

Looplessness in networks is linked to trophic coherence

Samuel Johnson^{1*} and Nick S. Jones²

¹Warwick Mathematics Institute, and Centre for Complexity Science,
University of Warwick, Coventry CV4 7AL, United Kingdom.

²Department of Mathematics, Imperial College London,
London SW7 2AZ, United Kingdom.

*E-mail: S.Johnson.2@warwick.ac.uk

Abstract

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, are 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. Since 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.

Keywords: Networks, feedback, stability, trophic coherence

<http://www.pnas.org/content/early/2017/05/15/1613786114>

Significance Statement

Complex systems such as cells, brains or ecosystems are made up of many interconnected elements, each one acting on its neighbours, and sometimes influencing its own state via feedback loops. Certain biological networks have surprisingly few such loops. While 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 which provide a benchmark against which to examine empirical data,

and conclude that ‘looplessness’ in nature is probably a consequence of trophic coherence.

Introduction

Feedback is a fundamental process in dynamical systems which 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 which 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 behaviour 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 optimisation 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]. In order 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 characterised 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 analyse 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: what mechanisms explain trophic coherence? In the case of food webs, there are probably evolutionary pressures leading predators to specialise 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 which meet certain 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, while node degrees (i.e. numbers of neighbours) in the Erdős-Rényi ensemble are Poisson distributed, real networks often have heavy-tailed degree distributions – a property which 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:

$$\tau = \ln \alpha + \frac{1}{2\tilde{q}^2} - \frac{1}{2q^2}, \quad (1)$$

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 $k_i^{out} = \sum_j a_{ji}$, 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, since these coincide: $\langle k \rangle = N^{-1} \sum_i k_i^{in} = N^{-1} \sum_i k_i^{out}$. An important magnitude which depends only on degrees is the branching factor:

$$\alpha = \frac{\langle k^{in} k^{out} \rangle}{\langle k \rangle}. \quad (2)$$

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 $\text{Tr}(A^n) = \sum_i \lambda_i^n$ [21]. Therefore, the distribution of eigenvalues, $p(\lambda)$, is related to powers of A via its moments:

$$\langle \lambda^\nu \rangle = \frac{1}{N} \text{Tr}(A^\nu). \quad (3)$$

Since 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 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 which includes a basal node, we can define the trophic level of each node i as

$$s_i = 1 + \frac{1}{k_i^{\text{in}}} \sum_j a_{ij} s_j. \quad (4)$$

With no loss of generality for subsequent results, we define the trophic level of basal nodes as $s_i = 1$ ($\forall i$ such that $k_i^{\text{in}} = 0$) [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 $s > 2$.¹ Note that Eq. (4) is a system of linear equations which can be solved whenever every node is on a path which 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,² and we can measure the graph’s ‘trophic coherence’ with its standard deviation:

$$q = \sqrt{\frac{1}{L} \sum_{ij} a_{ij} x_{ij}^2 - 1}. \quad (5)$$

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,

¹Eq. (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 normalised by k_i^{in} , whereas PageRank divides each term in the sum by k_j^{out} . Also, the small ‘teleportation’ additive term which PageRank includes to avoid problems with cycles is here the ‘+1’ term which induces the hierarchy of trophic levels. Both measures are related to diffusion processes; but while PageRank provides the probability of a node 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 travelled from the basal nodes.

²This can be easily seen by noting that, for any node i , the average difference over its incoming edges is $\sum_{ij} a_{ij} (s_i - s_j) / k_i^{\text{in}} = 1$.

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 $m_\nu = \text{Tr}(A^\nu)$, 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-neighbours connected to a basal node will be $k_i^{\text{in}} L_B / L$. In order 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 which satisfy the constraint that the proportion of neighbours connected to basal nodes is exactly $k_i^{\text{in}} L_B / L$ 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

$$\tilde{q} = \sqrt{\frac{L}{L_B} - 1} \quad (6)$$

and

$$\tilde{\alpha} = \frac{L - L_B}{N - B} \quad (7)$$

(where we use the notation $E(z) = \tilde{z}$ to refer to the expectation of magnitude z in the basal ensemble). The full derivation of these results can be found in SI Appendix. In the limit $N \rightarrow \infty$, with $L/N \rightarrow \infty$, 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. Equations (6) and (7) can therefore be considered reasonable null expectations for real networks given only N , L , B and L_B – i.e. in the absence of information regarding in-out-degree correlations or trophic coherence.

The coherence ensemble. Let us now consider the ensemble of directed graphs which 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 $E(z) = \bar{z}$ 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 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 $k = 1, \dots, \nu$ 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 . Since cycles are paths which satisfy $S = 0$, the expected proportion of paths of length

ν that are cycles, \bar{c}_ν , will be proportional to $p(S = 0)$. That is,

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

Not all 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 particularising for the basal ensemble case, for which q is given by Eq. (6), and $c_\nu = \tilde{\alpha}/L$ (see 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 \bar{c}_ν is

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

The expected proportion of paths of size ν which 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 which 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 $\bar{n}_\nu \simeq L\alpha^{\nu-1}$, as in the basal and the directed configuration ensembles – i.e. irrespective of q (see SI Appendix). This is a reasonable assumption, at least for low ν , since 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 :

$$\bar{m}_\nu = \frac{\tilde{\alpha}\tilde{q}}{\alpha q} e^{\tau\nu}, \quad (11)$$

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 which are more coherent than the random expectation ($q < \tilde{q}$), and positive for those which are less so; while 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 which 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

$$\bar{m}_\nu^u = \frac{\tilde{\alpha}\tilde{q}}{\alpha q} \frac{e^{\tau\nu}}{\nu}. \quad (12)$$

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

$$\langle \lambda^\nu \rangle = \frac{1}{N} \sum_i \bar{\lambda}_i^\nu = \frac{1}{N} \frac{\tilde{\alpha}\tilde{q}}{\alpha q} e^{\tau\nu}. \quad (13)$$

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 :

$$\lim_{\nu \rightarrow +\infty} \left(\sum_i \bar{\lambda}_i^\nu \right)^{\frac{1}{\nu}} = \bar{\lambda}_1 = e^\tau. \quad (14)$$

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 $\tilde{\lambda}_1^{Sym} = \langle k^2 \rangle / \langle k \rangle$; while 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

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

Empirical networks. We have obtained the adjacency matrices of 62 directed networks from various sources. Several details of each, including references, are listed in Tables S1–S4 of SI Appendix. There are three broad classes of biologically derived network in our data set: 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 analyse the same networks while conserving self-edges when reported, and the results do not differ significantly. Figure 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 the empirical values. The inset shows the positive quadrant on a semi-log scale. Of the 62 networks in our data set, 36 have $\tau < 0$ and 26 have $\tau > 0$. The mean values of λ_1 for these are, 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 standard deviation of several magnitudes for the three main classes of biologically-derived network in our data set. 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

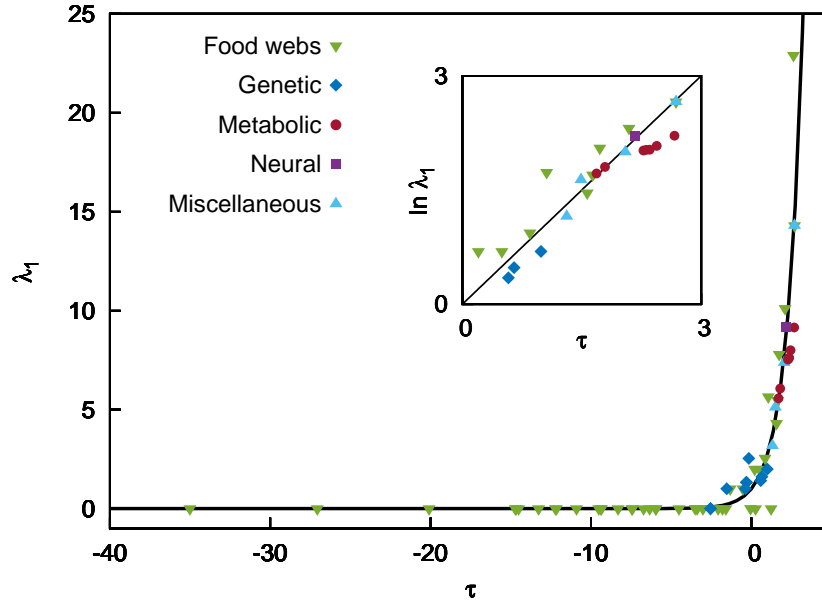


Figure 1. Loop exponents τ are informative of leading eigenvalues for a variety of empirical networks. Here we plot leading eigenvalues λ_1 of several directed networks, against τ as given by Eq. (1); symbols indicate graph data derived from food webs (green down-pointing triangles), gene regulatory networks (dark blue diamonds), metabolic networks (burgundy circles), a neural network (purple square), and other miscellaneous networks (light blue up-pointing triangles). Line: Expected leading eigenvalue $\bar{\lambda}_1$ in the coherence ensemble, as given by Eq. (14). Inset: Semi-log 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. Details for each network, including references, are listed in the tables of SI Appendix.

Table 1. Mean values and standard deviation 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 network; and fractions of these networks to have $\tau < 0$, and to be acyclic, of the total in our data set.

	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

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 which 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 which 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 standard deviation. 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 which have negative τ . For the food web and gene regulatory network data, it is 74% and 63%, respectively, while the metabolism-related networks are all in the positive τ regime. This leads to average leading eigenvalues, shown in the fourth row, which 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 which 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 \bar{c}_ν it is a cycle, while with $1 - \bar{c}_\nu$ it is not. The number of cycles m_ν will therefore be binomially distributed:

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

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

$$P_{acyclic} = \prod_{\nu=2}^{\infty} p(m_{\nu} = 0) = \prod_{\nu=2}^{\infty} \left\{ 1 - \frac{\tilde{\alpha} \tilde{q}}{L q} \exp \left[\frac{\nu}{2} \left(\frac{1}{\tilde{q}^2} - \frac{1}{q^2} \right) \right] \right\}^{L \alpha^{\nu-1}}. \quad (18)$$

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 $\tau \rightarrow -\infty$, 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.

Figure 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 while almost all 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 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 characterised 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). Since 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, while 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

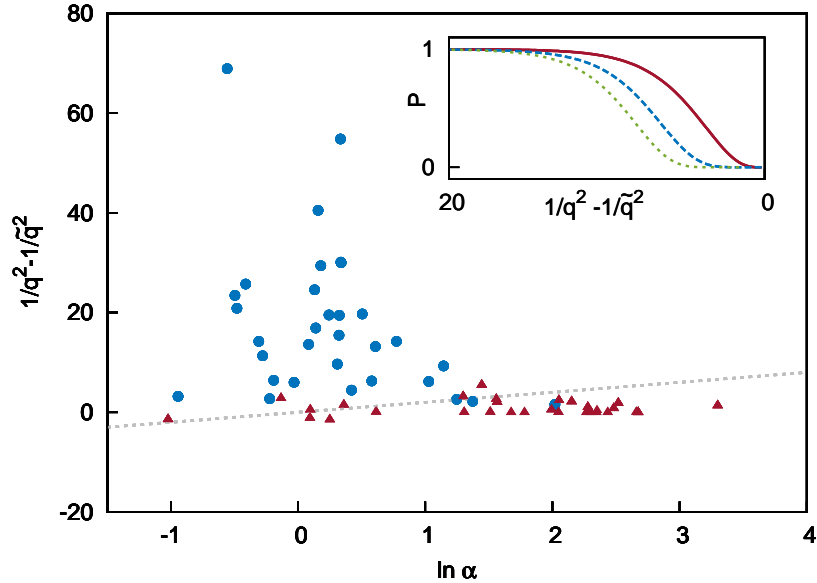


Figure 2. The components of the loop exponent τ , coherence vs in-degree-out-degree correlations, are predictive of whether empirical networks will have cycles. Here we show a scatter plot of several networks according to $1/q^2 - 1/\tilde{q}^2$ and $\ln \alpha$, the two terms in Eq. (1). Blue circles: 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 $\alpha = 2$.

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). While 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, since network directionality would ensue from trophic coherence (that is, since the distribution of differences $p(x)$ is centred 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 themselves 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]. On the other hand, our analysis indicates that any network formation processes which 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. Since predators often specialise in consuming prey with specific characteristics, they naturally focus on relatively narrow trophic ranges, a mechanism which could lead to networks that are more coherent than the random expectation. This idea is borne out by generative network models which 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 which might lead to trophic coherence more generally.

While 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 non-trophic 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 $s_{noun} = 1.4 \pm 1.2$, while that of verbs is $s_{verb} = 7.0 \pm 2.7$ (see Fig. S4 in SI Appendix). 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 non-trivial 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.

Acknowledgements

Many thanks to M.A. Muñoz and V. Domínguez-García for a fruitful collaboration from which some of these ideas sprouted. We are grateful to I. Johnston, M. Ibáñez Berganza and J. Klaise for useful discussions. Thanks also to 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.)

References

- [1] May RM (1973) Qualitative stability in model ecosystems. *Ecology* 54:638–41.
- [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 Science & Business Media).
- [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* 238:413–14.
- [8] Barrat A, Barthélemy M, Vespignani A (2008) *Dynamical Processes on Complex Networks*. (Cambridge University Press, Cambridge).
- [9] Zhang Z et al. (2012) Chaotic motifs in gene regulatory networks. *PLOS ONE* 7(7):e39355.
- [10] Johnson S, Marro J, Torres JJ (2013) Robust short-term memory without synaptic learning. *PLOS ONE* 8(1):e50276.
- [11] DasGupta B, Kaligounder L (2014) On global stability of financial networks. *Journal of Complex Networks* 2(3):313–354.
- [12] Milo R et al. (2002) Network motifs: Simple building blocks of complex networks. *Science* 298(5594):824–827.
- [13] Borrelli JJ (2015) Selection against instability: stable subgraphs are most frequent in empirical food webs. *Oikos*.
- [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(50):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) Intervals 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 Review* 45:167–256.
- [20] Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. *Random structures & algorithms* 6(2-3):161–180.
- [21] Cohn PM (1982) *Algebra*. (Wiley).
- [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. *Computer networks and ISDN systems* 30(1):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(11):6313–6318.
- [26] Nadakuditi RR, Newman MEJ (2013) Spectra of random graphs with arbitrary expected degrees. *Phys. Rev. E* 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 preprint 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. *Journal of The Royal Society Interface* 11(101):20140940.
- [31] Kéfi S et al. (2015) Network structure beyond food webs: mapping non-trophic and trophic interactions on chilean rocky shores. *Ecology* 96(1):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(8):e1002527.

Supporting Information (SI Appendix) 'Looplessness in networks is linked to trophic coherence'

Samuel Johnson^{1*} and Nick S. Jones²

¹Warwick Mathematics Institute, and Centre for Complexity Science,
University of Warwick, Coventry CV4 7AL, United Kingdom.

²Department of Mathematics, Imperial College London,
London SW7 2AZ, United Kingdom.

*E-mail: S.Johnson.2@warwick.ac.uk

CONTENTS

1 Graph ensembles

1.1 The directed configuration ensemble

1.2 The basal ensemble

1.3 Equivalence of ensembles

2 Networks with self-cycles

3 Network data

3.1 Green Eggs and Ham words network

1 Graph ensembles

1.1 The directed configuration ensemble

This is the set of all graphs with given sequences of in- and out-degrees [1]. Using this ensemble as a null-model, we can obtain the expected numbers of paths and cycles by inserting the expected value of the adjacency matrix for large graphs, $\hat{a}_{ij} = k_i^{out} k_j^{in} / L$, in the above definitions (we shall use the notation \hat{z} to refer to the expected value of a magnitude z in the directed configuration ensemble). Thus, the expected number of paths in this ensemble is

$$\hat{n}_\nu = L\alpha^{\nu-1}, \quad (1)$$

while the expected number of those paths which are cycles is

$$\hat{m}_\nu = \alpha^\nu, \quad (2)$$

where

$$\alpha = \frac{\langle k^{in} k^{out} \rangle}{\langle k \rangle}. \quad (3)$$

The branching factor α captures the correlation between the in- and out-degrees of nodes (i.e. $\alpha > \langle k \rangle$ indicates a positive correlation, with high in-degree nodes also tending to have high out-degree, while $\alpha < \langle k \rangle$ means this correlation is negative).

1.2 The basal ensemble

Let us consider the ensemble of random graphs which satisfy the following constraint: for every non-basal node, the proportion of in-coming edges which connect to basal nodes is the same. This is a sufficient condition for all non-basal nodes to have the same trophic level. More formally, if a network in this ensemble has B basal nodes, $N - B$ non-basal nodes, L edges, and L_B edges connecting to basal nodes, then every non-basal node i with in-degree k_i^{in} receives $k_i^{in} L_B / L$ edges from basal nodes and $(1 - L_B / L) k_i^{in}$ from non-basal nodes. Note that this constraint does not affect the expectations obtained above for the more general directed configuration ensemble, so the expected numbers of paths and cycles are, respectively,

$$\tilde{n}_\nu = L\alpha^{\nu-1} \quad (4)$$

and

$$\tilde{m}_\nu = \alpha^\nu \quad (5)$$

(where we have used the notation \tilde{z} for the expected value of a magnitude z in the basal ensemble). However, the fixed proportion of basal in-neighbors allows us also to derive expected values for several magnitudes in this ensemble, given $\{N, B, L, L_B\}$, as follows.

We recall from the main text [Eq. (4)] that the trophic level of each node i is

$$s_i = 1 + \frac{1}{k_i^{in}} \sum_j a_{ij} s_j \quad (6)$$

if i is non-basal (i.e. $k_i^{in} > 0$), and $s_i = 1$ if i is basal ($k_i^{in} = 0$) [2]. Let \tilde{s}_{nb} be the expected trophic level in the basal ensemble of a node given that it is not

basal. Since the expected proportion of in-neighbours which are basal, for such a node, will be L_B/L , we have from Eq. (6) that

$$\tilde{s}_{nb} = 1 + \frac{L_B + (L - L_B)\tilde{s}_{nb}}{L}, \quad (7)$$

and so $\tilde{s}_{nb} = L/L_B + 1$. The mean trophic level of the network is therefore

$$\tilde{s} = 1 + \left(1 - \frac{B}{N}\right) \frac{L}{L_B}. \quad (8)$$

In the basal ensemble there will be two types of edges: those which emanate from basal nodes, with trophic difference $x = \tilde{s}_{nb} - 1$, and those between non-basal nodes, with $x = 0$. In other words, the distribution of differences will be

$$p(x) = \frac{L_B}{L} \delta\left(x - \frac{L}{L_B}\right) + \left(1 - \frac{L_B}{L}\right) \delta(x), \quad (9)$$

where δ is the Dirac delta function. The trophic coherence associated with this distribution is

$$\tilde{q} = \sqrt{\frac{L}{L_B} - 1}. \quad (10)$$

We can also obtain the expected value of α in the basal ensemble. Every node with $k^{in} \neq 0$ (i.e. every non-basal node) has in-degree $k^{in} = L/(N - B)$; and while the out-degree of such a node is not determined, the expected value is $\tilde{k}^{out} = (L - L_B)/(N - B)$. Inserting these values into Eq. (3) yields

$$\tilde{\alpha} = \frac{L - L_B}{N - B}. \quad (11)$$

Note that this is the mean degree that would result if all basal nodes and edges emanating from basal nodes were eliminated from the network. As in the configuration ensemble, the expected proportion of paths of length ν which are cycles, in the basal ensemble, is

$$\tilde{c}_\nu = \frac{\tilde{\alpha}}{L}. \quad (12)$$

The basal ensemble does not include any structure which might lead to variance in the trophic levels of non-basal nodes, and so the bimodal $p(x)$ given by Eq. (9) is exact. For networks in which different non-basal nodes are connected to differing proportions of basal nodes, this expression will be a valid approximation only when the separation of the two modes of $p(x)$ is much larger than the spread about them.

1.3 Equivalence of ensembles

Let κ_i be the proportion of in-coming edges to node i which emanate from a basal node. In the directed configuration ensemble, the expectation for κ_i is $\hat{\kappa}_i = L_B/L, \forall i$. The basal ensemble is the subset of graphs from the directed configuration ensemble which satisfy $\kappa_i = L_B/L, \forall i$, exactly, not just in expectation. Thanks to this constraint, all non-basal nodes have the same trophic level, given by Eq. (7), in the basal ensemble. For finite graphs drawn from

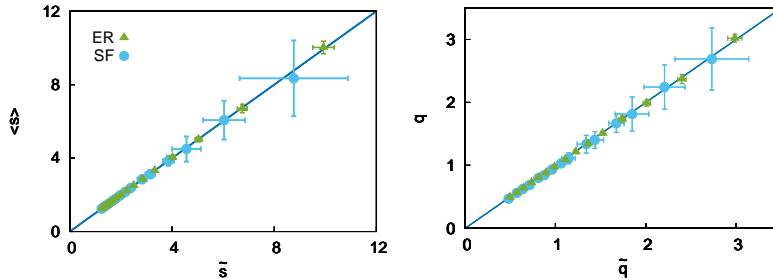


Figure S 1. Values of mean trophic level $\langle s \rangle$ (left), and incoherence parameter q (right), obtained numerically with two random graph models against the corresponding predictions for the basal ensemble, as given by Eqs. (8) and (10), respectively. The random graphs are sampled from the Erdős-Rényi ensemble (circles) with $N = 300$ and $\langle k \rangle = L/N = 10$; and from the directed configuration ensemble (triangles) with $N = 500$, $\langle k \rangle = L/N = 5$ and power-law degree sequences of exponent $\gamma = 3$. In both cases, the proportion of basal nodes, B/N , is varied from 10% (upper right hand corners) to 50% (lower left hand corners).

the directed configuration ensemble, there will be some variation in the trophic levels of non-basal nodes, leading to discrepancies with respect to the expectations of quantities such as q , α and $\langle s \rangle$. However, in the limit $N \rightarrow \infty$, with $L/N \rightarrow \infty$, expectations in the basal ensemble and the directed configuration ensemble converge, since $\kappa_i \rightarrow \hat{\kappa}_i, \forall i$. This suggests that the basal ensemble might provide a reasonable null model even for finite networks. In order to test this assumption numerically, we draw networks from the directed configuration and compare measured quantities with basal ensemble expectations. We note, however, that the results derived in the main text for the coherence ensemble do not depend on an equivalence between the basal ensemble and the directed configuration ensemble.

Figure S1 displays the results of Monte Carlo simulations which support the conjecture that the basal ensemble provides a good approximation to more general random graph ensembles as regards trophic structure. We consider a directed version of the Erdős-Rényi ensemble, in which L directed edges are distributed randomly among N nodes with the constraint that there must be B basal nodes. We also take the directed configuration ensemble of networks, which restricts both the in- and out-degrees of each node to specified values, and generate heavy-tailed (scale-free) networks by drawing those values from independent distributions $p(k) \sim k^{-\gamma}$, for $k = k^{in}$ and $k = k^{out}$. The left panel of Fig. S1 shows the mean trophic level $\langle s \rangle$ obtained numerically from Erdős-Rényi networks and heavy-tailed networks, as described, for varying proportions of basal nodes. This is plotted against the corresponding values given by Eq. (8) in each case. Similarly, the right panel of Fig. S1 shows the incoherence parameter q against the value given by Eq. (10), for the same networks. In both cases, the results fall very close to the $f(x) = x$ line, suggesting that the trophic structure of random graphs, regardless of their degree heterogeneity, is

well approximated by the basal ensemble we have defined above.

2 Networks with self-cycles

In the main text we ignore self-edges (cycles of length one) in those networks which exhibit them; this is mainly because self-edges are not reported in all networks, the nature of self-interaction often being considered fundamentally different to that of inter-element interaction. However, for completeness we also compute the values of the leading eigenvalue λ_1 , and of the loop exponent τ , defined as

$$\tau = \ln \alpha + \frac{1}{2\bar{q}^2} - \frac{1}{2q^2}, \quad (13)$$

when self-edges are allowed, and display these values in Fig. S2 (compare with Fig. 1 of the main text). We can observe that the good fit to the expression

$$\overline{\lambda_1} = e^\tau \quad (14)$$

– Eq. (14) of the main text, where $\overline{\lambda_1}$ is the coherence ensemble expectation for λ_1 – is not significantly affected by the inclusion of self-edges. The main difference is that several of the food webs which have $\lambda_1 = 0$ when self-edges are excluded now have $\lambda_1 = 1$, as a result of cannibalism.

3 Network data

In the main text we assess the validity of our analytical results through comparisons with a set of empirically-derived directed networks. Most of these are available online, but a few of them were shared with us in private correspondence. Below we list the most relevant details for each of these networks. Table 1 is for 42 food webs, Table S2 lists eight gene regulatory networks, and Table S3 contains information on seven metabolic networks. Table S4 is for six networks of various other kinds: the neural network of *C. elegans*, a P2P file sharing network, two networks of trade between nations (one of basic manufactured good and the other of minerals) and one of concatenated English words in the book *Green Eggs and Ham*, by Dr. Seuss; the last of these was obtained from the original text for this work [3]. The adjacency matrices of these networks are available at:

<http://www2.warwick.ac.uk/fac/sci/math/people/staff/sjohnson>

or upon request from the authors.

For each network, we list the number of nodes N , the number of basal nodes B , and the mean degree $\langle k \rangle$; the incoherence parameter q and its ratio to the expected value \bar{q} (low ratios mean more coherent networks than randomly expected); the mean trophic level $\langle s \rangle$ and the k^{in} - k^{out} correlation parameter α , both normalised by their expected values; the loop exponent τ , whose sign determines whether a network is in the high or low feedback regimes (see main text); the leading eigenvalue, λ_1 , of the adjacency matrix; and, finally, references to the sources of the data.

Figure S3 shows three example networks, one from each class: the Ythan Estuary food web [4], a metabolic network derived for *Chlamydia pneumoniae* [5], and a gene regulatory network derived for *E. coli* [6]. The height of each

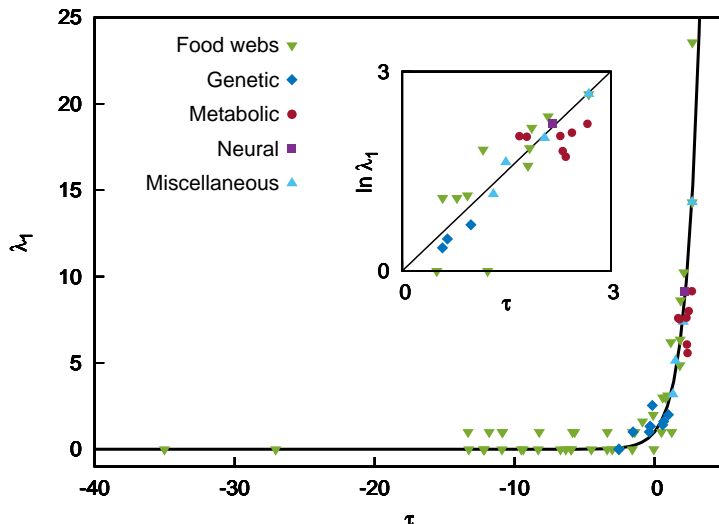


Figure S 2. Leading eigenvalues λ_1 of several directed networks when self-edges are not excluded, against τ as given by Eq. (13); symbols indicate food webs (green down-pointing triangles), gene regulatory networks (dark blue diamonds), metabolic networks (burgundy circles), a neural network (purple square), and other miscellaneous networks (light blue up-pointing triangles). Line: Expected leading eigenvalue $\bar{\lambda}_1$ in the coherence ensemble, as given by Eq. (14). Inset: Semi-log version of the positive quadrant of the main panel (Pearson's correlation coefficient: $r^2 = 0.80$). Compare with Fig. 2 of the main text, for which self-edges are excluded. Details for each network, including references, are listed in Tables S1, S2, S3 and S4 (though note that in these tables self-edges are excluded).

node on the vertical axis is proportional to its trophic level, and this visualisation is enough to show that the trophic structures of these systems can be highly informative. Note, in particular, that a network can display significant trophic coherence (the Ythan Estuary food web has $q/\tilde{q} = 0.15$) without all its nodes falling into clearly defined trophic levels, while it is also possible to be almost bipartite, as in the case of *E. coli*'s gene regulatory network, yet be less significantly coherent as compared to the random expectation ($q/\tilde{q} = 0.88$). These examples show that in order to determine which regime (of high or low feedback) a given system belongs to, it is insufficient to look only at trophic coherence: one must compute the loop exponent τ .

Food web	N	B	$\langle k \rangle$	q	q/\tilde{q}	$\langle s \rangle/\tilde{s}$	$\alpha/\tilde{\alpha}$	τ	λ_1	Ref.
Benguela Current	29	2	6.76	0.69	0.15	0.17	0.69	0.50	2	[7]
Berwick Stream	77	35	3.12	0.18	0.53	1.05	1.06	-12.21	0	[8, 9, 10]
Blackrock Stream	86	49	4.36	0.19	0.57	1.02	1.27	-9.51	0	[8, 9, 10]
Bridge Brook Lake	25	8	5.08	0.53	0.36	0.68	0.72	-0.53	1	[11]

Broad Stream	94	53	6.00	0.14	0.49	1.05	1.20	-20.10	0	[8, 9, 10]
Canton Creek	102	54	6.82	0.15	0.57	1.01	1.22	-14.52	0	[12]
Caribbean Reef	50	3	10.70	0.94	0.33	0.36	0.96	1.73	7.80	[13]
Cayman Islands	242	10	15.55	0.77	0.24	0.30	0.51	1.22	0	[14]
Catlins Stream	48	14	2.29	0.20	0.41	0.98	1.00	-10.90	0	[8, 9, 10]
Chesapeake Bay	31	5	2.16	0.45	0.33	0.73	0.90	-1.81	0	[15, 16]
Coachella Valley	29	3	8.38	1.20	0.48	0.47	0.91	1.63	5.48	[17]
Coweeta 1	58	28	2.17	0.30	0.64	1.00	1.08	-3.39	0	[8, 9, 10]
Coweeta 17	71	38	2.08	0.24	0.60	1.00	1.25	-5.94	0	[8, 9, 10]
Dempsters (Au)	83	46	4.99	0.21	0.57	1.08	1.02	-7.42	0	[8, 9, 10]
Dempsters (Sp)	93	50	5.78	0.13	0.38	1.07	1.11	-27.07	0	[8, 9, 10]
Dempsters (Su)	107	50	9.02	0.27	0.57	1.05	1.04	-3.51	0.01	[8, 9, 10]
El Verde Rainforest	155	28	9.72	1.01	0.45	0.49	1.21	2.09	10.12	[18]
German Stream	84	48	4.19	0.20	0.47	1.02	1.10	-9.35	0	[8, 9, 10]
Healy Stream	96	47	6.60	0.22	0.53	1.03	1.12	-6.34	0	[8, 9, 10]
Kyeburn Stream	98	58	6.42	0.18	0.62	1.02	1.18	-9.39	0	[8, 9, 10]
LilKyeburn Stream	78	42	4.81	0.23	0.53	1.01	1.10	-5.97	0	[8, 9, 10]
Little Rock Lake	92	12	10.7	0.67	0.22	0.25	0.77	1.06	5.66	[19]
Lough Hyne	349	49	14.62	0.60	0.37	0.59	0.63	0.85	2.56	[20, 21]
Martins Stream	105	48	3.27	0.32	0.58	0.99	1.26	-2.56	0	[8, 9, 10]
Narrowdale Stream	71	28	2.17	0.23	0.50	0.98	1.17	-7.45	0	[8, 9, 10]
NE Shelf	79	2	17.44	0.73	0.13	0.10	0.71	1.57	4.32	[22]
North Col Stream	78	25	3.09	0.28	0.52	0.98	1.36	-4.52	0	[8, 9, 10]
Powder Stream	78	32	3.44	0.22	0.47	0.99	1.12	-8.32	0	[8, 9, 10]
Scotch Broom	85	1	2.58	0.40	0.14	0.30	1.20	-2.08	0	[23]
Skipwith Pond	25	1	7.56	0.61	0.15	0.16	0.64	0.20	2	[24]
St Marks Estuary	48	6	4.54	0.63	0.37	0.63	1.02	0.26	0	[25]
St Martin Island	42	6	4.88	0.59	0.32	0.54	0.79	-0.05	0.01	[26]
Stony Stream	109	61	7.59	0.15	0.55	1.03	1.16	-14.66	0	[27]
Stony Stream 2	112	63	7.41	0.15	0.55	1.04	1.18	-14.72	0	[8, 9, 10]
Sutton (Au)	80	49	4.19	0.15	0.66	1.08	1.28	-13.27	0	[8, 9, 10]
Sutton (Sp)	74	50	5.28	0.10	0.56	1.11	1.15	-35.01	0	[8, 9, 10]
Sutton (Su)	87	63	4.87	0.28	0.89	1.19	0.52	-1.59	0	[8, 9, 10]
Troy Stream	77	40	2.35	0.19	0.37	1.01	1.14	-12.16	0	[8, 9, 10]
UK Grassland	61	8	1.59	0.40	0.18	0.42	0.63	-3.03	0	[28]
Venlaw Stream	66	30	2.83	0.23	0.54	1.06	1.35	-6.72	0	[8, 9, 10]
Weddel Sea	483	61	31.71	0.72	0.55	0.75	1.17	2.63	22.91	[29]
Ythan Estuary	82	5	4.77	0.42	0.15	0.28	0.93	-1.32	1	[4]

Table S 1. Details of 42 food webs used in the main text. Columns are for number of nodes N , number of basal nodes B , mean degree $\langle k \rangle$, and incoherence parameter q ; ratios of q , mean trophic level $\langle s \rangle$, and correlation parameter α to their expected values in the basal ensemble; loop exponent τ , leading eigenvalue λ_1 , and references to the data sources. Many of the data are available online at:

https://www.nceas.ucsb.edu/interactionweb/html/thomps_towns.html

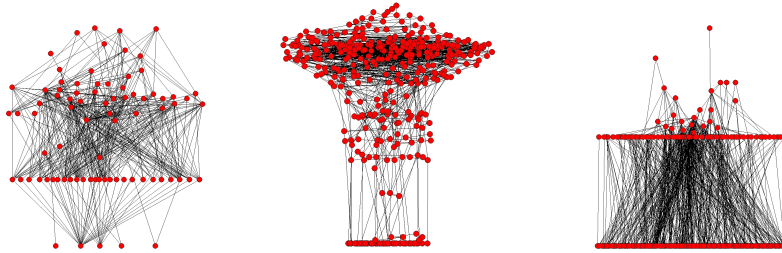


Figure S 3. Empirical networks display a rich diversity of trophic structures. Examples of three kinds of network are plotted here in such a way that the height of each node on the vertical axis is proportional to its trophic level (the scale used for each network is different because of the disparity in mean trophic level). Left: Ythan Estuary food web [4], which is significantly more trophically coherent than the random expectation ($q/\tilde{q} = 0.147$) and has no significant $k^{in}-k^{out}$ correlations ($\alpha/\tilde{\alpha} = 0.935$); it is in the negative τ regime: $\tau = -1.319$. Centre: A network derived from observations of the *Chlamydia pneumoniae* metabolism [5], which is significantly less trophically coherent than the random expectation ($q/\tilde{q} = 1.621$), and has positive $k^{in}-k^{out}$ correlations ($\alpha/\tilde{\alpha} = 2.550$); it is in the positive τ regime: $\tau = 1.686$. Right: A network derived from gene regulation in *E. coli* [6], which is only slightly more trophically coherent than the random expectation ($q/\tilde{q} = 0.878$) and has no significant $k^{in}-k^{out}$ correlations ($\alpha/\tilde{\alpha} = 0.938$); it is in the negative τ regime: $\tau = -2.543$. Details for each network, including references, are listed in the tables of SI.

Gene regulatory network	N	B	$\langle k \rangle$	q	q/\tilde{q}	$\langle s \rangle/\tilde{s}$	$\alpha/\tilde{\alpha}$	τ	λ_1	Ref.
Human (healthy)	4071	4004	2.08	0.08	0.99	1.00	0.99	-1.54	1	[30, 31]
Human (cancer)	4049	3967	2.89	0.08	1.00	1.00	1.07	-0.16	2.54	[30, 31]
<i>E. coli</i> (Salgado)	1470	1316	1.98	0.23	1.03	1.00	1.21	0.65	1.62	[32, 31]
<i>E. coli</i> (Thieffry)	418	312	1.24	0.27	0.88	1.01	0.94	-2.54	0	[6, 33]
<i>S. cerevisiae</i> (Harbison)	2933	2764	2.10	0.17	0.98	1.00	1.29	-0.38	1	[34, 31]
<i>S. cerevisiae</i> (Costanzo)	688	557	1.57	0.25	1.04	1.00	0.81	-0.31	1.32	[35, 33]
<i>P. aeruginosa</i>	691	606	1.43	0.30	1.00	1.03	1.94	0.58	1.41	[36, 31]
<i>M. tuberculosis</i>	1624	1542	1.95	0.17	1.02	1.00	1.24	0.99	2.00	[37, 31]

Table S 2. Details of eight gene regulatory networks (GRN) used in the main text. The *E. coli* (Salgado) and Yeast (Harbison) are available online at: <http://wvs.weizmann.ac.il/mcb/UriAlon/download/collection-complex-networks>. The others were shared with us by Luca Albergante, and some of them can be obtained from various websites: <http://regulondb.ccg.unam.mx/> (*E. coli*, Salgado); http://younglab.wi.mit.edu/regulatory_code (Yeast, Harbison); <http://www.genome.gov/ENCODE/> (Human, both the non-cancer GM12878 cell line and the K562 leukaemia cell line). Columns as in Table S1.

Metabolic network	N	B	$\langle k \rangle$	q	q/\bar{q}	$\langle s \rangle/\bar{s}$	$\alpha/\bar{\alpha}$	τ	λ_1	Ref.
<i>A. fulgidus</i>	1267	36	2.38	13.79	1.88	2.06	4.34	2.35	7.62	[5]
<i>M. thermoautotrophicum</i>	1111	30	2.43	12.17	1.77	1.90	4.08	2.31	7.59	[5]
<i>M. jannaschii</i>	1081	32	2.40	12.47	1.86	1.98	4.00	2.27	7.53	[5]
<i>C. pneumoniae</i>	386	20	2.05	8.98	1.62	1.71	2.55	1.69	5.57	[5]
<i>C. trachomatis</i>	446	19	2.11	11.77	1.95	2.02	2.77	1.79	6.07	[5]
<i>S. cerevisiae</i> (yeast)	1510	43	2.54	14.61	1.73	1.82	5.54	2.66	9.15	[5]
<i>C. elegans</i>	1172	40	2.44	13.29	1.86	2.04	4.60	2.44	8.00	[5]

Table S 3. Details of seven metabolic networks used in the main text, downloaded from <http://www3.nd.edu/~networks/resources.htm>. Columns as in Table S1.

Network (miscellaneous)	N	B	$\langle k \rangle$	q	q/\bar{q}	$\langle s \rangle/\bar{s}$	$\alpha/\bar{\alpha}$	τ	λ_1	Ref.
Neural (<i>C. elegans</i>)	297	3	7.90	1.49	0.42	0.39	1.42	2.17	9.15	[38, 39]
P2P (Gnutella 2008)	6301	3836	3.30	0.98	0.98	1.00	1.07	1.49	5.12	[40, 41]
Trade (manufactured goods)	24	2	12.92	4.24	1.14	1.14	1.10	2.68	14.3	[42]
Trade (minerals)	24	3	5.63	4.04	1.02	0.97	1.28	2.05	7.38	[42]
Words	50	16	2.02	2.04	1.01	1.16	1.55	1.31	3.17	[3]

Table S 4. Details of six other networks used in the main text. The network of words was obtained for this work from *Green Eggs and Ham* [3], and is available upon request from s.johnson.2@warwick.ac.uk. The other data can be found on various websites:
<http://www-personal.umich.edu/~mejn/netdata/> (neural network);
<https://snap.stanford.edu/data/p2p-Gnutella08.html> (P2P network);
and <http://vlado.fmf.uni-lj.si/pub/networks/data/esna/metalWT.htm> (trade networks). Columns as in Table S1

3.1 Green Eggs and Ham words network

We obtained the network of concatenated words from Dr Seuss’s masterpiece *Green Eggs and Ham* in the following way [3]. Every node in the text was assigned a node, and a directed edge $a_{ij} = 1$ was placed whenever word i preceded word j in a sentence. Figure S4 displays this network in such a way that the height of each node on the y-axis is proportional to its trophic level. Note that the arrows we have placed between nodes in this visualisation are from the preceding word to the succeeding one, so that one can obtain sentences (some of them grammatically correct) by following the arrows. We have used colours to indicate syntactic function, as described in the caption.

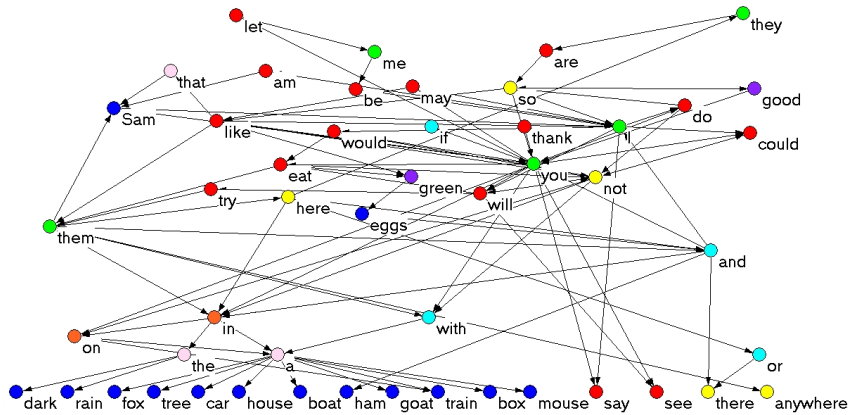


Figure S 4. Network of concatenated words from *Green Eggs and Ham*, by Dr Seuss [3]. The height of each word is proportional to its trophic level. Colours indicate syntactic function; from lowest to highest mean trophic level: nouns (blue), prepositions and conjunctions (cyan), determiners (pink), adverbs (yellow), pronouns (green), verbs (red), and adjectives (purple). When a word has more than one function, the one most common in the text is used.

References

- [1] H. Kim, C. I. Del Genio, K. E. Bassler, and Z. Toroczkai, “Constructing and sampling directed graphs with given degree sequences,” *New J. Phys.*, vol. 14, p. 023012, 2012.
- [2] S. Levine, “Several measures of trophic structure applicable to complex food webs,” *J. Theor. Biol.*, vol. 83, pp. 195–207, 1980.
- [3] Dr Seuss, *Green Eggs and Ham*. Random House, 1960.
- [4] M. Huxham, S. Beaney, and D. Raffaelli, “Do parasites reduce the chances of triangulation in a real food web?,” *Oikos*, vol. 76, pp. 284–300, 1996.
- [5] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, “The large-scale organization of metabolic networks,” *Nature*, vol. 407, no. 6804, pp. 651–654, 2000.
- [6] D. Thieffry, A. M. Huerta, E. Pérez-Rueda, and J. Collado-Vides, “From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in escherichia coli,” *Bioessays*, vol. 20, no. 5, pp. 433–440, 1998.
- [7] P. Yodzis, “Local trophodynamics and the interaction of marine mammals and fisheries in the benguela ecosystem,” *Journal of Animal Ecology*, vol. 67, no. 4, pp. 635–658, 1998.

- [8] R. M. Thompson and C. R. Townsend, “Impacts on stream food webs of native and exotic forest: An intercontinental comparison,” *Ecology*, vol. 84, pp. 145–161, 2003.
- [9] R. M. T. and C. R. Townsend, “Energy availability, spatial heterogeneity and ecosystem size predict food-web structure in stream,” *Oikos*, vol. 108, p. 137–148, 2005.
- [10] Townsend, Thompson, McIntosh, Kilroy, Edwards, and Scarsbrook, “Disturbance, resource supply, and food-web architecture in streams,” *Ecology Letters*, vol. 1, no. 3, pp. 200–209, 1998.
- [11] K. Havens, “Scale and structure in natural food webs,” *Science*, vol. 257, no. 5073, pp. 1107–1109, 1992.
- [12] Townsend, Thompson, McIntosh, Kilroy, Edwards, and Scarsbrook, “Disturbance, resource supply, and food-web architecture in streams,” *Ecology Letters*, vol. 1, no. 3, pp. 200–209, 1998.
- [13] S. Opitz, “Trophic interactions in Caribbean coral reefs,” *ICLARM Tech. Rep.*, vol. 43, p. 341, 1996.
- [14] J. Bascompte, C. Melián, and E. Sala, “Interaction strength combinations and the overfishing of a marine food web,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 15, pp. 5443–5447, 2005.
- [15] R. E. Ulanowicz and D. Baird, “Nutrient controls on ecosystem dynamics: the chesapeake mesohaline community,” *Journal of Marine Systems*, vol. 19, no. 1–3, pp. 159 – 172, 1999.
- [16] L. G. Abarca-Arenas and R. E. Ulanowicz, “The effects of taxonomic aggregation on network analysis,” *Ecological Modelling*, vol. 149, no. 3, pp. 285 – 296, 2002.
- [17] G. Polis, “Complex trophic interactions in deserts: an empirical critique of food-web theory,” *Am. Nat.*, vol. 138, pp. 123–125, 1991.
- [18] R. B. Waide and W. B. R. (eds.), *The Food Web of a Tropical Rainforest*. Chicago: University of Chicago Press, 1996.
- [19] N. D. Martinez, “Artifacts or attributes? Effects of resolution on the Little Rock Lake food web,” *Ecol. Monogr.*, vol. 61, pp. 367–392, 1991.
- [20] J. Riede, U. Brose, B. Ebenman, U. Jacob, R. Thompson, C. Townsend, and T. Jonsson, “Stepping in Elton’s footprints: a general scaling model for body masses and trophic levels across ecosystems,” *Ecology Letters*, vol. 14, pp. 169–178, 2011.
- [21] A. Eklöf, U. Jacob, J. Kopp, J. Bosch, R. Castro-Urgal, B. Dalsgaard, N. Chacoff, C. deSassi, M. Galetti, P. Guimaraes, S. Lomáscolo, A. Martín González, M. Pizo, R. Rader, A. Rodrigo, J. Tylianakis, D. Vazquez, and S. Allesina, “The dimensionality of ecological networks,” *Ecology Letters*, vol. 16, pp. 577–583, 2013.

- [22] J. Link, “Does food web theory work for marine ecosystems?,” *Mar. Ecol. Prog. Ser.*, vol. 230, pp. 1–9, 2002.
- [23] J. Memmott, N. D. Martinez, and J. E. Cohen, “Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural food web,” *J. Anim. Ecol.*, vol. 69, pp. 1–15, 2000.
- [24] P. H. Warren, “Spatial and temporal variation in the structure of a freshwater food web,” *Oikos*, vol. 55, pp. 299–311, 1989.
- [25] R. R. Christian and J. J. Luczkovich, “Organizing and understanding a winter’s Seagrass foodweb network through effective trophic levels,” *Ecol. Model.*, vol. 117, pp. 99–124, 1999.
- [26] L. Goldwasser and J. A. Roughgarden, “Construction of a large Caribbean food web,” *Ecology*, vol. 74, pp. 1216–1233, 1993.
- [27] Townsend, Thompson, McIntosh, Kilroy, Edwards, and Scarsbrook, “Disturbance, resource supply, and food-web architecture in streams,” *Ecology Letters*, vol. 1, no. 3, pp. 200–209, 1998.
- [28] N. D. Martinez, B. A. Hawkins, H. A. Dawah, and B. P. Feifarek, “Effects of sampling effort on characterization of food-web structure,” *Ecology*, vol. 80, p. 1044–1055, 1999.
- [29] U. Jacob, A. Thierry, U. Brose, W. Arntz, S. Berg, T. Brey, I. Fetzer, T. Jonsson, K. Mintenbeck, C. Möllmann, O. Petchey, J. Riede, and J. Dunne, “The role of body size in complex food webs,” *Advances in Ecological Research*, vol. 45, pp. 181–223, 2011.
- [30] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, *et al.*, “Architecture of the human regulatory network derived from encode data,” *Nature*, vol. 489, no. 7414, pp. 91–100, 2012.
- [31] L. Albergante, J. J. Blow, and T. J. Newman, “Buffered Qualitative Stability explains the robustness and evolvability of transcriptional networks,” *eLife*, vol. 3, p. e02863, 2014.
- [32] H. Salgado, M. Peralta-Gil, S. Gama-Castro, A. Santos-Zavaleta, L. Muñoz-Rascado, J. S. García-Sotelo, V. Weiss, H. Solano-Lira, I. Martínez-Flores, A. Medina-Rivera, *et al.*, “Regulondb v8. 0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more,” *Nucleic Acids Research*, vol. 41, no. D1, pp. D203–D213, 2013.
- [33] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, “Superfamilies of evolved and designed networks,” *Science*, vol. 303, no. 5663, pp. 1538–1542, 2004.
- [34] C. T. Harbison, D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J.-B. Tagne, D. B. Reynolds, J. Yoo, *et al.*, “Transcriptional regulatory code of a eukaryotic genome,” *Nature*, vol. 431, no. 7004, pp. 99–104, 2004.

- [35] M. C. Costanzo, M. E. Crawford, J. E. Hirschman, J. E. Kranz, P. Olsen, L. S. Robertson, M. S. Skrzypek, B. R. Braun, K. L. Hopkins, P. Kondu, *et al.*, “YpdTM, pombepdTM and wormpdTM: model organism volumes of the bioknowledgeTM library, an integrated resource for protein information,” *Nucleic Acids Research*, vol. 29, no. 1, pp. 75–79, 2001.
- [36] E. Galán-Vásquez, B. Luna, and A. Martínez-Antonio, “The regulatory network of *pseudomonas aeruginosa*,” *Microb Inform Exp*, vol. 1, no. 1, pp. 3–3, 2011.
- [37] J. Sanz, J. Navarro, A. Arbués, C. Martín, P. C. Marijuán, and Y. Moreno, “The transcriptional regulatory network of *mycobacterium tuberculosis*,” *PLoS one*, vol. 6, no. 7, p. e22178, 2011.
- [38] J. G. White, E. Southgate, J. N. Thompson, and S. Brenner, “The structure of the nervous system of the nematode *caenorhabditis elegans*,” *Phil. Trans. R. Soc. London*, vol. 314, pp. 1–340, 1986.
- [39] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, pp. 440–442, 1998.
- [40] J. Leskovec, J. Kleinberg, and C. Faloutsos, “Graph evolution: Densification and shrinking diameters,” *ACM Transactions on Knowledge Discovery from Data*, vol. 1, 2007.
- [41] M. Ripeanu, I. Foster, and A. Iamnitchi, “Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design,” *IEEE Internet Computing Journal*, 2002.
- [42] W. de Nooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press, 2004.