Looplessness in networks is linked to trophic coherence

Samuel Johnson[a,b,1] **and Nick S. Jones**[c]

[a]Warwick Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom; [b]Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, United Kingdom; and [c]Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom

Edited by Alan Hastings, University of California, Davis, CA, and approved April 17, 2017 (received for review August 18, 2016)

Many natural, complex systems are remarkably stable thanks to an absence of feedback acting on their elements. When described as networks these exhibit few or no cycles, and associated matrices have small leading eigenvalues. It has been suggested that this architecture can confer advantages to the system as a whole, such as "qualitative stability," but this observation does not in itself explain how a loopless structure might arise. We show here that the number of feedback loops in a network, as well as the eigenvalues of associated matrices, is determined by a structural property called trophic coherence, a measure of how neatly nodes fall into distinct levels. Our theory correctly classifies a variety of networks—including those derived from genes, metabolites, species, neurons, words, computers, and trading nations—into two distinct regimes of high and low feedback and provides a null model to gauge the significance of related magnitudes. Because trophic coherence suppresses feedback, whereas an absence of feedback alone does not lead to coherence, our work suggests that the reasons for "looplessness" in nature should be sought in coherence-inducing mechanisms.

networks | stability | feedback | trophic coherence | food webs

Feedback is a fundamental process in dynamical systems that occurs when the output of an element is coupled to its input. In complex systems this coupling can happen via feedback loops (or cycles) involving many elements, and hence the number and structure of such loops often determine important properties of the system as a whole (1). In systems that can be represented as graphs, or networks, the combined effects of feedback loops are described by the spectrum of eigenvalues of the adjacency matrix (the matrix of ones and zeros representing the existence or absence of edges between nodes) (2, 3). These eigenvalues can be related to fundamental questions regarding both structure (4, 5) and dynamical processes—including percolation (6), stability of dynamical elements (7), diffusion (8), or synchronization of coupled oscillators (2). Feedback loops also play a role in the behavior of many specific systems, such as robustness in gene regulatory networks (9), short-term memory in neural networks (10), or systemic risk in financial networks (11).

It has been observed that many biologically derived networks, such as food webs (12, 13) and gene transcription networks (14), have far fewer feedback loops than would be randomly expected, or even none at all. Given that acyclicity is the main requirement for being "qualitatively stable," or stable regardless of the details of dynamics (1), one might suppose that this "loopless" architecture is an adaptation for stability or some other functional advantage. However, in some cases it is not clear what the optimization mechanism behind loop suppression might be. In an ecosystem, for instance, how would a feedback cycle be eliminated if it happened to benefit the particular organisms involved?

It has recently been shown that the high linear stability of food webs is determined mainly by a structural feature called "trophic coherence," a measure of how neatly nodes fall into distinct levels (15). Trophic coherence, moreover, has been found to play an important role in other structural and dynamical properties of networks (16, 17). To investigate the relationship between trophic coherence and feedback, we here define the "coherence ensemble" of graphs and obtain expressions for various magnitudes relating to the cycle structure and spectrum of eigenvalues of coherent but otherwise random networks. We find that the number of cycles of length ν in a network can either grow or decay exponentially with ν, according to a "loop exponent," τ, which is a function of trophic coherence. A corollary is that the expectation for the leading eigenvalue is $\overline{\lambda_1} = e^{\tau}$. Thus, depending on the sign of τ and hence on trophic coherence, a network can belong either to a "loopful" regime characterized by many cycles and high leading eigenvalues or a loopless one in which cycles become scarcer with length, and all eigenvalues have real parts close to zero. In the loopless regime, the probability of drawing a directed acyclic graph tends to one with decreasing τ. We analyze a collection of empirically derived networks of several kinds and find that they conform to our theoretical predictions, with those networks with negative loop exponents displaying very few or no cycles. The observation of scarcity of feedback in certain natural systems is therefore unsurprising, given their trophic coherence.

Our work also suggests the question of what mechanisms explain trophic coherence. In the case of food webs, there are probably evolutionary pressures leading predators to specialize on prey on a narrow range of trophic levels (15). However, further research is needed to reveal other coherence-inducing mechanisms.

Results

The results we present here are for graph ensembles; that is, we make statements about expected values or probability distributions over the sets of all possible graphs that meet certain

Significance

Complex systems such as cells, brains, or ecosystems are made up of many interconnected elements, each one acting on its neighbors, and sometimes influencing its own state via feedback loops. Certain biological networks have surprisingly few such loops. Although this may be advantageous in various ways, it is not known how feedback is suppressed. We show that trophic coherence, a structural property of ecosystems, is key to the extent of feedback in these as well as in many other systems, including networks related to genes, neurons, metabolites, words, computers, and trading nations. We derive mathematical expressions that provide a benchmark against which to examine empirical data, and conclude that "looplessness" in nature is probably a consequence of trophic coherence.

Author contributions: S.J. designed research; S.J. and N.S.J. performed research; S.J. and N.S.J. analyzed data; and S.J. and N.S.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: S.Johnson.2@warwick.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613786114/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1613786114

Downloaded at Palestinian Territory, occupied on December 9, 2021

www.manaraa.com

constraints. As we shall see, these results can shed light on the relationships between the structural features of many real-world networks, to the extent that they can be regarded as random draws from a particular ensemble. Paul Erdős and Alfred Rényi pioneered this approach to graph theory with their analysis of the ensemble of all graphs with a given number of nodes N and edges L (18). The Erdős–Rényi ensemble has two regimes, one for $L > N$, in which almost all graphs will include a giant connected component, and another at $L < N$, in which no component will have more than $O(\ln N)$ nodes. Sometimes other characteristics are important. For example, whereas node degrees (i.e., numbers of neighbors) in the Erdős–Rényi ensemble are Poisson-distributed, real networks often have heavy-tailed degree distributions—a property that affects many other topological features (19). Such systems might thus be better studied by means of the configuration ensemble, the set of all graphs with not only given N and L but also a given degree sequence (20).

In the same spirit, we here define the coherence ensemble as the set of directed graphs with a given number of nodes and degree sequence, plus a specified trophic coherence. We go on to show that in this ensemble there are also two regimes, depending on a single parameter, τ, called the loop exponent:

$$\tau = \ln \alpha + \frac{1}{2\tilde{q}^2} - \frac{1}{2q^2}, \qquad [1]$$

where the branching factor α depends on the degree sequence, and q and $\tilde{q}$ capture the trophic coherence of a given network and that of its random expectation, respectively. The coherence ensemble expectations for magnitudes such as the number of cycles of given length, or the leading eigenvalue, depend on τ exponentially. Therefore, as we shall go on to show, the sign of τ determines whether a network belongs to the loopless ($\tau < 0$) or the loopful ($\tau > 0$) regimes.

Definitions. Consider the directed, unweighted graph given by the $N \times N$ adjacency matrix $A = (a_{ij})$, which has $L = \sum_{ij} a_{ij}$ directed edges. The in and out degrees of node i are $k_i^{in} = \sum_j a_{ij}$ and $k_x i^{out} = \sum_j a_{ji}$, respectively, and the mean degree is $\langle k \rangle = L/N$ (we shall use the notation $\langle \cdot \rangle$ to refer to averages over nodes in a given graph, as opposed to ensemble averages). Note that the mean degree can be regarded as either the mean in degree or the mean out degree, because these coincide: $\langle k \rangle = N^{-1} \sum_i k_i^{in} = N^{-1} \sum_i k_i^{out}$. An important magnitude that depends only on degrees is the branching factor:

$$\alpha = \frac{\langle k^{in} k^{out} \rangle}{\langle k \rangle}. \qquad [2]$$

Note that this magnitude, which together with trophic coherence determines the loop exponent τ, only depends on the mean degree $\langle k \rangle$ and the correlations between in and out degrees, and not on other aspects of the degree distributions.

The eigenspectrum of A is $\{\lambda_i\}$. The trace of the n-th power of any square matrix A can be expressed in terms of its eigenvalues as $\mathrm{Tr}(A^n) = \sum_i \lambda_i^n$ (21). Therefore, the distribution of eigenvalues, $p(\lambda)$, is related to powers of A via its moments:

$$\langle \lambda^\nu \rangle = \frac{1}{N} \mathrm{Tr}(A^\nu). \qquad [3]$$

Because A is not, in general, symmetric, its eigenvalues will be complex. The trace of A is real and invariant with respect to a change of basis, so the eigenvalues of A will always be distributed symmetrically around the real axis (21). Of particular interest is the eigenvalue with the largest real part, λ_1—usually referred to as A's leading eigenvalue.

A "basal node" is one with in degree equal to zero. If a graph has at least one basal node (our assumption throughout), and every node belongs to at least one directed path that includes a basal node, we can define the trophic level of each node i as

$$s_i = 1 + \frac{1}{k_i^{in}} \sum_j a_{ij} s_j. \qquad [4]$$

With no loss of generality for subsequent results, we define the trophic level of basal nodes as $s_i = 1$ ($\forall i$ such that $k_i^{in} = 0$) (22). This is the convention in ecology, where the trophic level of a species informs as to its ecological function: Typically, plants have $s = 1$, herbivores $s = 2$, and omnivores and carnivores $s > 2$.* Note that Eq. **4** is a system of linear equations that can be solved whenever every node is on a path that begins at a basal node (15). Hence, despite the recurrent nature of this definition of trophic levels, the presence of cycles does not pose a problem.

In ref. 15 we defined the "trophic difference" associated to each edge: $x_{ij} = s_i - s_j$. The distribution of trophic differences over edges, $p(x)$, has mean $L^{-1} \sum_{ij} a_{ij} x_{ij} = 1$ by definition,† and we can measure the graph's trophic coherence with its SD:

$$q = \sqrt{\frac{1}{L} \sum_{ij} a_{ij} x_{ij}^2 - 1}. \qquad [5]$$

A graph will be more trophically coherent the closer q is to zero, so we refer to q as an "incoherence parameter." Maximal coherence, $q = 0$, corresponds to a "layered" network in which every node has an integer trophic level, and, as q increases, the further the system departs from this ordered configuration (15, 16).

The number of directed paths (henceforth "paths") of length ν in A is $n_\nu = \sum_{ij} (A^\nu)_{ij}$. The number of directed cycles (henceforth "cycles") of length ν is $m_\nu = \mathrm{Tr}(A^\nu)$, which, according to Eq. **3** can be expressed as $m_\nu = N \langle \lambda^\nu \rangle$. (Note that we are not referring here to simple paths and simple cycles, in which no node can be repeated.) This definition of m_ν counts every unique cycle ν times, so the number of unique cycles will be $m_\nu^u = m_\nu / \nu$.

The "directed configuration ensemble" is the set of all possible graphs with given in- and out-degree sequences (24). If the number of basal edges connected to basal nodes in a graph drawn from this ensemble is L_B, then for any node i the expected proportion of in neighbors connected to a basal node will be $k_i^{in} L_B / L$. To obtain several expectations related to trophic coherence exactly we define a modified version of this ensemble called the "basal ensemble," which is the subset of graphs from the directed configuration ensemble that satisfy the constraint that the proportion of neighbors connected to basal nodes is exactly $k_i^{in} L_B / L$ for every node i. It is straightforward to determine that in this ensemble the expectations for the trophic coherence and for the branching factor are given, respectively, by

$$\tilde{q} = \sqrt{\frac{L}{L_B} - 1} \qquad [6]$$

and

$$\tilde{\alpha} = \frac{L - L_B}{N - B} \qquad [7]$$

(where we use the notation $E(z) = \tilde{z}$ to refer to the expectation of magnitude z in the basal ensemble). The full derivation

*Eq. **4** is similar to the definition of the PageRank algorithm used by the search engine Google (23). The main difference is that the sum in Eq. **4** is normalized by k_i^{in}, whereas PageRank divides each term in the sum by k_j^{out}. Also, the small "teleportation" additive term that PageRank includes to avoid problems with cycles is here the "+1" term that induces the hierarchy of trophic levels. Both measures are related to diffusion processes, but whereas PageRank provides the probability of a node's being reached by a "random surfer" (a random walker with some chance of teleportation), Eq. **4** provides a measure of how far the biomass arriving at a given node has traveled from the basal nodes.

†This can be easily seen by noting that, for any node i, the average difference over its incoming edges is $\sum_{ij} a_{ij}(s_i - s_j)/k_i^{in} = 1$.

Downloaded at Palestinian Territory, occupied on December 9, 2021
ECOLOGY
MATHEMATICS
www.manaraa.com

of these results can be found in *SI Appendix*. In the limit $N \to \infty$, with $L/N \to \infty$, expectations in the basal ensemble and the directed configuration ensemble converge. For finite graphs, we show numerically in *SI Appendix* that expectations in the two ensembles are close. Eqs. **6** and **7** can therefore be considered reasonable null expectations for real networks given only N, L, B and L_B, that is, in the absence of information regarding in- and out-degree correlations or trophic coherence.

The Coherence Ensemble. Let us now consider the ensemble of directed graphs that not only have given in- and out-degree distributions (as in the directed configuration ensemble) but also given trophic coherence. We shall refer to this as the "coherence ensemble" and use the notation $E(z) = \overline{z}$ for the expected values of quantities z in this ensemble. For networks in the coherence ensemble, the probability of a randomly chosen path of length ν's being a cycle can be obtained by considering a random walk along the edges of the graph and computing the probability that it returns to the initial node after ν hops. This constraint implies that the sum of the trophic differences x_k over the $k = 1, ...\nu$ edges involved, $S = \sum_k x_k$, must be zero. Let us approximate the differences x_k as independent random variables drawn from the trophic difference distribution $p(x)$. According to the central limit theorem, the distribution $p(S)$ will tend, with increasing ν, to a Gaussian with mean $\nu\langle x\rangle = \nu$ and variance νq^2. Because cycles are paths that satisfy $S = 0$, the expected proportion of paths of length ν that are cycles, $\overline{c}_\nu$, will be proportional to $p(S = 0)$. That is,

$$\overline{c}_\nu = B_\nu \frac{1}{\sqrt{\nu} q} \exp\left(-\frac{\nu}{2q^2}\right). \quad [8]$$

Not all of the paths satisfying $S = 0$ will return to the initial node, and this effect is accounted for by the factor B_ν. We can obtain B_ν by particularizing for the basal ensemble case, for which q is given by Eq. **6**, and $c_\nu = \tilde{\alpha}/L$ (*SI Appendix*). Inserting these values into Eq. **8**, we have

$$B_\nu = \frac{\tilde{\alpha}}{L} \sqrt{\nu} \tilde{q} \exp\left(\frac{\nu}{2\tilde{q}^2}\right). \quad [9]$$

Therefore, an approximate expression for $\overline{c}_\nu$ is

$$\overline{c}_\nu = \frac{\tilde{\alpha}}{L} \frac{\tilde{q}}{q} \exp\left[\frac{\nu}{2}\left(\frac{1}{\tilde{q}^2} - \frac{1}{q^2}\right)\right]. \quad [10]$$

The expected proportion of paths of size ν that are cycles can thus either decrease or increase exponentially with ν, depending on whether a particular graph is more or less trophically coherent than the null expectation given its degree sequence. Eq. **10** was obtained using the central limit theorem and so should only be valid for moderately large ν. However, if the distribution of differences, $p(x)$, is approximately normal, it will be a good approximation also at low values of ν. We have also assumed the trophic differences of each path to be independent random variables drawn from $p(x)$, an approximation that will hold as long as there are no significant correlations between these differences.

Let us now assume that the total number of paths in the coherence ensemble is given approximately by $\overline{n}_\nu \simeq L\alpha^{\nu-1}$, as in the basal and the directed configuration ensembles—that is, irrespective of q (*SI Appendix*). This is a reasonable assumption, at least for low ν, because α is the key element determining the number of ways a set of edges can be concatenated. (For finite N, the approximation may break down at high ν and low q, because the maximum path length will be shorter in highly coherent graphs than in random ones.) Combining this with Eq. **10** we obtain the expected number of cycles of length ν:

$$\overline{m}_\nu = \frac{\tilde{\alpha}\tilde{q}}{\alpha q} e^{\tau\nu}, \quad [11]$$

where the "loop exponent" τ has already been supplied in Eq. **1**. The term $1/\tilde{q}^2 - 1/q^2$ in Eq. **1** will be negative for networks that are more coherent than the random expectation ($q < \tilde{q}$) and positive for those that are less so, whereas the sign of $\ln\alpha$ depends on whether α is greater or less than 1. Eq. **11** implies that the expected number of cycles of length ν in a graph can either grow exponentially with ν, when $\tau > 0$, or decrease exponentially, if $\tau < 0$. Thus, which of these two regimes a given graph finds itself in is determined by the correlation between in and out degrees, $\alpha = \langle k^{in} k^{out}\rangle/\langle k\rangle$; the proportion of edges that connect to basal nodes, L_B/L (via $\tilde{q} = \sqrt{L/L_B - 1}$); and the trophic coherence, given by q. Note that, as mentioned above, the definition of m_ν counts each cycle ν times, so the expected number of unique cycles is

$$\overline{m}^u_\nu = \frac{\tilde{\alpha}\tilde{q}}{\alpha q} \frac{e^{\tau\nu}}{\nu}. \quad [12]$$

The number of cycles is related to the eigenspectrum of the adjacency matrix through Eq. **3**. Therefore, from Eq. **11** we have that the expected value of the ν-th moment of the distribution of eigenvalues is

$$\overline{\langle\lambda^\nu\rangle} = \frac{1}{N}\sum_i \overline{\lambda_i}^\nu = \frac{1}{N}\frac{\tilde{\alpha}\tilde{q}}{\alpha q} e^{\tau\nu}. \quad [13]$$

We can use this relation to obtain, for the coherence ensemble, the expected value of the leading eigenvalue by considering the limit of large ν:

$$\lim_{\nu\to+\infty}\left(\sum_i \overline{\lambda_i}^\nu\right)^{\frac{1}{\nu}} = \overline{\lambda_1} = e^\tau. \quad [14]$$

The expressions for the configuration ensemble can be recovered by taking $q = \tilde{q}$, which, according to Eq. **1**, implies $\tau = ln\alpha$. Thus, the leading eigenvalue in the directed configuration ensemble is $\tilde{\lambda}_1 = \alpha = \langle k^{in} k^{out}\rangle/\langle k\rangle$. If the graph were symmetric ($k_i^{in} = k_i^{out} = k_i$, $\forall i$), we would have $\tilde{\lambda}_1^{Sym} = \langle k^2\rangle/\langle k\rangle$, whereas for the Erdős–Rényi ensemble we obtain $\tilde{\lambda}_1^{ER} = 1 + \langle k\rangle$. These particular cases are in agreement with previous mean-field results for these ensembles (25, 26). The expected distribution of eigenvalues is entirely defined by its full set of moments, as given by Eq. **13**. For instance, the moment-generating function for graphs with given τ is

$$M_\lambda(t) = \sum_{\nu=0}^{\infty} \frac{t^\nu}{\nu!}\overline{\langle\lambda^\nu\rangle} = \left(1 - \frac{1}{N}\frac{\tilde{\alpha}\tilde{q}}{\alpha q}\right) + \frac{1}{N}\frac{\tilde{\alpha}\tilde{q}}{\alpha q}\exp\left(te^\tau\right). \quad [15]$$

Empirical Networks. We have obtained the adjacency matrices of 62 directed networks from various sources. Several details of each, including references, are listed in *SI Appendix*, Tables S1–S4. There are three broad classes of biologically derived network in our dataset: food webs, gene regulatory networks, and metabolic networks. We also include a neural network, and several man-made networks: two of international trade, a P2P file-sharing network, and a network of concatenated English words. In all cases we have removed self-edges if present, mainly because these are not reported for many of the networks, and the nature of self-interaction is often different from that occurring between elements. However, in *SI Appendix* we also analyze the same networks while conserving self-edges when reported, and the results do not differ significantly. Fig. 1 displays the leading eigenvalues, λ_1, against τ for the 62 networks, with different classes of network identified by the symbols, as indicated. The coherence ensemble expected value given by Eq. **1**, shown with a line, provides a good estimate of almost all of the empirical values. The inset shows the positive quadrant on a semilog scale. Of the 62 networks in our dataset, 36 have $\tau < 0$ and 26 have $\tau > 0$. The mean values of λ_1 for these are,

Downloaded at Palestinian Territory, occupied on December 9, 2021
www.manaraa.com

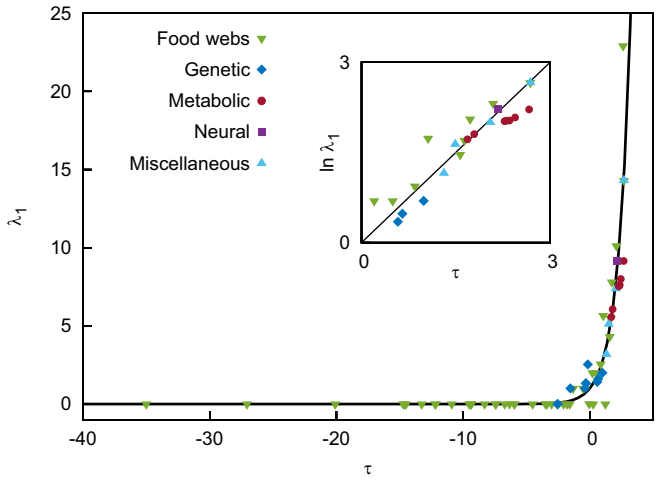

Fig. 1. Loop exponents τ are informative of leading eigenvalues for a variety of empirical networks. Here we plot leading eigenvalues λ_1 of several directed networks, against τ as given by Eq. **1**; symbols indicate graph data derived from food webs (green down-pointing triangles), gene regulatory networks (dark blue diamonds), metabolic networks (burgundy circles), a neural network (purple square), and other miscellaneous networks (light blue upward-pointing triangles). The line shows expected leading eigenvalue $\overline{\lambda_1}$ in the coherence ensemble, as given by Eq. **14**. (*Inset*) Semilog version of the positive quadrant of the main panel (Pearson's correlation coefficient: $r^2 = 0.87$). For these results, self-edges were removed from the networks; a similar figure in which self-edges are included can be found in *SI Appendix*. Details for each network, including references, are listed in the tables of *SI Appendix*.

respectively, $\lambda_1(\tau<0) = 0.22 \pm 0.54$ and $\lambda_1(\tau>0) = 6.1 \pm 7.4$. In other words, the two regimes are separated by an order of magnitude in the leading eigenvalue.

Table 1 shows the mean and SD of several magnitudes for the three main classes of biologically derived network in our dataset. The first three rows are for the ratios of measured values to the basal ensemble expectations. The graphs corresponding to food webs are significantly coherent ($q/\tilde{q} < 1$) and have slightly lower mean trophic levels than the expectation ($\langle s\rangle/\tilde{s} \lesssim 1$). The networks derived from gene regulation have coherence and mean trophic levels that are very close to their expected values. Meanwhile, the networks linked to metabolism are significantly incoherent ($q/\tilde{q} > 1$) and have mean trophic levels that are higher than expected ($\langle s\rangle/\tilde{s} > 1$). The measured values of α are in all three classes slightly higher than the expectation, but in the cases of food webs and gene regulatory networks the difference is within one SD. However, the metabolism-related networks display marked positive correlations between in and out degrees ($\alpha/\tilde{\alpha} > 1$). The fifth row shows the proportion of networks in each class that have negative τ. For the food web and gene regulatory network data, it is 74 and 63%, respectively, whereas the metabolism-related networks are all in the positive τ regime. This leads to average leading eigenvalues, shown in the fourth row, that are much greater for metabolic network data than for food webs or gene regulatory-related networks. The sixth row gives the proportion of networks in each class that are acyclic, a feature we discuss in the next section. In *SI Appendix* we show an example of each kind of network to illustrate the wide variety of trophic structures found among natural systems.

Directed Acyclic Graphs. Let us consider the probability that a graph randomly chosen from the coherence ensemble will have exactly m_ν cycles of length ν. We shall assume that each path is an independent random event with two possible outcomes: With probability $\overline{c}_\nu$ it is a cycle, whereas with $1-\overline{c}_\nu$ it is not. The number of cycles m_ν will therefore be binomially distributed:

$$p(m_\nu) = \binom{\tilde{n}_\nu}{m_\nu} \overline{c}_\nu^{m_\nu} (1-\overline{c}_\nu)^{\tilde{n}_\nu - m_\nu}. \quad [16]$$

We can use this distribution to obtain the probability that a network from the coherence ensemble would have no directed cycles of length greater or equal to n:

$$P_n = \prod_{\nu=n}^{\infty} p(m_\nu = 0). \quad [17]$$

For instance, the probability that a network drawn randomly from this ensemble would be acyclic is

$$P_{acyclic} = \prod_{\nu=2}^{\infty} p(m_\nu = 0) \quad [18]$$

$$= \prod_{\nu=2}^{\infty} \left\{ 1 - \frac{\tilde{\alpha}}{L}\frac{\tilde{q}}{q} \exp\left[\frac{\nu}{2}\left(\frac{1}{\tilde{q}^2} - \frac{1}{q^2}\right)\right]\right\}^{L\alpha^{\nu-1}}.$$

Taking logarithms and considering graphs with sufficiently negative τ that we can use the approximation $\ln(1-x) \simeq -x$, we have

$$\ln P_{acyclic} \simeq -\frac{\tilde{\alpha}\tilde{q}}{\alpha q} \sum_{\nu=1}^{\infty} e^{\tau\nu}; \quad [19]$$

and, after performing the sum and some algebra,

$$P_{acyclic} \simeq \exp\left[-\frac{\tilde{\alpha}\tilde{q}}{\alpha q}\frac{1}{(e^{-\tau}-1)}\right]. \quad [20]$$

Therefore, as $\tau \to -\infty$, networks are almost always acyclic. We note that these sums include small values of ν for which results are only approximate unless the distribution of trophic differences, $p(x)$, is Gaussian.

Fig. 2 is a scatter plot of our set of empirical networks according to the terms in Eq. **1**: $1/q^2 - 1/\tilde{q}^2$ and $\ln\alpha$. Each network is represented with either a triangle or a circle depending on whether it has cycles or not, respectively. The curve $\tau = 0$ separates the two regimes, and it is clear that whereas almost all of the networks in the positive τ regime have cycles (the exceptions being two food webs), as one moves into the negative τ regime most of the examples are acyclic. The inset shows the probability of a network randomly drawn from the coherence ensemble being acyclic, as given by Eq. **20** and indicated in the caption. One can compute, for each empirical network, the probability that it is acyclic according to Eq. **20**. Thus, given only a network's degree sequence and trophic coherence, we would expect it to be acyclic if $P_{acyclic} > 0.5$ (note that a network might be in the $\tau < 0$ regime yet still be predicted to have cycles by this decision rule). We find that, out of the 62 networks, eight are incorrectly classified: Seven food webs are acyclic despite being predicted to have cycles, and one gene regulatory network would be expected (by a

Table 1. Mean values and SD of the ratios $q/\tilde{q}$, $\langle s\rangle/\tilde{s}$, and $\alpha/\tilde{\alpha}$, and of the leading eigenvalue λ_1, for three classes of biologically derived networks and fractions of these networks to have $\tau < 0$, and to be acyclic, of the total in our dataset

Quantity	Food webs	Genetic	Metabolic
$q/\tilde{q}$	0.44 ± 0.17	0.99 ± 0.05	1.81 ± 0.11
$\langle s\rangle/\tilde{s}$	0.88 ± 0.18	1.00 ± 0.001	2.05 ± 0.01
$\alpha/\tilde{\alpha}$	1.02 ± 0.23	1.19 ± 0.34	3.98 ± 1.04
λ_1	1.54 ± 4.09	1.36 ± 0.75	7.36 ± 1.20
$\tau < 0$	31/42	5/8	0/7
Acyclic	31/42	1/8	0/7

Downloaded at Palestinian Territory, occupied on December 9, 2021
www.manaraa.com

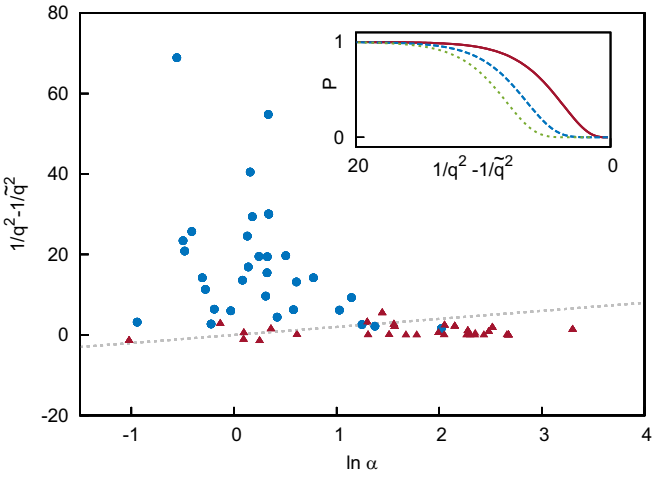

Fig. 2. The components of the loop exponent τ, coherence vs. in-degree-out-degree correlations, are predictive of whether empirical networks will have cycles. Here we show a scatter plot of several networks according to $1/q^2 - 1/\tilde{q}^2$ and $\ln \alpha$, the two terms in Eq. **1**. Blue circles: networks with no cycles. Burgundy triangles: networks with at least one cycle. The $\tau = 0$ line is shown with a dashed line (the negative τ regime falls above the line). Details for each network, including references, are listed in the tables of *SI Appendix*. (*Inset*) Probabilities of networks in the coherence ensemble being acyclic, according to Eq. **20**, as a function of $1/q^2 - 1/\tilde{q}^2$. Solid line: $L/L_B = 10$ and $\alpha = 1$; dashed line: $L/L_B = 100$ and $\alpha = 1$; dotted line: $L/L_B = 100$ and $\alpha = 2$.

small margin) to be acyclic but is not. The prediction is therefore accurate in 87% of instances.

Discussion

We have shown that a directed network can belong to either of two regimes characterized by fundamentally different cycle structures, depending on the sign of a single parameter, τ, which is a function of the trophic coherence and the branching factor, as given by Eq. **1**. Because the expected number of cycles of length ν is proportional to $e^{\tau\nu}$, positive τ implies an exponentially growing number of cycles with length, whereas for negative τ the probability of finding cycles is vanishing. This, in turn, has a crucial effect on the spectral properties of graphs: In particular, the expected value of the leading eigenvalue of the adjacency matrix is $\overline{\lambda_1} = e^{\tau}$. A corollary is that graphs drawn randomly from the negative τ regime have a high probability of being directed acyclic graphs, the main requisite for qualitative stability (1).

Our results provide expected values for what we have called the coherence ensemble—the set of directed graphs with a given degree sequence and trophic coherence—and do not, therefore, place bounds on the possible values a given network can exhibit. However, analysis of a set of empirically derived networks of various kinds shows that in most cases these expected values are very good approximations to the ones measured, suggesting that the coherence ensemble may be an appropriate null model to use in many cases. We should note also that we have focused on binary networks (those with adjacency matrices of only ones and zeros). Although some of the results could be extended to weighted networks in a straightforward way, it is not so obvious how concepts such as trophic coherence should be understood when a distinction between excitatory and inhibitory interactions is made. We leave such questions for future work.

The fact that many biologically derived networks have surprisingly few feedback loops has recently been attributed to considerations of robustness (14), stability (13), and to an "inherent directionality" (27). Our results are compatible with the latter, because network directionality would ensue from trophic coherence (that is, because the distribution of differences $p(x)$ is centered at 1 and has variance q^2, the expected number of edges with $x < 0$ is $L\int_{-\infty}^{0} p(x)dx$, a monotonically increasing function of q). However, neither a suppression of cycles nor an imposed directionality will in itself induce trophic coherence, as can be easily seen in the case of the "cascade model" (28). In this network assembly model there is a strict hierarchy of nodes; directed edges are placed at random with the sole constraint that the out node must be below the in node in the hierarchy, thus emulating the situation in many food webs where predators tend to be larger than their prey. Such networks are by construction acyclic and directional, yet they do not exhibit significant trophic coherence (15). However, our analysis indicates that any network formation processes that tended to induce a certain trophic coherence would confer the properties of a low or negative τ on a system, without necessarily being the result of an optimization for low feedback. For example, in ecosystems many features of species, such as body size and metabolic rate, are related to trophic levels. Because predators often specialize in consuming prey with specific characteristics, they naturally focus on relatively narrow trophic ranges, a mechanism that could lead to networks that are more coherent than the random expectation. This idea is borne out by generative network models that capture this effect—namely, the "preferential preying model" presented in ref. 15 (which produces acyclic graphs with tunable trophic coherence) and an extension of this model studied in ref. 16 (which can set the trophic coherence of graphs with cycles). However, relatively little is yet known about the mechanisms that might lead to trophic coherence more generally.

Although we have argued here that looplessness should be regarded as an effect of trophic coherence, this naturally moves the challenge to establishing the origins of trophic coherence. Further research is needed to address this issue, possibly involving the relation between trophic levels and the functional roles of nodes. This view has interesting parallels with recent work on node roles in generic directed networks, based on topological similarity, which when applied to food webs reveals trophic structure (29, 30). Functional groups have also been uncovered in ecosystems using stochastic block models, which can take nontrophic interactions into account (31, 32).

A relation between node function and trophic level may exist in systems other than ecological ones. For instance, in the word adjacency network of the children's book *Green Eggs and Ham*, by Dr. Seuss, we find that the mean trophic level of common nouns is $s_{noun} = 1.4 \pm 1.2$, whereas that of verbs is $s_{verb} = 7.0 \pm 2.7$ (*SI Appendix*, Fig. S4). This shows that in networks where node function is encoded in trophic levels any mechanism whereby edges tended to occur between nodes with specific functions might develop nontrivial coherence (or incoherence). More broadly, it also suggests that the trophic structure of directed networks may provide insights into their function and dynamics. Classifying nodes by trophic level, as has long been standard in ecology, might also tell us something about the functions of, say, genes, metabolites, neurons, economic agents, or words in unknown languages. In view of these considerations, we believe that further exploration of the trophic structure of networks, and its relation to function and dynamics, will prove a fruitful avenue for learning about many complex systems.

ACKNOWLEDGMENTS. We thank M. A. Muñoz and V. Domínguez-García for a fruitful collaboration from which some of these ideas sprouted; I. Johnston, M. Ibáñez Berganza, and J. Klaise for useful discussions; and L. Albergante, J. Dunne, U. Jacob, R. M. Thompson, C. R. Townsend, U. Alon, M. E. J. Newman, W. de Nooy, and J. Leskovec for providing data or making them available online (see *SI Appendix*.)

Downloaded at Palestinian Territory, occupied on December 9, 2021
www.manaraa.com

1. May RM (1973) Qualitative stability in model ecosystems. *Ecology* 54:638–641.
2. Arenas A, Díaz-Guilera A, Kurths J, Moreno Y, Zhou C (2008) Synchronization in complex networks. *Phys Rep* 469:93–153.
3. Brouwer AE, Haemers WH (2011) *Spectra of Graphs* (Springer, New York).
4. Fiedler M (1973) Algebraic connectivity of graphs. *Czech Math J* 23:298–305.
5. Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103:8577–8582.
6. Bollobás B, Borgs C, Chayes J, Riordan O (2010) Percolation on dense graph sequences. *Ann Prob* 38:150–183.
7. May RM (1972) Will a large complex system be stable. *Nature* 238:413–414.
8. Barrat A, Barthélemy M, Vespignani A (2008) *Dynamical Processes on Complex Networks* (Cambridge Univ Press, Cambridge, UK).
9. Zhang Z, et al. (2012) Chaotic motifs in gene regulatory networks. *PLoS One* 7: e39355.
10. Johnson S, Marro J, Torres JJ (2013) Robust short-term memory without synaptic learning. *PLoS One* 8:e50276.
11. DasGupta B, Kaligounder L (2014) On global stability of financial networks. *J Complex Networks* 2:313–354.
12. Milo R, et al. (2002) Network motifs: Simple building blocks of complex networks. *Science* 298:824–827. Available at science.sciencemag.org/content/298/5594/824.
13. Borrelli JJ (2015) Selection against instability: Stable subgraphs are most frequent in empirical food webs. *Oikos* 124:1583–1588.
14. Albergante L, Blow JJ, Newman TJ (2014) Buffered Qualitative Stability explains the robustness and evolvability of transcriptional networks. *Elife* 3:e02863.
15. Johnson S, Domínguez-García V, Donetti L, Muñoz MA (2014) Trophic coherence determines food-web stability. *Proc Natl Acad Sci USA* 111:17923–17928.
16. Klaise J, Johnson S (2016) From neurons to epidemics: How trophic coherence affects spreading processes. *Chaos* 26:065310.
17. Domínguez-García V, Johnson S, Muñoz MA (2016) Intervality and coherence in complex networks. *Chaos* 26:065308.
18. Erdős P, Rényi A (1959) On random graphs I. *Publ Math Debrecen* 6:290–297.
19. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45:167–256.
20. Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. *Random Struct Algorithm* 6:161–180.
21. Cohn PM (1982) *Algebra* (Wiley, New York).
22. Levine S (1980) Several measures of trophic structure applicable to complex food webs. *J Theor Biol* 83:195–207.
23. Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. *Comput Networks ISDN Syst* 30:107–117.
24. Kim H, Del Genio CI, Bassler KE, Toroczkai Z (2012) Constructing and sampling directed graphs with given degree sequences. *New J Phys* 14:023012.
25. Chung F, Lu L, Vu V (2003) Spectra of random graphs with given expected degrees. *Proc Natl Acad Sci USA* 100:6313–6318.
26. Nadakuditi RR, Newman MEJ (2013) Spectra of random graphs with arbitrary expected degrees. *Phys Rev E Stat Nonlin Soft Matter Phys* 87:012803.
27. Domínguez-García V, Pigolotti S, Muñoz MA (2014) Inherent directionality explains the lack of feedback loops in empirical networks. *Sci Rep* 4:7497.
28. Cohen JE, Newman CM (1985) A stochastic theory of community food webs I. Models and aggregated data. *Proc R Soc London Ser B* 224:421–448.
29. Cooper K, Barahona M (2010) Role-based similarity in directed networks. arXiv: 1012.2726.
30. Beguerisse-Díaz M, Garduno-Hernández G, Vangelov B, Yaliraki SN, Barahona M (2014) Interest communities and flow roles in directed networks: The twitter network of the uk riots. *J Royal Soc Interface* 11:20140940.
31. Kéfi S, et al. (2015) Network structure beyond food webs: Mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology* 96:291–303.
32. Kéfi S, Miele V, Wieters EA, Navarrete SA, Berlow EL (2016) How structured is the entangled bank? The surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience. *PLoS Biol* 14:e1002527.

Downloaded at Palestinian Territory, occupied on December 9, 2021

www.manaraa.com