match rates of Secondary words to Primary (33%) and even to Solo (44%) words of any of the other disciplines. In keyword terms this may make ISci the most general of the disciplines.

The keyword *data science* is found among all the computing disciplines in this research except ISci.

6 CONCLUSION

There is likely a pronounced difference in how practioners and the public view computing research. Practitioners are more likely to intra-classify and include other computing disciplines in their own research, while keeping the general pool of classifications narrow. The public, however, includes a much broader usage of computing, sometimes novel, as computing research. And the spectrum of other disciplines willing to include a computing discipline as part of their resesearch is very wide.

The computing disciplines do overlap. But, the amount of overlap is modest. The computing disciplines have a tendency towards each other. All have a tendency to co-Classify with CS, with CE being the largest co-classifier among the computing disciplines. But, maybe not surprisingly, IT and ISci also tend to have a larger base of connections with each other than IT to CE or ISci to CE. And, whether the theses have the computing discipline as the primary focus of the thesis makes a marked difference in the co-Classifications that are associated with the theses.

Fein's statement, "the computer thus provides a significant link among various established disciplines as well as those fields of endeavor of intense present interest," (1959) shows a prescient understanding of the role of computing in academia, industry, and society. Along with the public versus practitioner roles, computing research may also be viewed as *research in* versus *research using* computing. The sheer variety of Classifications including a computing discipline demonstrates a large base of *research using* computing. On the other hand, disciplines like CS, with its seeming insular focus and next to no co-Classifications, shows a large amount of *research in* computing. The computing disciplines as they co-Classify with each other is a mix of *research in* and *research using*.

However, *research using* should never be shunned because it is not *research in*. The public *uses* computing, the practitioner *creates* it. And in fact, it is the use, or often the *desired use*, that drives the creation—as demonstrated by the simple history of computing and the need for machine help in making problems involving large amounts of data or calculations tractable. At some level everyone is part of the public as well as being a practitioner.

As stated previously, the inclusion of an author into academia does not automatically remove them from the public realm into the role of practioner of all discplines. A dance student may view research in dance as choreography or performance. A student of computing that co-classifies with dance may include it if there are elements of motion capture. Therefore, the fact that other academic areas may have looser interpretations of the computing disciplines, coming from a public view, is not surprising.

As the numbers of computing theses and dissertations appear to have leveled off in the time period from 2009-2014 it will be interesting to see the potential growth in either the total number of computing related theses or the creation of new computing related disciplines.

It was seen how the introduction of CE, IT, and ISci have given CS and other disciplines more avenues to classify their computing related work. What will the addition of even more computing disciplines do?

From public to practitioner and from its use to its creation, computing is a multi-faceted, integral part of research, academia, and industry. Classifications, keywords, and Departments from a large portion of the academic world has painted a vivid picture of what computing research currently consists of and where it may be headed.

REFERENCES

Accreditation Board for Engineering and Technology. ABET. *ABET Accredited Program Search*. Accessed May 07, 2016. http://main.abet.org/aps/Accreditedprogramsearch.aspx.

Agarwal, A. *6.002 Circuits and Electronics, Spring 2007*. Massachusetts Institute of Technology: MIT OpenCourseWare. 2007. Accessed May 23, 2016. http://ocw.mit.edu.

AntiqueTech. *The ENIAC vs. The Cell Phone*. 2013. Accessed May 07, 2016. http://www.antiquetech.com/?page_id=1438.

Association of Research Libraries. ARL. *About | Association of Research Libraries*. Accessed May 07, 2016. http://www.arl.org/about#.VyqwTPkrJhE.

Association of Research Libraries. ARL. *List of ARL Members*. Accessed May 07, 2016. http://www.arl.org/membership/list-of-arl-members#P.

Association for Computing Machinery. ACM. *About the ACM Organization.* Accessed May 21, 2016. http://www.acm.org/about-acm/about-the-acm-organization.

Association for Computing Machinery. ACM. *About SIGS.* Accessed May 21, 2016. http://www.acm.org/special-interest-groups/about-special-interest-groups

Association for Computing Machinery. ACM. *ACM History.* Accessed May 21, 2016. http://www.acm.org/about-acm/acm-history.

Association for Computing Machinery. ACM. *Association for Computing Machinery.* Accessed May 21, 2016. http://www.acm.org.

Association for Computing Machinery. ACM. *Computing Disciplines & Majors*. Accessed May 07, 2016. http://computingcareers.acm.org/?page_id=6.

Association for Computing Machinery. ACM. *Curricula Recommendations*. Accessed May 09, 2016. http://www.acm.org/education/curricula-recommendations.

Automattic. *Happiness Engineer*. Accessed May 11, 2016. https://automattic.com/work-with-us/happiness-engineer/

Cipra, B. A. *The Best of the 20th Century: Editors Name Top 10 Algorithms*. SIAM News, Volume 33, Number 4. May 16, 2000. Accessed May 23, 2016. https://www.siam.org/pdf/news/637.pdf

Clauset, A. and Newman, M. E. J. and Christopher, M. *Finding Community Structure in Very Large Networks*. Institute for Infrastructure Surety. University of New Mexico. 2004. Accessed May 09, 2016. http://ece-research.unm.edu/ifis/papers/community-moore.pdf

Cole, C. J., *Research in Information Technology: Analysis of Existing Graduate Research* (2009). All Theses and Dissertations. Paper 1935. http://scholarsarchive.byu.edu/etd/1935

Computer History Museum. *1933 | Timeline of Computing History.* 2016. Accessed May 07, 2016. http://www.computerhistory.org/timeline/1933/.

Computer History Museum. *1944 | Timeline of Computing History.* 2016.Accessed May 07, 2016. http://www.computerhistory.org/timeline/1944/.

Computer History Museum. *1960 | Timeline of Computing History.* 2016. Accessed May 07, 2016. http://www.computerhistory.org/timeline/1960/.

David O. McKay School of Education. *Instructional Psychology & Technology*. Brigham Young University. 2015. Accessed May 09, 2016. http://education.byu.edu/ipt

Dhar, V. *Data Science and Prediction*. Communications of the ACM, Vol. 56 No. 12, Pages 64-73. 2013. Accessed May 21, 2016. http://dx.doi.org/10.1145/2500499.

Dongarra, J. *Visit to the National University for Defense Technology Changsha, China*. Oak Ridge National Laboratory. University of Tennessee. June 3, 2013. http://www.netlib.org/utk/people/JackDongarra/PAPERS/tianhe-2-dongarra-report.pdf

Falk, J. *Early Evolution of the Modern Calculator. Part 2*: The Modern Era. Last Modified November 14, 2014. Accessed May 21, 2016. http://metastudies.net/pmwiki/pmwiki.php?n=Site.TheModernEpochAndTheEmergenceOfTheModernCalculator#pascaline

Fein, L. *Oral History Interview with Louis Fein*. By Pamela McCorduck. Charles Babbage Institute. University of Minnesota Digital Conservancy. 1984. Accessed May 23, 2016. http://purl.umn.edu/107284.

Fein, L. *The Role of the University in Computers, Data Processing, and Related Fields*. Communications of the ACM. Vol. 2 No. Pages 7-14. ACM. New York. USA. 9 Sept. 1959. http://dx.doi.org/10.1145/368424.368427

Georgia Tech. *MS in Digital Media - Digital Media at Georgia Tech*. 2016. Accessed May 07, 2016. http://dm.lmc.gatech.edu/program/ms-program/.

Gross, M. and Cassidy, B. *Information Technology and Today's Health Care Management*. National Center for Biotechnology Information. 1992. Accessed May 09, 2016. http://www.ncbi.nlm.nih.gov/pubmed/10122865.

Hansen, D., Shneiderman, B., and Smith, M.. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann Publishers, Burlington, 2011.

Hosny, A., Vera-Licona, P., Laubenbacher, R,. and Favre, T. *AlgoRun: a Docker-Based Packaging System for Platform-Agnostic Implemented Algorithms*. National Center for Biotechnology Information. 2016. Accessed May 09, 2016. http://www.ncbi.nlm.nih.gov/pubmed/27153722

Howe, B. *Introduction to Data Science*. University of Washington. Coursera. 2014. Accessed May 23, 2016. https://www.coursera.org/course/datasci

Hoyles, C., Professor Dame, Sabin, M., and Macey, J. *Why Does Mathematics Matter in the Digital Arts?* Accessed May 07, 2017. http://www.ima.org.uk/_db/_documents/MT June 14 Why Does Mathematics Matter in the Digital Arts.pdf.

Hu, C. P., Hu, J. M., Deng, S. L.., and Liu, Y. 2013. *A Coword Analysis of Library and Information Science in China*. Scientometrics, 97(2), pp 369-382.

Ira A. Fulton College Engineering and Technology. *Technology and Engineering Education*. Brigham Young University. 2012. Accessed May 09, 2016. http://tee.byu.edu/

InfoBrief. *Time to Degree of U.S. Research Doctorate Recipients*. National Science Foundation. NSF 06-312. March 2006. Accessed May 24, 2016. http://www.nsf.gov/statistics/infbrief/nsf06312/nsf06312.pdf

Johnson, R. C. *Microsoft, Google Beat Humans at Image Recognition*. EE Times. 2015. Accessed May 07, 2016. http://www.eetimes.com/document.asp?doc_id=1325712.

Kaser, D. *ProQuest Goes Global*. Information Today. 2014. Accessed May 07, 2016. http://www.infotoday.com/it/apr14/Kaser--ProQuest-Goes-Global.shtml.

Kelsky, K. *The Perils of Publishing Your Dissertation Online*. The Professor Is In. August 24, 2011. Accessed May 21, 2016. http://theprofessorisin.com/2011/08/24/the-perils-of-publishing-your-dissertation-online/.

Kim, S., Hansen, D., and Helps, C. R. G. *Information Technology Research in the Academy: Insights from Theses and Dissertations*. 2015. SIGITE/RIIT Conference'15, September 30–October 4, 2015, Chicago, Illinois, United States.

LinkedIn. *Marketing Engineer Jobs*. Accessed May 11, 2016. https://www.linkedin.com/jobs/marketing-engineer-jobs

LabWrite Resources. *Tables vs. Graphs*. NC State University. 2004. Accessed May 09, 2016. https://www.ncsu.edu/labwrite/res/tablevsgraph/res-tablevsgraph.html

LaSalle, D. and Karypis, G. *Multi-Threaded Modularity Based Graph Clustering using the Multilevel Paradigm*. Department of Computer Science & Engineering. University of Minnesota. 2014. Accessed May 09, 2016. http://www-users.cs.umn.edu/~lasalle/nerstrand/publications/nerstrand-journal-r3.pdf

Lai, E."*The '640K' Quote Won't Go Away -- but Did Gates Really Say It?* Computerworld. 2008. Accessed May 07, 2016. http://www.computerworld.com/article/2534312/operating-systems/the--640k--quote-won-t-go-away----but-did-gates-really-say-it-.html.

Lee, B. and Jeong, Y. I. 2008. *Mapping Korea's National R&D Domain of Robot Technology by Using the Co-Word Analysis*. Scientometrics, 77(1), pp 3-19.

Lee, P. C. and Su, H. N. 2010. *Investigating the Structure of Regional Innovation System Research Through Keyword Cooccurrence and Social Network Analysis*. Innovation : Management, Policy & Practice 12(1):26-40.

Library of Congress. *News from the Library of Congress*. *Library and Copyright Office Sign Landmark Agreement with UMI*. 1999. Accessed May 07, 2016. http://www.loc.gov/today/pr/1999/99-007.html.

National Center for Education Statistics. NCES. *Introduction to the Classification of Instructional Programs: 2010 Edition (CIP-2010).* Accessed May 21, 2016. http://nces.ed.gov/ipeds/cipcode/Files/Introduction_CIP2010.pdf

New York University. NYU. *Claude Shannon*. Accessed May 23, 2016. https://www.nyu.edu/pages/linguistics/courses/v610003/shan.html.

Newman, M. E. J. *Modularity and Community Structure in Networks*. Department of Physics and Center for the Study of Complex Systems. University of Michigan. Feb. 17, 2006. Accessed May 24, 2016. http://arxiv.org/pdf/physics/0602124.pdf

Parker, C. B. *Great Recession Spurred Student Interest in Higher Education, Stanford Expert Says*. Stanford News. 2015. Accessed May 07, 2016. https://news.stanford.edu/2015/03/06/higher-ed-hoxby-030615/.

O'Shaughnessy, L. *12 Reasons Not to Get a PhD*. CBSNews. 2012. Accessed May 07, 2016. http://www.cbsnews.com/news/12-reasons-not-to-get-a-phd/.

ProQuest. *Dissertations Abstracts International*. Accessed May 07, 2016 http://www.proquest.com/products-services/dissertations/Dissertations-Abstract-International.html.

ProQuest. *Dissertations and Theses Dissemination and Ordering*. Accessed May 07, 2016 from ProQuest: http://www.proquest.com/products-services/dissertations/.

ProQuest. *History & Milestones*. Accessed May 07, 2016 from ProQuest: http://www.proquest.com/about/history-milestones/.

ProQuest. *ProQuest 2015-2016 Publishing Agreement*. Accessed Apr 19, 2016. http://media2.proquest.com/documents/umi_agreement_papersub.pdf.

ProQuest. *ProQuest Subject Categories – 2015-2016 Academic Year*. Accessed May21, 2016. http*://media2.proquest.com/documents/Subject+Categories+2015-2016+Academic+Year.pdf*

ProQuest. *ProQuest Dissertations & Theses Global*. Accessed May 07, 2016 from ProQuest: http://www.proquest.com/products-services/pqdtglobal.html.

Rice, J. R. and Rosen, S. *History of the Department of Computer Sciences at Purdue University*. Computer Sciences History. Purdue University. Accessed May 21, 2016. https://www.cs.purdue.edu/history/history.html.

SIGITE. *IT Discipline*. Accessed May 07, 2016. http://www.sigite.org/?page_id=22.

SIGKDD. *About SIGKDD.* Accessed May 21, 2016. http://www.kdd.org/about

UMassAmherst. *Master's Degree: Learning, Media and Technology Concentration*. 2016. Accessed May 07, 2016. https://www.umass.edu/education/departments/tecs/learning-media-tech.

Wakita, K. and Tsurumi, T. (2007) Finding Community Structure in Mega-scale Social Networks. arXiv:cs/0702048

Wayne, K. and Sedgewick, R. *Algorithms Part I*. Princeton University. Coursera. 2016. Accessed May 23, 2016. https://www.coursera.org/course/algs4partI

White, S. K. *8 Universities at the Forefront of Big Data*. CIO. 2016. Accessed May 07, 2016. http://www.cio.com/article/3058841/it-skills-training/8-universities-at-the-forefront-of-big-data.html#slide2.

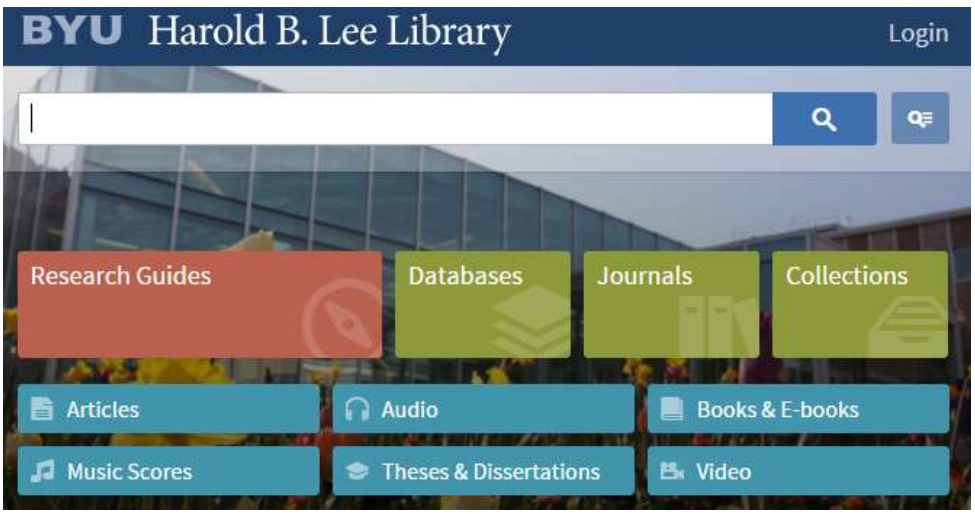
Wilson, A. *You Say Engineer, I Say Gas Fitter – Plumbers Protest at BAE Remarks*. The Telegraph. 2010. Accessed May 09, 2016. http://www.telegraph.co.uk/finance/newsbysector/industry/engineering/7347635/You-say-engineer-I-say-gas-fitter-plumbers-protest-at-BAE-remarks.html.

APPENDICES

APPENDIX A. DOWNLOADING PROQUEST DATA

Go to the Brigham Young University Harold B. Lee Library site

(http://lib.byu.edu)



From the homepage (as of Jan. 17, 2015), click on Theses & Dissertations

(http://lib.byu.edu/theses-and-dissertations/)

Under the Databases section, click on Dissertations & Theses (ProQuest)

(http://dbs.lib.byu.edu/dissertations)

If not logged in as an authorized user will be redirected to login with BYU NetId and password.

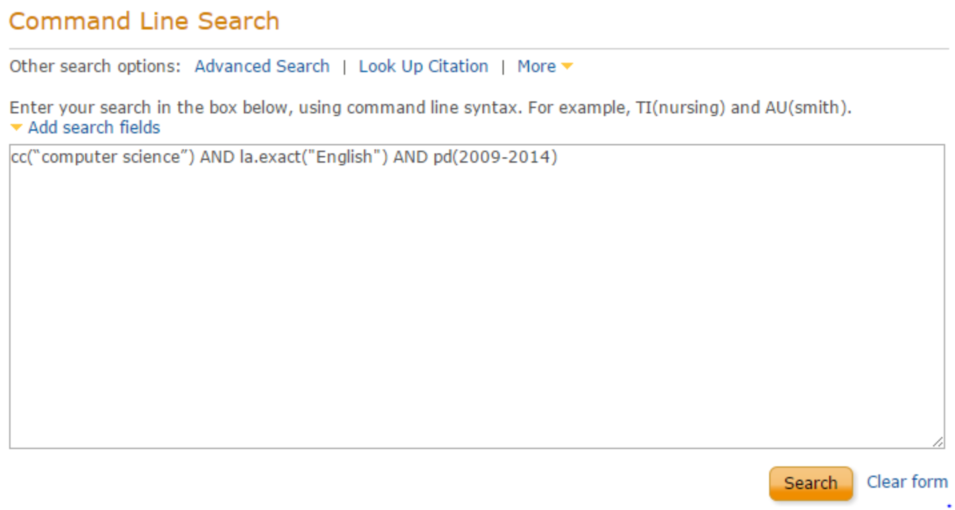
After login will be redirected to the ProQuest Dissertations & Global page

(http://search.proquest.com.erl.lib.byu.edu/pqdtglobal/index?accountid=4488)

Click on the Advanced Search link

Will bring up:

Click on *Command Line*



Enter the following in the search text area:

cc("information technology") AND la.exact("English") AND pd(2009)

Explanation of codes:

cc: Classification

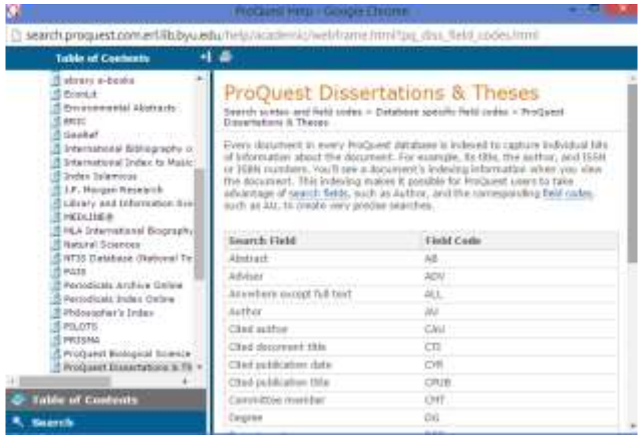
la: Language

pd: Publication date

Note that "cl" is the code used for "Classification" when switching from the Advanced Search to the Command Line Search after building a query in Advanced Search.

For further field codes, from the Command Line page click *Field codes*.

Note that this does not show all the available codes.



In the Table of Contents, click *Common field codes* and choose *Command Line Search field codes*.

As Information Technology was added as a ProQuest classification in 2009 and ProQuest did not allow for month-level granularity as of Jan. 17, 2015 for theses and dissertations (although the 'pd' field should allow it), the date range for theses and dissertation extraction is from 2009 till 2014.

Although "Information Technology" was added as a ProQuest classification in 2009 it appears that previously published works can be reclassified, or have classifications added. The earliest ProQuest publication date with "Information Technology" as a classification is now 1956. Although possibly informative to discover the types of theses and dissertations that have now been reclassified to include IT, as the number of relatively small (128 from 1956-2008), this will only be broached in passing. Note that this also means that the exact number of IT theses and dissertations may change, likely increase, as authors or publishers determine that a previous publication is in fact appropriately classified as IT.

One year at a time is downloaded because ProQuest sets a max limit of 4000 results. While this isn't a problem with IT, it is a problem with "computer science" which returns 20170 results for the same 2009-2014 period.

3483 Results

Under *Sort results by* choose *Publication date (oldest first)*.

To download

NOTE: ProQuest has modified their number of theses metadata that can be downloaded at one time. The previous limit was 500, now it is set to 50. It is unknown whether the authors heavy downloading of metadata was a factor in causing ProQuest to change this function, or whether this limitation only applies to the author.

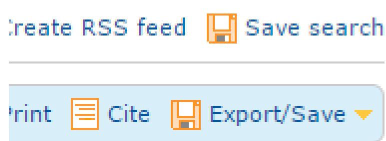
Scroll down to the bottom of the results page and change Items per page from 20 to 50.

Sort a sort order of “Oldest First.” This will allow resuming downloads from a known spot versus having to figure out which theses have and have not been downloaded, if a whole year cannot be downloaded in a single session.

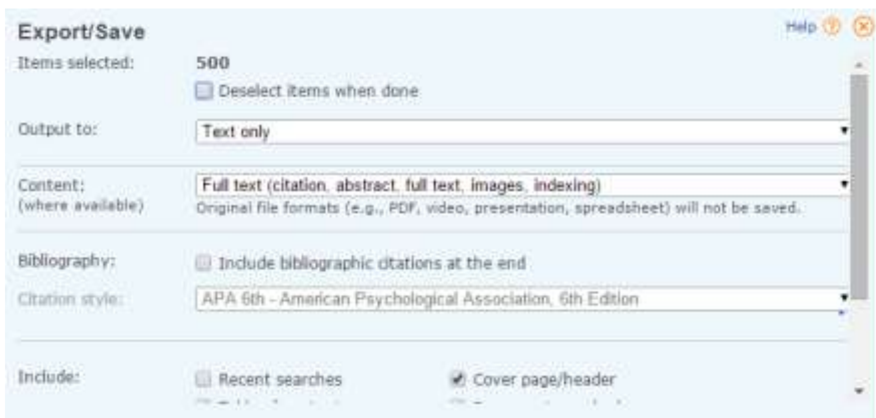
From the top of the results page click Select 1-50 (or the currently shown set of fifty)

Scroll down to the bottom and click *Next page*

Hover over *Export/Save*



Select Text only



Select *Deselect items when done*

Under *Content*: Leave as is.

Select *Continue*

Documents take a few seconds to be prepped on ProQuest's side.

Repeat the query and download procedure for "computer science" and “computer engineering” and “information science.”

The theses can also be searched for using their classification codes in the cc field.

cc("computer science") AND la.exact("English") AND pd(2009)

ProQuest did not know when computer science was added as a category. Just like IT theses and dissertations were reclassified so it can be that CS theses & dissertations were also reclassified. The first "computer science" classified thesis or dissertation has a publish year of 1951.

ProQuest changing the maximum number of theses for which information can be downloaded to 50 greatly increases the amount of time required to download theses information. As a hint from the author, using a laptop with a touchscreen proved very useful.

APPENDIX B. CREATING NODEXL GRAPHS

Download and Install NodeXL

Download the template for Excel 2014 from:

http://nodexl.codeplex.com/downloads/get/806203

Filename: *NodeXLBasicTemplate2014Setup.exe*

NOTE: NodeXL now has a basic version and a registered/paid version. A student copy of the Pro version was used for this work.

The filename may vary.

This release is compatible with Excel 2007, 2010, and 2013.

http://nodexl.codeplex.com/releases/view/117659

Run the executable to install the template. A compatible Excel version must already be installed. Excel 2010 was used for this study.

To Start NodeXL

From the Windows 8[.1] start menu, type "NodeXL" and choose NodeXL Excel Template.

For Windows 10: Start Excel start menu -> All Apps -> NodeXL Excel Template (if the above method used for Windows 8.x does not work).

Prepare ProQuest Data

The Classification field of ProQuest text extracts are filtered out and cleaned to created *n choose 2*, unique, un-ordered (AB is the same as BA) permutations of all given Classifications from individual theses. These permutations become the edges of our graphs, while the Classifications themselves are the nodes.

For example, if a thesis has a Classification line containing "Information Technology," "Computer Science," and "Education," the combinations will be:

- "Information Technology" "Computer Science"
- "Information Technology" "Education"
- "Computer Science" "Education"

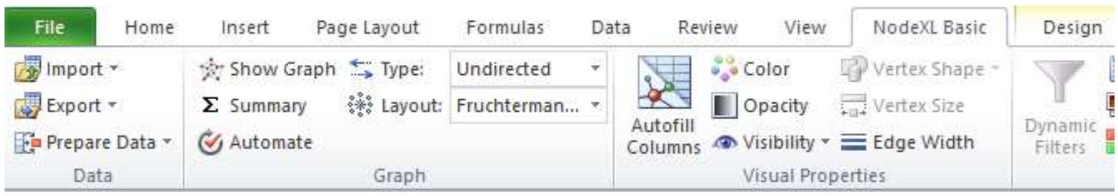
When only one Classification is found it will always be the focal Classification. For the purposes of this paper these singles will be filtered out as it is a common node to all other Classifications. Also, after combinations, any combinations with the focal Classification are also eliminated. Thus from our three combinations above, only "Computer Science" "Education" will be graphed. The goal is to show how theses and dissertations that are connected to the focal Classification are related to each other. Again, it is given that all nodes are connected to the focal Classification. By doing so we can see how academic research perceives each of the computing disciplines.

To create the actual graph:

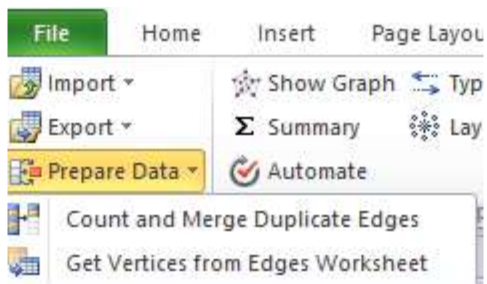
1. Create a new blank worksheet to hold *n choose 2* data.

2. Copy and paste the *n choose 2* data into the Vertex 1 and Vertex 2 columns of the **Edges** worksheet.

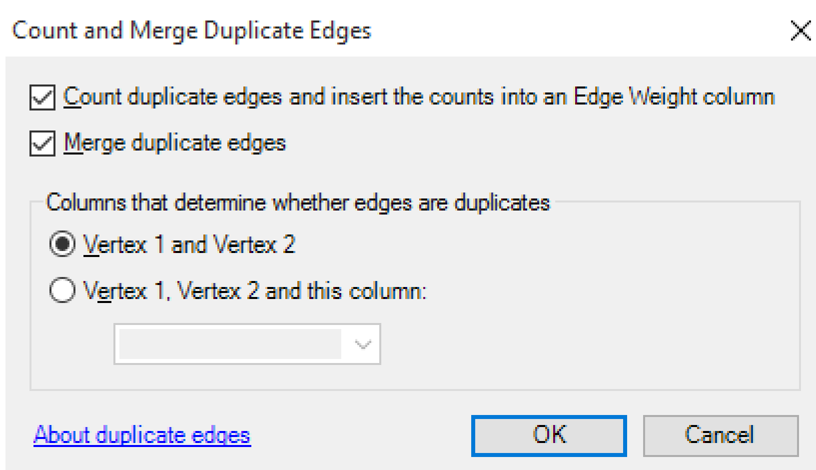
In the Ribbon, under the *NodeXL Basic* tab, in the *Data* panel, select *Prepare Data.*



Select *Count and Merge Duplicate Edges*.



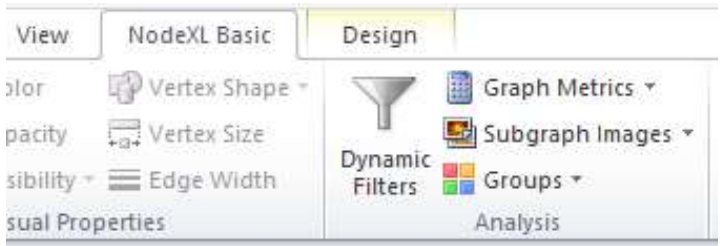
Check the *Merge duplicate edges* box.



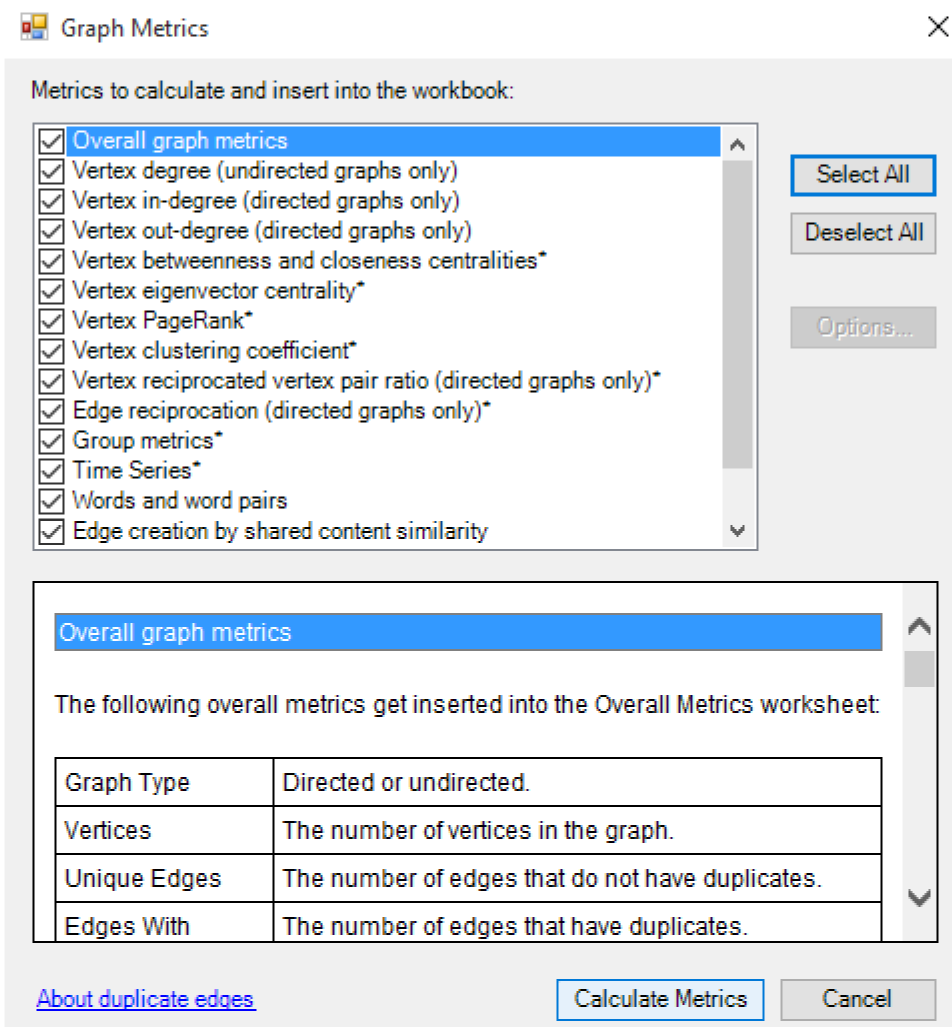
3. Also under *Prepare Data* click *Get the Vertices from Edges Worksheet*

4. Get more information on the graph characteristics via Graph Metrics

5. Under the NodeXL tab, *Analysis* panel -> *Graph Metrics* -> *Select All* -> *Calculate Metrics*

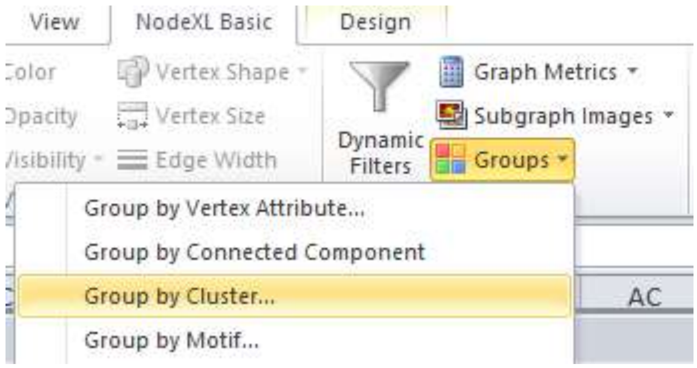


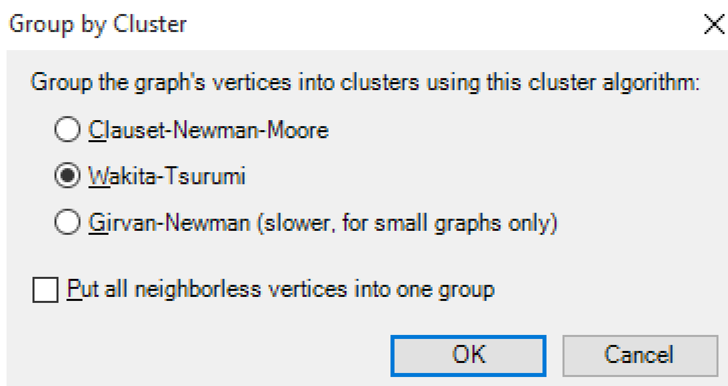
6. Run metrics on all edges, including singles



7. Calculate any Groupings in the graph

8. Under the NodeXL tab, *Analysis* panel -> *Group by Cluster….* -> *Wakita-Tsurumi*





9. Format the graph

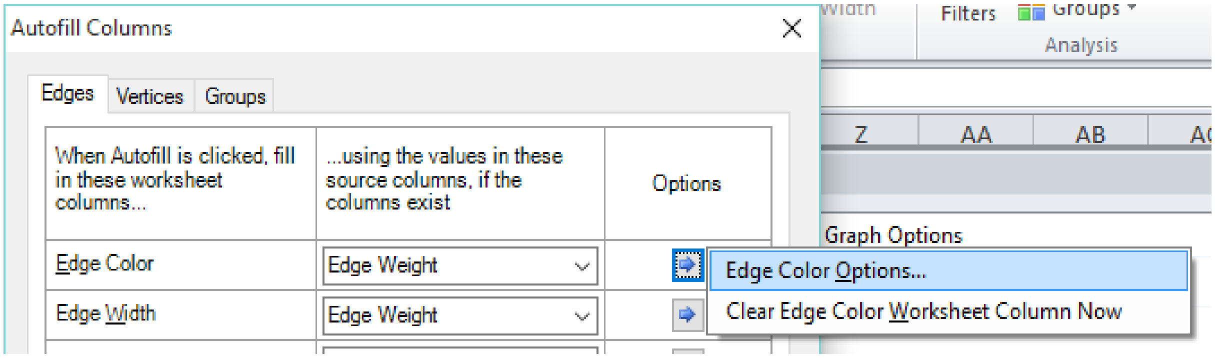
10. To format the graph use the *NodeXL Basic* tab -> *Visual Properties* panel -> click *AutoFill Columns*

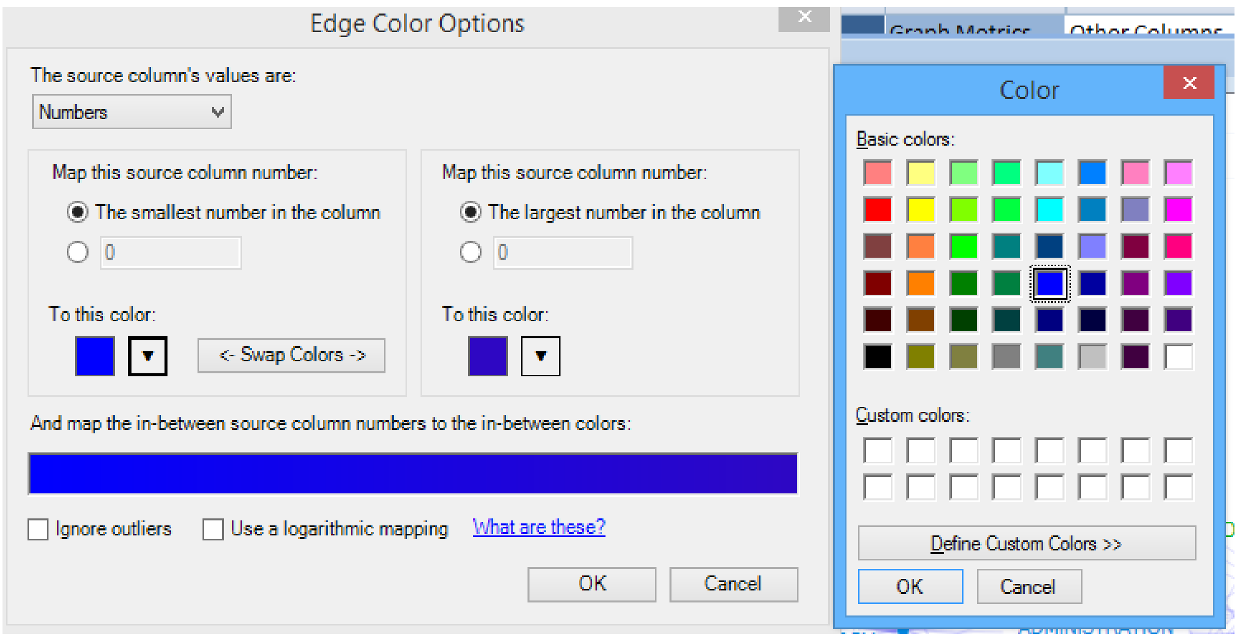
11. Under the Edges tab set:

    a. Edge Width -> Edge Weight

    b. Edge Color -> Edge Weight

        i. Under *Options* for Edge Color set the first color from to dark blue like the second color.





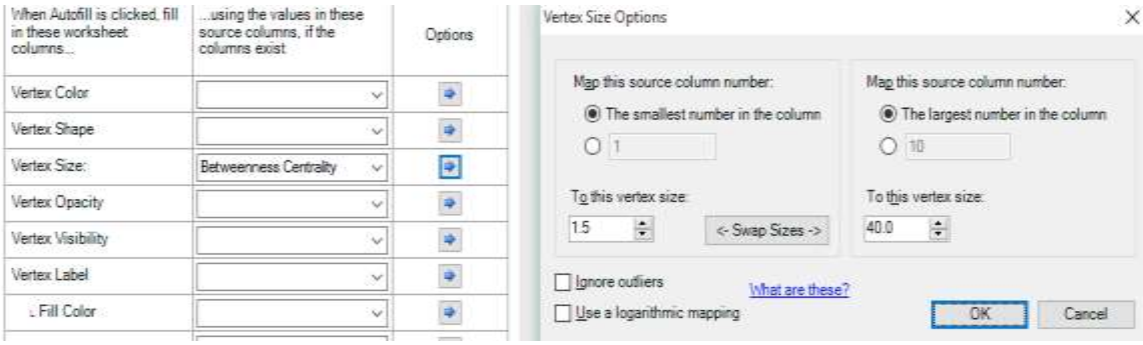
c. Edge Opacity -> Edge Weight

d. Click *AutoFill*

12. Under the Vertices tab set:

a. Vertex Size -> Betweenness Centrality

i. Under *Options* for Betweenness Centrality set the max size to -> 40.0



b. Vertex Label -> Vertex

c. Click *AutoFill*

13. *Close*Return to the **Edges** worksheet

a. Sort by Edge Weight

i. Click the arrow box in cell O2.

ii. Sort Largest to Smallest

14. Showing all nodes and edges can excessively clutter the graph so only edges (and corresponding nodes) with weight >= 10 will be shown.

a. To hide edges with weight < 10 go to column G (Visibility) and the first row with weight of 9.

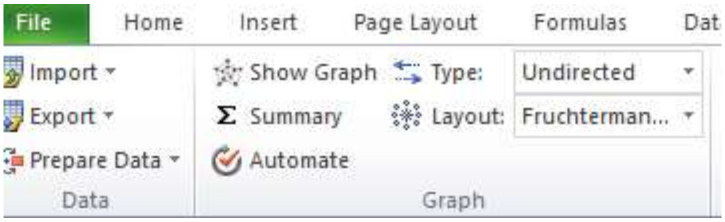
b. Choose "Skip" from the arrow-box choices in that cell and copy and paste "Skip" down to the last edge of the worksheet.

15. Show the Graph

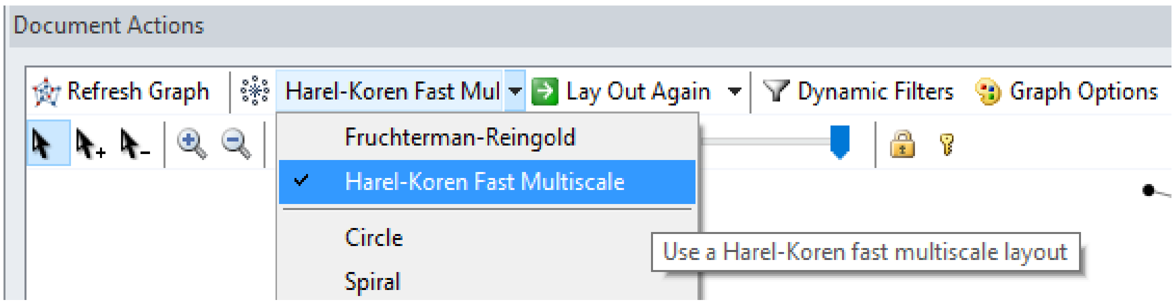
16. Note that a 1080p monitor or higher pixel density monitor is best for showing the graph.

a. A second monitor of the resolution above or greater works best.

17. Under the *Graph* panel of the *NodeXL Basic* tab choose *Show Graph*.

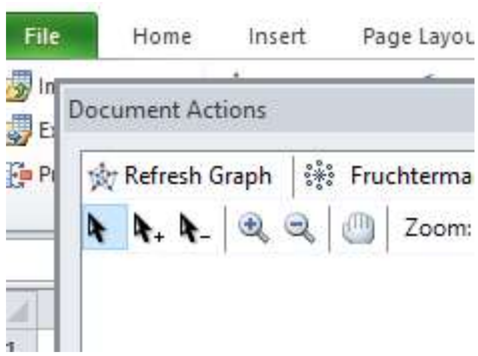


18. Change the algorithm to *Harel-Koren Fast Multiscale*.

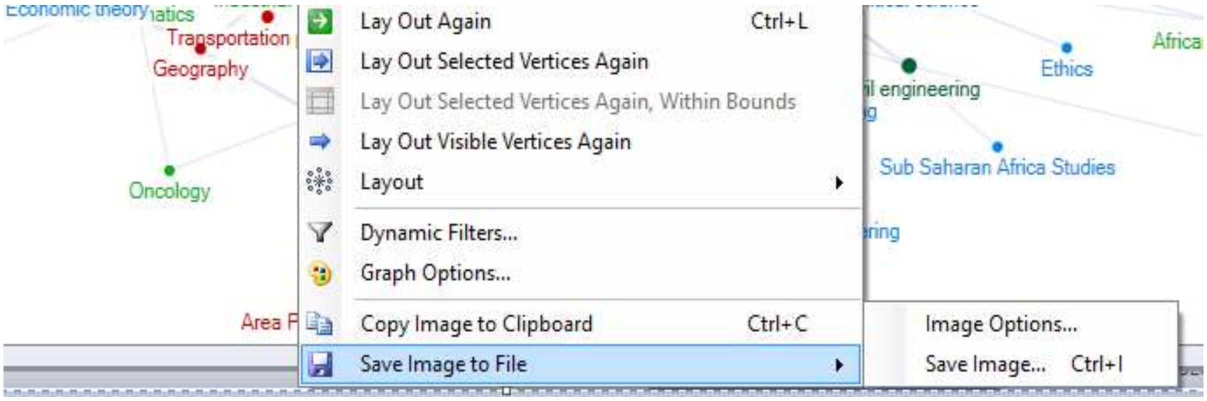


19. Select *Lay Out Again* in the graph window to see an initial graph

    a. *Refresh Graph* may need to be clicked multiple times to get a satisfactory grouping.



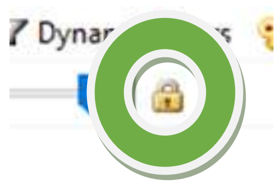
20. To save the image to file, right click on the graph itself -> *Save Image to File* -> *Image Options…* to whatever size is desired (or keep the current size), -> *OK*.



a. Then right click again and -> *Save Image to File* -> *Save Image…*

21. The individual vertices must be pulled and placed at locations that help show the information most vividly. Again, *Refresh Graph* or *Lay Out Again* can be clicked multiple times to get an initial starting point that is closer to the desired final representation.

    a. After a desired layout instance is found select all the nodes and lock them by clicking on the following icon:



while all the desired nodes are selected.

22. *Vertex size* represents the *betweenness centrality* of each classification. This was deemed as most interesting since many vertices have many connections but the bridge spanners with fewer connections happen to be the only links to some obscure topics that may not usually be associated with a computing discipline.

23. *Vertex color* represents the groups the Wakita-Tsurumi algorithm placed the vertices into. The graph is very tightly clustered. Pulling on any highly connected point created a taffy-like effect, stretching out and obscuring other parts of the graph.

24. *Line Width, Color, and Opacity* represent the edge weight, or the number of times the two classifications are mentioned together in a thesis or dissertation. Since there was so much overlap, having line-width alone made the graph just a mass of color. But, only using opacity--even though this made the graph slightly more eye-pleasing--did not adequately represent the number of connections between the most commonly associated classifications. Opacity was harder to discern with smaller lines, as was color. Lines alone also did not do justice to the number connections between highly connected classifications. It did, however, have a cleaner appearance. Color was added to add more visibility to the mass of connections in the center of the graph. And, line width was adjusted to have a maximum width of six so as to not excessively clutter the center of the graph, while still visually representing the density that is present.

25. To show and hide individual groups

    a. The Groups worksheet