Many complex networks depend upon biological entities for their preservation. Such entities, from human cognition to evolution, must first encode and then replicate those networks under marked resource constraints. Networks that survive are those that are amenable to constrained encoding, or, in other words, compressible. But how compressible is a network? And what features make one network more compressible than another? Here we answer these questions by modeling networks as information sources before compressing them using rate-distortion theory. Each network yields a unique rate-distortion curve, which specifies the minimal amount of information that remains at a given scale of description. A natural definition then emerges for the compressibility of a network: the amount of information that can be removed via compression, averaged across all scales. Analyzing an array of real and model networks, we demonstrate that compressibility increases with two common network properties: transitivity (or clustering) and degree heterogeneity. These results indicate that hierarchical organization – which is characterized by modular structure and heavy-tailed degrees – facilitates compression in complex networks. Generally, our framework sheds light on the interplay between a network’s structure and its capacity to be compressed, enabling investigations into the role of compression in shaping real-world networks.

Information theory | Complex networks | Rate-distortion | Compression

Complex networks are often represented in biology, and thereby utilized and replicated by biological systems. The brain represents language (1), knowledge (2), music (3), social (4, 5), and transportation networks (6); the human mind uses those representations to engage in linguistic communication, build on existing understanding, sing a victorious melody, strengthen a valuable friendship, and walk the covered holloways (7). Similarly, biological networks among molecular and cellular components are represented by genetic material (8–11); and evolution uses those representations to propagate precise network topologies in a surviving species. From brains to genes, the biological materials that represent complex networks operate under marked constraints on time, energy, metabolism, and physical extent, among others. Such constraints determine which networks persist into the future; in particular, those whose topology can be efficiently encoded. These shared constraints raise a fundamental question: How does the structure of a network facilitate efficient encodings?

Encoding a network (indeed, encoding any piece of information) involves a natural trade-off between simplicity and accuracy. One could construct a simple representation that omits the fine-scale details of a network. Or one could build a representation that captures a network’s intricate structure, but is complicated and unwieldy. An efficient encoding strikes an optimal balance between simplicity and accuracy; that is, it is a compression (12, 13). In fact, compression – a foundational branch of information theory – has provided key insights into optimal network representations, yielding principled algorithms for constructing coarse-grained maps of complex systems (14–16).

Building upon this progress, here we investigate how the structure of complex networks facilitates compression. Intuitively, just as natural images are easier to compress than white noise due to their visual patterns and regularities, so too should networks with strong structural regularities be more compressible than random networks. But do homogeneous topologies, such as those found in lattice-like networks, make systems more compressible, or is compression facilitated by the hierarchical organization found in many real networks? To answer these questions, here we develop a framework for quantifying the compressibility of complex networks. Applying our framework to several real and model networks, we identify specific network features that facilitate compression. Together, these results elucidate how a network’s topology impacts its compressibility, and suggest that many real-world networks may be shaped by the pressure to be compressed.

Rate-distortion theory of network clustering

In compression (13), one begins with an information source, a sequence of items that defines the object of interest. In the
A Random walk

Clustering

H(x) = 2 bits

yt
yt
yt

I(x,y) = 0 bits

I(x,y) = 0.52 bits

I(x,y) = 2 bits

15 clusters

3 clusters

1 cluster

Scale S

Information rate R (bits)

More compressible

Less compressible

Fig. 1. Rate-distortion theory of random walks on networks. (A) A simple network with N = 15 nodes, each with constant degree k = 4. A random walk generates information at a rate H(x) = 2 bits. (B) Network clusterings across various scales of description. For n = 15 clusters, each containing its own node, the sequence communicates I(x,y) = H(x) = 2 bits of information (Top). For n = 3 clusters, each corresponding to one of the three modules in the original network, the information rate is I(x,y) = 0.52 bits (Middle). For n = 1 cluster containing the entire network, the sequence no longer communicates information (Bottom). (C) Schematic of the optimal information rate R as a function of the scale of description S for networks that are either more compressible (black) or less compressible (grey). For more compressible networks, one can achieve a lower information rate at a given scale of description (vertical line), and one can achieve a more fine-grained description for a given information rate (horizontal line).

context of networks, a logical choice for the information source is a random walk, which contains all of the information in a network and nothing more (14). One then seeks to reduce the amount of information in the sequence by constructing a coarse-grained representation of the network. We remark that the lossless compression of random walks has provided important information-theoretic perspectives on the problem of community detection (14). By contrast, here we study lossy compression, which allows us to build upon the elegant foundation of rate-distortion theory and thereby develop an intuitive definition for compressibility.

Compressing random walks. To see how compression unfolds in practice, consider the network in Fig. 1A. A random walk on the network defines a sequence of nodes \( \mathbf{x} = (x_1, x_2, \ldots) \), with each node transitioning to one of its four neighbors uniformly at random. The rate at which this sequence generates information is given by the entropy \( H(x) \), which (because there are four possible nodes at each step) equals 2 bits. To reduce the amount of information in the sequence, we can construct a coarse-grained representation by clustering nodes together (14–16). This clustering yields a new sequence \( \mathbf{y} = (y_1, y_2, \ldots) \), where \( y_i \) is the cluster containing node \( x_i \) (Fig. 1B), which communicates information at a rate equal to the mutual information \( I(x, y) = H(y) - H(y|x) \) (12, 13, 15, 16). If the clusters are chosen deterministically, as is common (4, 14, 17), then the conditional entropy \( H(y|x) \) vanishes, and the information rate simplifies to the entropy of the clustered sequence, \( I(x, y) = H(y) \).

Consider, for example, a trivial clustering in which each node belongs to its own cluster (Fig. 1B, Top). In this case, we maintain a complete description of the network, but we have not reduced the information rate, since \( I(x, y) = H(x) = 2 \) bits. By contrast, consider the opposite setting in which all nodes belong to the same large cluster (Fig. 1B, Bottom). Now we have reduced the information rate to zero \( I(x, y) = 0 \) bits, but all details about the network structure have been lost. Between these two extremes lies a range of clusterings (such as that in Fig. 1B, Middle), each inducing its own information rate and yielding a unique distortion of the network structure.

Scale as a measure of distortion. Building representations that strike an optimal balance between minimizing information rate while also minimizing distortion is precisely the purview of rate-distortion theory (12, 13). As in any rate-distortion problem, one must choose a specific definition for the distortion of the object of interest. When clustering a network, a natural choice for the distortion presents itself: the scale of description. Specifically, for a network with \( N \) nodes and a clustering with \( n \) clusters, we define the scale to be \( S = \frac{1}{N/n} \) and achieve a more fine-grained description at a given scale of description (vertical line), and one can achieve a more fine-grained description for a given information rate (horizontal line).

Computing the rate-distortion curve of a network

Computing the rate-distortion curve \( R(S) \) of a network – in particular, doing so efficiently to enable applications to large systems – poses two distinct challenges. First, we must estimate the mutual information \( I(x, y) \) for different clusterings; and second, we must identify the clusterings that minimize
Fig. 2. Properties of optimal clusterings. (A) Upper bound (solid line) and lower bound (dashed line) on the optimal information rate $\bar{R}(S)$ as a function of the scale of description $S$ for Zachary’s karate club network (18). Across all scales, the optimal compression includes one large cluster, which we illustrate for $S = 0.25, 0.5,$ and 0.75 (Right). (B) Size of the largest cluster in a compression, normalized by the size of the network $N$, as a function of the scale $S$ for the real networks in Supporting Table 1. The median over real networks (solid line) matches the largest possible normalized cluster size, $(N − n + 1)/N = S$. (C) Illustration of edges within the one large cluster (blue), on the boundary of the cluster (purple), and outside the cluster (red). (D) Fraction of the $k_c$ edges emanating from the large cluster that either connect to nodes outside the cluster $1 − G_{cc}/k_c$ (purple) or remain within the cluster $G_{cc}/k_c$ (blue) as a function of the scale $S$. (E) Average degree of nodes inside (blue) and outside (red) the large cluster, normalized by the average degree of the network, as a function of the scale $S$. In panels (D) and (E), solid lines and shaded regions represent averages and one-standard-deviation error bars, respectively, over the real networks (Supporting Table 1), and dashed lines correspond to clusters with nodes selected at random.

this information rate across all scales.

Although estimating mutual information is generally difficult (19), the simplicity of our setup allows for tractable upper and lower bounds (see Materials and Methods). Of particular interest is the upper bound $\bar{I}(x, y) \geq I(x, y)$, which follows by approximating the clustered sequence $y$ as Markovian (a property that we note is not guaranteed, even though the original random walk $x$ is Markovian (13)). Rather than minimizing the information rate $I(x, y)$ directly, we instead minimize the upper bound $\bar{I}(x, y)$, yielding an upper bound $\bar{R}(S)$ on the rate-distortion curve. For simplicity, in what follows we often refer to $\bar{I}(x, y)$ as the information rate and $\bar{R}(S)$ as the rate-distortion curve.

To compute $\bar{R}(S)$ – that is, to find clusterings that minimize the information rate $\bar{I}(x, y)$ – we employ a greedy clustering algorithm. Beginning with $n = N$ clusters, each containing its own node, we combine the pair of clusters that yields the largest reduction in the information rate $\bar{I}(x, y)$. Repeating this agglomerative process across all scales $S$ (until only one cluster remains), we arrive at an estimate for the rate-distortion curve $\bar{R}(S)$. To speed up the calculation, rather than searching through all $\binom{n}{2}$ pairs of clusters at each step, we only consider a limited number of pairs chosen via principled heuristics (see Materials and Methods). In practice, not only do these heuristics enable applications to networks of approximately $10^8$ nodes, they also improve the accuracy of the rate-distortion estimates themselves (see Supporting Fig. 1).

We are now prepared to compute the rate-distortion curve for a specific system. In Fig. 2A, we plot the upper and lower bounds on the rate-distortion curve $\bar{R}(S)$ for Zachary’s karate club network (18). As is true for all networks (see Materials and Methods), the two bounds are exact at both the minimum scale $S = 1/N$ (when the information rate simply equals the entropy of random walks $H(x)$) and the maximum scale $S = 1$ (when the information rate is zero). Moreover, the two bounds remain close across all intermediate scales, demonstrating that the upper bound $\bar{R}(S)$ provides a good approximation to the true rate-distortion curve. To understand how the rate-distortion curve depends on the structure of a network, however, it helps to examine the properties of optimal compressions themselves.

Properties of optimal compressions

Using the framework developed above, we are ultimately interested in studying compression in real systems. The networks chosen for analysis span from communication networks (including semantic, language, and music networks) and information networks (including hyperlinks on the web and citations in science) to social networks, animal and protein interactions, transportation networks, and structural and functional connections in the brain (see Materials and Methods; Supporting...
Table 1. Although these networks encompass a wide range of systems bridging several orders of magnitude in size, they are all encoded biologically, either in genetic material or in the neural code.

**Emergence of one large cluster.** Computing the rate-distortion curve $R(S)$ for each of the above networks, we find a striking consistency in the structure of optimal compressions. As can be observed in Zachary’s karate club (Fig. 2A, Right), rather than dividing the network into multiple clusters of moderate size, optimal compressions tend to comprise one large cluster containing $N - n + 1 = SN$ nodes and $n - 1$ minimal clusters each containing one node. In fact, among the networks studied, this tendency to form one large cluster is a nearly ubiquitous feature of optimal compressions (Fig. 2B).

We remark that the clustering that minimizes the information rate need not (and indeed, does not) provide a faithful characterization of a network’s community structure, as is the goal in community detection (14–16). Instead, we find that optimal compressions seek to identify the group of nodes that can be combined to maximally reduce the information rate. By dividing the network into two parts—the one inside the large cluster and the other outside—the challenge of compressing random walks thus resembles the graph partitioning problem (20), which has generated key insights about the modular structure of networks across scales (17). This simplification, in turn, allows us to develop analytic predictions about the properties of optimal compressions and the structures of compressible networks.

**Information rate of optimal compressions.** Although our framework is general, applying to any weighted, directed network (see Materials and Methods), in order to make analytic progress, here we focus on the special case of an unweighted, undirected network with adjacency matrix $G_{ij}$. For such a network, the entropy of random walks takes the simple form $H(x) = \frac{1}{2E} \sum_i k_i \log k_i$, where $k_i = \sum_j G_{ij}$ is the degree of node $i$, $E = \frac{1}{2} \sum_{i,j} G_{ij}$ is the number of edges in the network, and $\log(.)$ is base 2 such that information is measured in bits.

Now consider forming one large cluster $c$. One can show (see Materials and Methods) that the information rate of the clustered network is given by

$$\bar{I}(x, y) = \frac{1}{2E} \left[ \sum_{i \in c} k_i \log k_i + k_c \log k_c \right] - 2 \sum_{i \in c} G_{ic} \log G_{ic} + G_{cc} \log G_{cc},$$

where $k_c = \sum_{i \in c} k_i$ is the sum of the degrees of the nodes in $c$, $G_{ic} = \sum_{j \in c} G_{ij}$ is the number of edges connecting nodes in $c$ to a given node $i$, and $G_{cc} = \sum_{i \in c} G_{ij}$ is the number of edges connecting nodes within $c$.

**Information content of different edges.** Using Eq. 20, can we predict the properties of the optimal cluster $c'$? More broadly, can we anticipate the types of network topologies that facilitate compression? To answer these questions, it helps to group the edges in a network into three distinct categories (Fig. 2C): those connecting nodes within $c$, those connecting nodes outside of $c$, and those on the boundary of $c$ (connecting nodes within $c$ to nodes outside of $c$). We can gauge which type of edge is preferred over the others by comparing their contributions to the information rate (Eq. 20). An optimal compression will maximize the number of edges that are informationally preferred (contributing only weakly to the information rate), while limiting edges that are informationally costly.

For example, adding an edge within $c$ increases the information rate by $\Delta I_{\text{within}} \approx -\frac{1}{N} \left(2 \log k_c - 2 \log G_{cc}\right)$. By contrast, adding an edge on the boundary of $c$ (say, connecting $c$ to a node $i \notin c$) yields an increase of roughly $\Delta I_{\text{boundary}} \approx \frac{1}{N} \left(\log k_i + \log k_c - 2 \log G_{ic}\right)$. For a large cluster $c$, we have $k_c G_{cc} \gg k_i G_{ic}$, from which one can show that $\Delta I_{\text{within}} \lesssim \Delta I_{\text{boundary}}$ (see Supporting Information). Thus, edges within the large cluster are informationally preferred to those on the boundary, suggesting that optimal compressions will combine tightly-connected groups of nodes that are sparsely connected to the rest of the network. Indeed, in real networks, we find that among the $k_c$ edges emanating from the large cluster, the proportion $1 - G_{cc}/k_c$ that connects to the rest of the network is much smaller than chance (Fig. 2D). This proportion of edges leaving the cluster is a well-studied quantity, known as the conductance or Cheeger constant of a network (17). Thus, networks with low conductance—such as those with modular structure and strong transitivity (the tendency for nodes to form triangles, also known as clustering)—should be highly compressible (17, 21). This is our first hypothesis about the impact of network structure on compressibility.

We now consider an edge connecting two nodes $i$ and $j$ outside of $c$, which increases the information rate by approximately $\Delta I_{\text{outside}} \approx -\frac{1}{N} \left(\log k_i + \log k_j\right)$. As before, one can show that $\Delta I_{\text{within}} \lesssim \Delta I_{\text{outside}}$ (see Supporting Information), demonstrating that edges within the large cluster are informationally preferred to those outside the cluster. In turn, this preference for the large cluster to include as many edges as possible suggests that $c$ will favor high-degree nodes over low-degree nodes, which we confirm in real networks (Fig. 2E). This result leads to our second hypothesis: networks should be more compressible if they have heterogeneous (or heavily-tailed) degrees, containing “rich-clubs” of high-degree hub nodes (22, 23). Given the predictions that modular and heterogeneous topologies facilitate compression, we now propose a quantitative definition for the compressibility of a network.

**Quantifying network compressibility.** Intuitively, a network should be compressible if one can achieve a large reduction in the information rate at a given scale (Fig. 1C). However, rather than choosing a specific scale $S$ (equivalently, a specific number of clusters $n$), we would like our definition of compressibility to be a property of the network itself. We therefore define the compressibility of a network to be the amount of information that can be removed via compression, averaged across all scales,

$$C = H(x) - \frac{1}{N} \sum_{S} R(S).$$

Visually, the compressibility represents the area above a network’s rate-distortion curve (Fig. 3A). In practice, plugging our tractable upper bound on the rate-distortion curve $R(S)$ into Eq. 25 yields a lower bound $C$, which (for simplicity) we will refer to as compressibility.
Fig. 3. Quantifying compressibility. (A) The compressibility of a network (shaded region) is the area between the rate-distortion curve (solid line) and the entropy of random walks (dashed line). (B) A k-regular network, characterized only by the requirement that all nodes have constant degree k. (C) Rate-distortion curves $R(S)$ for k-regular networks with different degrees k. (D) Compressibility $C$ of k-regular networks versus degree k. In panels (C) and (D), solid lines and data points are averages over 50 randomly-generated networks, each of size $N = 10^5$, and dashed lines indicate analytic predictions (Eqs. 3 and 4). (E) Compressibility $C$ versus average degree for the real networks in Supporting Table 1. We note that average degree is plotted on a log scale. Dashed line indicates a logarithmic fit. For networks of size $N > 10^4$, data points and error bars represent means and standard deviations over 50 randomly-sampled subnetworks of $10^3$ nodes each (see Materials and Methods).

To make the notion of compressibility concrete, consider the class of random k-regular networks, which have no structure besides the requirement that the nodes have uniform degree k (Fig. 3B). For these lattice-like networks, one can derive (see Supporting Information) an analytic approximation to the rate-distortion curve,

$$R(S) \approx (1 - S)^2 \log k + S(1 - S) \log N - S \log S,$$  

which becomes accurate in the limit of large k (Fig. 3C). Noting that the entropy of a k-regular network is $\log k$, and approximating the average in Eq. 25 by an integral over S, we arrive at an analytic form for the compressibility,

$$C \approx \frac{2}{3} \log k - \frac{1}{6} \log N - \frac{1}{4 \ln 2}. \quad [4]$$

We confirm this prediction numerically (Fig. 3D), noting that the compressibility grows logarithmically with the degree k. Computing the compressibility of the real networks in Supporting Table 1, we find precisely the same logarithmic dependence on the average degree (Fig. 3E). Furthermore, we verify that this logarithmic dependence generalizes to directed versions of the networks (Supporting Fig. 3) and is not simply due to our clustering heuristics (Supporting Fig. 4). These results demonstrate that the compressibility of a network increases predictably with average degree. But how does compressibility depend on the topology of a complex network?

Impact of network structure on compressibility

Based on the properties of optimal compressions (Fig. 2), we hypothesized that the compressibility of a network should increase with both (i) transitivity and (ii) degree heterogeneity. These two features are frequently observed across an array of real-world networks, from social, scientific, and biological interactions (24–26) to the Internet (2), language (24), music (27), and the brain (28). Moreover, the combination of transitivity (with tightly-connected modules) and heterogeneous degrees (with well-connected hubs) defines hierarchical organization (24), which has been shown to support multi-scale representations of complex networks (29, 30) and enable efficient information processing in neural and communication systems (31, 32). In fact, when encoding information about the world, the brain itself often employs hierarchical representations (33–35). If correct, our hypotheses will lead to these perspectives a new outlook on the role of hierarchical structure: that it supports the efficient compression of complex networks.

To investigate the impact of transitivity and modular structure on compressibility, we consider a class of stochastic block networks (Fig. 3A), wherein nodes are grouped into communities of equal size and a specified fraction $f$ of the edges in the network connect nodes within the same community. Computing the rate-distortion curves of stochastic block networks, we find that increasing the fraction $f$ of within-community edges – that is, strengthening the modular structure and transitivity in the networks – decreases the rate-distortion curve $R(S)$ at intermediate scales S (Fig. 3B). We therefore find that compressibility increases with both modularity (Fig. 3C) and transitivity (Fig. 3D). Importantly, these results on stochastic block networks generalize to real networks, with increases in transitivity yielding significant improvements in network compressibility (Fig. 3E).

To examine the dependence of compressibility on degree heterogeneity, we study scale-free networks (Fig. 3F), which have heavy-tailed degree distributions $P(k) \sim k^{-\gamma}$ characterized by a power-law exponent $\gamma$ (23). As $\gamma$ decreases, accentuating the networks’ heavy-tailed structure, the rate-distortion curve $R(S)$ increases at small scales and decreases at intermediate and large scales (Fig. 3G). Both of these rate-distortion effects serve to improve the compressibility of scale-free networks (Fig. 3H). Moreover, rather than indirectly investigating the impact of heavy-tailed structure via the scale-free exponent $\gamma$, we can directly quantify the degree heterogeneity of a given network $h = \langle |k_i - k_j| \rangle / \langle k \rangle$, where $\langle |k_i - k_j| \rangle$ is the absolute difference in degrees averaged over all pairs of nodes and $\langle k \rangle$ is the
We find that the compressibility of scale-free networks grows linearly with degree heterogeneity (Fig. 3J), a result that generalizes to real networks (Fig. 3J). Furthermore, we confirm that the dependence of compressibility on both transitivity and degree heterogeneity extends to directed networks (Supporting Fig. 3) and is robust to our choice of clustering heuristics (Supporting Fig. 4).

Together, the above results demonstrate that transitivity and degree heterogeneity—the two defining features of hierarchical organization—increase the compressibility of complex networks. By focusing on specific families of networks, we also discover variations in compressibility that reflect a network’s function. Road networks, for example, exhibit the lowest transitivity and degree heterogeneity, and therefore the lowest compressibility, among the networks studied. This low compressibility is likely due to the fact that, unlike the other networks, road networks are confined to exist in two dimensions, severely constraining their topology (37). Besides road networks, we find that protein interactions have the lowest transitivity and brain networks have the lowest degree heterogeneity, leading both classes of networks to be relatively incompressible. Interestingly, these two families are unique among the networks studied in that they are only encoded genetically and need not be represented cognitively by a human or animal. By contrast, language networks are highly compressible, perhaps reflecting the primary function of language as a means for encoding and communicating information. Thus, although many networks are encoded biologically, the pressure for these encodings to be efficient manifests to varying degrees in different families of networks, yielding a...
spectrum of compressibilities.

Discussion

Complex networks perform an astonishing array of functions, which are supported by a multitude of topological structures. Many networks, however, are unified by a common constraint: that they rely on biological entities to encode them and pass them on. Encoding a network efficiently – that is, striking an optimal balance between simplicity and accuracy – requires compression, an insight that has provided information-theoretic perspectives on network structure (14–16). Naturally, some networks should be more compressible than others, with structural regularities enabling efficient representations across multiple scales. To investigate this hypothesis, here we introduce a rate-distortion theory of network compression (Fig. 1) and propose a quantitative definition for the compressibility of a network (Eq. 25; Fig. 3A). Applying our framework to a number of real and model networks, we demonstrate that network compressibility increases with both transitivity and degree heterogeneity (Fig. 3), two features that together characterize hierarchical organization (24).

The interplay between network structure and compressibility hints at one possible factor contributing to the hierarchical organization observed in many real networks (2, 24–29, 31, 32, 38, 39): that it enables efficient representations across scales. But do the actual encodings themselves – from representations in the minds of humans and animals to the information stored in the genetic code – take advantage of this hierarchical organization? Moreover, do the encodings employed in nature approach the limit of optimal efficiency specified by rate-distortion theory? Answering these questions will require new exciting investigations into the information-theoretic pressures and constraints on complex networks.

Materials and Methods

Bounding the information rate. Given a (possibly weighted, directed) network with adjacency matrix $G_{ij}$, the probability of one node $i$ transitioning to another node $j$ in a random walk is $P_{ij} = G_{ij}/k_i$, where $k_i = \sum_j G_{ij}$ is the (out) degree of node $i$ (Fig. 1A). The entropy of random walks is given by

$$H(x) = -\sum_i \pi_i \sum_j P_{ij} \log P_{ij}, \quad [5]$$

where $\pi_i$ is the stationary distribution defined by the condition $\pi = P \pi$ (which we note is uniquely defined if the network is strongly-connected and aperiodic). After clustering the network, we arrive at a new sequence $y = (y_1, y_2, \ldots)$, where $y_i$ is the cluster containing node $x_i$ (Fig. 1B). The information rate of this new sequence is given by the mutual information $I(x, y)$, which for deterministic clusterings (such as those considered here) is equivalent to the entropy $H(y)$. However, even though the random walk $x$ is Markovian (yielding a simple form for the entropy [Eq. 5]), the clustered sequence $y$ need not be (13), and thus it is generally difficult to derive an analytic form for $H(y)$.

Despite this hurdle, there exist simple bounds on the information rate $I(x, y) = H(y)$, summarized by the inequalities

$$H(y_{t+1} | x_t) \leq H(y) \leq H(y_{t+1} | y_t), \quad [6]$$

where $H(y_{t+1} | x_t)$ and $H(y_{t+1} | y_t)$ are the conditional entropies of $y_{t+1}$ on $x_t$ and $y_t$, respectively (13). These bounds are tight at the minimum scale $S = 1/N$, when each cluster contains one node and so $H(y) = H(x) = H(x_{t+1} | x_t)$. The bounds are also tight at the maximum scale $S = 1$, when there is one cluster and so $H(y) = H(y_{t+1} | x_t) = H(y_{t+1} | y_t) = 0$.

To compute the lower bound at intermediate scales, we begin with the conditional probability of node $i$ in the random walk $x$ transitioning to cluster $c$ in the clustered sequence $y$, $P_{ic} = \sum_j G_{ij}$. Then, the lower bound is given by

$$I(x, y) \geq I(x, y) = H(y_{t+1} | x_t) = -\sum_i \pi_i \sum_c P_{ic} \log P_{ic}, \quad [7]$$

where the second sum runs over all clusters $c$. Similarly, to compute the upper bound, we consider the probability of one cluster $c$ transitioning to another cluster $c'$,

$$P_{cc'} = \frac{1}{\pi_c} \sum_{i \in c} \pi_i \sum_{j \in c'} P_{ij}, \quad [8]$$

where $\pi_c = \sum_{i \in c} \pi_i$ is the stationary distribution over clusters. We then arrive at the following upper bound,

$$I(x, y) \leq I(x, y) = H(y_{t+1} | y_t) = -\sum c \pi_c \sum_{c'} P_{cc'} \log P_{cc'}, \quad [9]$$

where $\pi_y = \sum_{i \in c} \pi_i$ is Markovian. In practice, when estimating the optimal information rate for a network, we minimize the upper bound in Eq. 9 over clusterings, resulting in an upper bound $\bar{R}(S)$ on the rate-distortion curve.

The upper bound $I(x, y)$ simplifies significantly for unweighted, undirected networks. In this case, the stationary distribution is proportional to the node degrees $\pi_i = k_i/2E$, where $E = \frac{1}{2} \sum_{i,j} G_{ij}$ is the number of edges in the network. Additionally, the cluster transition probabilities take the form $P_{cc'} = G_{cc'}/k_c$, where $G_{cc'} = \sum_{i \in c} \sum_{j \in c'} G_{ij}$ is the induced network of clusters and $k_c = \sum_{i \in c} k_i$ is the sum of the degrees of the nodes in $c$. One can then manipulate Eq. 9 into the form

$$I(x, y) = \frac{1}{2E} \sum_{c \in \mathcal{C}} k_c \log k_c - \sum_{c \in \mathcal{C}} G_{cc'} \log G_{cc'}. \quad [10]$$

Under the further simplification of a clustering with one large cluster $c$ and $n - 1$ minimal clusters of one node each (Fig. 2), this upper bound can be fashionable into Eq. 20.

Clustering algorithm. To compute the rate-distortion curve $\bar{R}(S)$, we use an agglomerative clustering algorithm. Beginning with $n = N$ clusters (corresponding to the minimum scale $S = 1/N$), each containing an individual node, we iteratively consider pairs of clusters until we eventually arrive at one large cluster containing the entire network (corresponding to the maximum scale $S = 1$). At each step, we greedily select the pair of clusters to combine that minimizes the information rate $I(x, y)$ (Eq. 9). However, rather than searching through all $\binom{n}{2}$ pairs of clusters at each iteration (which would limit applications to small networks), we instead focus on a subset of $m$ pairs chosen through one of two heuristics.

The first heuristic, motivated by the observation that optimal clusterings tend to combine clusters with large degrees (Fig. 2B), selects the $m$ pairs of clusters $c$ and $c'$ with the largest combined stationary probabilities $\pi_c + \pi_{c'}$. For unweighted, undirected networks, we note that this choice is equivalent to selecting the pairs of clusters with the largest combined degrees, since $\pi_c + \pi_{c'} = \frac{1}{k_c} (k_c + k_{c'})$. The second heuristic, motivated by the fact that optimal compressions tend to form clusters with tight intra-cluster connectivity (Fig. 2D), selects the pairs of clusters $c$ and $c'$ with the largest combined joint transition probabilities $\pi_c P_{c'c} + \pi_{c'} P_{cc'}$. For unweighted, undirected networks, we remark that this second heuristic is equivalent to selecting the pairs of clusters with the largest number of connecting edges, since $\pi_c P_{c'c} + \pi_{c'} P_{cc'} = \frac{1}{k_c} (G_{cc'} + G_{c'c'})$. In practice, we consider $m = 100$ pairs of clusters at each iteration. In Supporting Fig. 1, we compare these two heuristics to the brute-force approach that searches through all pairs of clusters at each iteration of the clustering algorithm. In addition to significantly speeding up the algorithm, we find that these two heuristics often yield more accurate estimates of the rate-distortion curve $\bar{R}(S)$ than the brute-force implementation.
Network datasets. The networks analyzed in this paper are listed and described in Supporting Table 1. While we study unweighted, undirected versions of the networks in Figs. 2, 3E, 3E, and 3J, similar results hold for directed versions of the networks (Supporting Figs. 2 and 3). For networks of size $N \leq 10^3$, we perform analyses directly. For larger networks with $N > 10^3$, we analyze 50 subnetworks of $10^3$ nodes each. Each subnetwork is generated by performing a random walk beginning at a randomly-selected node until $10^3$ nodes have been reached. This sampling method has been shown to give accurate estimates of network statistics (40).

Data and code availability. The data analyzed in this paper and the code used to perform the analyses are openly available at github.com/ChrisWLynn/Network_compressibility.

ACKNOWLEDGMENTS. The authors thank Christopher Kroninger, Lia Papadopoulos, Dr. Pragya Srivastava, Mathieu Ouellet, and Dale Zhou for helpful feedback on earlier versions of this manuscript. C.W.L. acknowledges support from the James S. McDonnell Foundation 21st Century Science Initiative Understanding Dynamic and Multi-scale Systems - Postdoctoral Fellowship Award. The authors also acknowledge support from the John D. and Catherine T. MacArthur Foundation, the ISI Foundation, the Paul G. Allen Family Foundation, the Army Research Laboratory (W911NF-10-2-0022), the Army Research Office (Bassett-W911NF-14-1-0679, Falk-W911NF-18-1-0244, Grafton-W911NF-16-1-0474, DCIST- W911NF-17-2-0181), the Office of Naval Research, the National Institute of Mental Health (2-R01-DC-009209-11, R01-MH112847, R01-MH107235, R21-M MH-106790), and the National Science Foundation (NSF PHY-1554488, BCS-1631550, and NCS-FO-1926829).

CITATION DIVERSITY STATEMENT. Recent work in several fields of science has identified a bias in citation practices such that papers from women and other minorities are under-cited relative to the number of such papers in the field (41–46). Here we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, and other factors. We obtained predicted gender of the first and last author of each reference by using databases that store the probability of a name being carried by a woman (45, 47). By this measure (and excluding self-citations to the first and last authors of our current paper), our references contain 16% woman(first)/woman(last), 17% man/woman, 18% woman/man, and 50% man/man. This method is limited in that a) names, pronouns, and social media profiles used to construct the databases may not, in every case, be indicative of gender identity, and b) it cannot account for intersex, non-binary, or transgender people. Second, we obtained predicted racial/ethnic category of the first and last author of each reference by databases that store the probability of a first and last name being carried by an author of color (48, 49). By this measure (and excluding self-citations), our references contain 9% author of color(first)/author of color(last), 14% white author/author of color, 15% author of color/white author, and 62% white author/white author. This method is limited in that a) names, Census entries, and Wikipedia profiles used to make the predictions may not be indicative of racial/ethnic identity, and b) it cannot account for Indigenous and mixed-race authors, or those who may face differential biases due to the ambiguous racialization or ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.
Supporting Information

1. Introduction

In this Supporting Information, we provide extended analysis and discussion to support the results presented in the main text. In Sec. 2, we discuss the information costs of the different edge types in Fig. 2C in the main text. In Sec. 3, we derive an analytic approximation of the rate-distortion curve for k-regular networks (Eq. 3 and Fig. 3C in the main text). In Sec. 4, we discuss the different heuristics used to speed up the clustering algorithm and compare their estimates of optimal information rates with those of a brute-force implementation. In Sec. 5, we demonstrate that the central results from the main text generalize to directed networks. In Sec. 6, we show that the dependencies of compressibility on average degree, transitivity, and degree heterogeneity do not depend on the heuristics used to speed up the clustering algorithm. Finally, in Sec. 7, we list the real networks analyzed in this work and describe how we sample large networks.

2. Information content of different edges

When analyzing optimal clusterings (that is, clusterings that minimize the information rate $I(x,y)$, Eq. 9 in the main text), we find that they tend to consist of one large cluster containing $N - n + 1 = SN$ nodes and $n - 1$ clusters containing one node each (see Fig. 2A, B in the main text). This observation allows us to group the edges in a network into three categories (Fig. 2C in the main text): those connecting nodes within $c$, those connecting nodes outside of $c$, and those on the boundary of $c$ (connecting nodes within $c$ to nodes outside of $c$).

In order to predict the structure of the one large cluster $c$, we wish to compare the contributions of different edge types to the information rate $I(x,y)$. For an unweighted, undirected network with adjacency matrix $G_{ij}$, and a clustering with one large cluster $c$, the information rate can be written as

$$I(x,y) = \frac{1}{2E} \left( \sum_{i \in c} k_i \log k_i + k_c \log k_c - 2 \sum_{i \in c} G_{ic} \log G_{ic} - G_{cc} \log G_{cc} \right), \quad [11]$$

where the sums run over all nodes $i$ not in $c$, $E = \frac{1}{2} \sum_{i,j} G_{ij}$ is the number of edges in the network, $k_i = \sum_{j \in c} G_{ij}$ is the degree of node $i$, $k_c = \sum_{i \in c} k_i$ is the combined degrees of nodes in $c$, $G_{ic} = \sum_{j \in c} G_{ij}$ is the number of edges connecting a node $i$ to nodes in $c$, and $G_{cc} = \sum_{i,j \in c} G_{ij}$ is the number of edges connecting nodes within $c$ (see Materials and Methods in the main text).

Within versus boundary edges. If we add an edge between two nodes within $c$, then both $k_c$ and $G_{cc}$ increase by two, and the change in the information rate is given by

$$\Delta I^{\text{within}} = \frac{1}{2E} \left[ (k_c + 2) \log (k_c + 2) - (G_{cc} + 2) \log (G_{cc} + 2) - k_c \log k_c + G_{cc} \log G_{cc} \right],$$

$$\approx \frac{1}{2E} \left[ 2 \log k_c - 2 \log G_{cc} \right], \quad [12]$$

where the approximation follows by letting $\log (k_c + 2) \approx \log k_c$ and $\log (G_{cc} + 2) \approx \log G_{cc}$. By contrast, adding an edge on the boundary of $c$ (say, connecting a node $i$ outside of $c$ to a node in $c$) yields a contribution to the information rate of

$$\Delta I^{\text{boundary}} = \frac{1}{2E} \left[ (k_i + 1) \log (k_i + 1) + (k_c + 1) \log (k_c + 1) - 2(G_{ic} + 1) \log (G_{ic} + 1) - k_i \log k_i - k_c \log k_c + 2G_{ic} \log G_{ic} \right],$$

$$\approx \frac{1}{2E} \left[ \log k_i + \log k_c - 2 \log G_{ic} \right], \quad [13]$$

where the approximation follows from $\log(k_i + 1) \approx \log k_i$, $\log(k_c + 1) \approx \log k_c$, and $\log(G_{ic} + 1) \approx \log G_{ic}$.

It is clear that $\Delta I^{\text{within}}$ will be less than $\Delta I^{\text{boundary}}$ if $k_i/G_{ic}^2 < k_i/SN$. Given a random selection of nodes to include in both, the fraction $G_{ic}/k_i$ of edges emanating from $c$ that end in $c$ and the fraction $G_{ic}/k_i$ of edges emanating from $i$ that end in $c$ are approximated by the proportional size of $c$, namely, $SN/k_i = S$. Because $G_{ic} \gg k_i$, for a randomly-selected cluster $c$ we have $k_i/G_{ic}^2 \approx 1/SG_{ic} \ll 1/SG_{ic} \approx k_i/G_{ic}^2$. Thus, the information cost of edges within $c$ ($\Delta I^{\text{within}}$) is likely to be lower than the cost of an edge on the boundary of $c$ ($\Delta I^{\text{boundary}}$), leading to the prediction that optimal clusters $c$ will seek to include tightly-knit communities with sparse connectivity to the rest of the network. We confirm this prediction in real networks (Fig. 2D in the main text), and we further demonstrate that modular structure and tight clustering serve to increase network compressibility (Fig. 4A-E in the main text).

Within versus outside edges. Now consider an edge connecting two nodes $i$ and $j$ outside of $c$. Adding such an edge increases the information rate by an amount

$$\Delta I^{\text{outside}} = \frac{1}{2E} \left[ (k_i + 1) \log (k_i + 1) + (k_j + 1) \log (k_j + 1) - k_i \log k_i - k_j \log k_j \right],$$

$$\approx \frac{1}{2E} \left[ \log k_i + \log k_j \right], \quad [14]$$

where the approximation follows from $\log(k_i + 1) \approx \log k_i$ and $\log(k_j + 1) \approx \log k_j$. Comparing Eqs. 12 and 14, we see that $\Delta I^{\text{within}}$ will be less than $\Delta I^{\text{outside}}$ if $k_i^2/G_{ic}^2 \ll k_i/k_j$. Based on the above result that optimal clusters $c$ tend to include tight within-cluster connectivity, we know that the proportion $G_{ic}/k_i$ is greater than the scale $S$ (see Fig. 2D in the main text). Thus, we will have $k_i^2/G_{ic}^2 \ll k_i/k_j$ (and therefore $\Delta I^{\text{within}} \leq \Delta I^{\text{outside}}$) across most scales (specifically, for all scales $S \geq 1/\sqrt{k_i/k_j}$). This result indicates that optimal clusters $c$ will seek to include high-degree nodes and exclude low-degree nodes, which we confirm in Fig. 2E in the main text. Moreover, because compression leverages differences in node degrees, we find that networks with heterogeneous (or heavy-tailed) degrees are highly compressible (Fig. 4F-J in the main text).

3. Rate-distortion curve for k-regular networks

We wish to derive an analytic approximation to the rate-distortion curve $R(S)$ for k-regular networks (Fig. 3B in the main text). For unweighted, undirected networks and a clustering including one large cluster $c$ of size $SN$, the information rate $I(x,y)$ is given in Eq. 11. Let us examine each term individually:

- The large cluster $c$ contains $N - n + 1 = SN$ nodes, and so the number of nodes outside of $c$ is $n - 1 = (1 - S)N$. Since each node in the network has degree $k_c$, we have

$$\sum_{i \in c} k_i \log k_i = (1 - S)Nk_c \log k_c, \quad [15]$$

and

$$k_c \log k_c = SNk_c \log(SNk_c). \quad [16]$$

- Assuming that each edge has a probability $S$ of connecting to nodes in $c$, we can approximate $G_{ic} \approx Sk_i = Sk$ and $G_{cc} \approx Sk_c = S^2Nk_c$. Thus, we can approximate

$$\sum_{i \in c} G_{ic} \log G_{ic} \approx (1 - S)NSk_c \log(Sk), \quad [17]$$

and

$$G_{cc} \log G_{cc} \approx S^2Nk_c \log(S^2Nk_c). \quad [18]$$

Plugging Eqs. 15-18 into Eq. 11, and noting that $2E = kN$, after some algebra we arrive at an analytic approximation to the rate-distortion curve for a k-regular network,

$$R(S) \approx (1 - S)^2 \log k + S(1 - S) \log N - S \log S. \quad [19]$$

We demonstrate that this prediction provides a good approximation to true rate-distortion curves, and becomes accurate in the high-degree limit (Fig. 3C in the main text).
4. Speeding up the clustering algorithm

To compute the rate-distortion curve $\bar{R}(S)$ for a given network, one must identify clusterings that minimize the information rate $\bar{I}(x, y)$ across all scales $S$ (equivalently, for all numbers of clusters $n = N, \ldots, 1$). Here, we employ a greedy clustering algorithm that iteratively combines pairs of clusters so as to minimize the information rate (see Materials and Methods in the main text). Specifically, beginning with $n = N$ clusters (one for each node in the network), we attempt to combine different pairs of clusters and compute the resulting changes to the information rate $\bar{I}(x, y)$. Combining the pair of clusters that yields the largest reduction in information rate, we arrive at a new clustering with $n-1$ clusters. Repeating this process for all numbers of clusters $n = N, \ldots, 1$, we arrive at estimates of the optimal clusterings and information rates across all scales $S$.

A brute force implementation of this algorithm would attempt to combine all $\binom{N}{2} = O(n^2) = O(N^2)$ pairs of clusters at each iteration. For each pair of clusters, one would then compute the new information rate

$$\bar{I}(x, y) = -\sum_c \pi_c \sum_{c'} P_{cc'} \log P_{cc'}, \tag{20}$$

where $\pi_c = \sum_i \pi_i$ is the stationary distribution over clusters, and

$$P_{cc'} = \frac{1}{\pi_c} \sum_{i \in c} \sum_{j \in c'} P_{ij} \tag{21}$$

is the conditional probability of transitioning from cluster $c$ to cluster $c'$. Since Eq. 20 involves summing over all pairs of clusters $c$ and $c'$, computing the information rate requires $O(n^2)$ computations. Finally, the algorithm repeats this process for all numbers of clusters $n = N, \ldots, 1$, requiring a total of $O(N^3)$ computations. This $N^3$ dependence limits a naive implementation of our clustering algorithm to small networks. In what follows, we will show how to reduce this size dependence to $N^2$, significantly improving the efficiency of the algorithm and enabling applications to networks of reasonable size.

**Change in information rate.** As discussed above, a naive implementation of the algorithm would re-compute the information rate (Eq. 20) after attempting to combine each pair of clusters, requiring $O(n^2)$ computations. However, we only require the change in the information rate, a computation that we will see takes $O(n)$ time.

Consider attempting to combine two clusters $\alpha$ and $\beta$. Before combining the clusters, the information rate is given by

$$\bar{I}_{\text{old}} = -\sum_{c \neq \alpha, \beta} \pi_c \sum_{c' \neq \alpha, \beta} P_{cc'} \log P_{cc'}$$

$$- \sum_{c \neq \alpha, \beta} \pi_c (P_{ca} \log P_{ca} + P_{\beta c} \log P_{\beta c})$$

$$- \pi_\alpha \sum_{c \neq \alpha, \beta} P_{ac} \log P_{ac} - \pi_\beta \sum_{c \neq \alpha, \beta} P_{\beta c} \log P_{\beta c}$$

$$- \pi_\alpha (P_{\alpha a} \log P_{\alpha a} + P_{\beta a} \log P_{\beta a})$$

$$- \pi_\beta (P_{\beta a} \log P_{\beta a} + P_{\alpha a} \log P_{\alpha a}). \tag{22}$$

After combining the clusters $\alpha$ and $\beta$, the new information rate is given by

$$\bar{I}_{\text{new}} = -\sum_{c \neq \alpha, \beta} \pi_c \sum_{c' \neq \alpha, \beta} P_{cc'} \log P_{cc'}$$

$$- \sum_{c \neq \alpha, \beta} \pi_c (P_{ca} \log P_{ca} + P_{\beta c} \log P_{\beta c})$$

$$- (\pi_\alpha + \pi_\beta) \sum_{c \neq \alpha, \beta} \pi_\alpha P_{ac} + \pi_\beta P_{\beta c} \log \frac{\pi_\alpha P_{ac} + \pi_\beta P_{\beta c}}{\pi_\alpha + \pi_\beta}$$

$$- (\pi_\alpha + \beta) \frac{P_{\alpha a} + P_{\beta a}}{\pi_\alpha + \pi_\beta} + \pi_\alpha (P_{\beta a} + P_{\alpha a}) \frac{P_{\alpha a} + P_{\beta a}}{\pi_\alpha + \pi_\beta}$$

$$= -\sum_{c \neq \alpha, \beta} \pi_c \sum_{c' \neq \alpha, \beta} P_{cc'} \log P_{cc'}$$

$$- \sum_{c \neq \alpha, \beta} (\pi_\alpha P_{ac} + \pi_\beta P_{\beta c}) \log \frac{\pi_\alpha P_{ac} + \pi_\beta P_{\beta c}}{\pi_\alpha + \pi_\beta}$$

$$- (\pi_\alpha + \beta) \frac{P_{\alpha a} + P_{\beta a}}{\pi_\alpha + \pi_\beta} + \pi_\alpha (P_{\beta a} + P_{\alpha a}) \frac{P_{\alpha a} + P_{\beta a}}{\pi_\alpha + \pi_\beta} \tag{23}$$

Thus, the change in the information after combining clusters $\alpha$ and $\beta$ is given by

$$\Delta \bar{I} = \bar{I}_{\text{new}} - \bar{I}_{\text{old}}$$

$$= -\sum_{c \neq \alpha, \beta} \pi_c (P_{ca} \log P_{ca} + P_{\beta c} \log P_{\beta c})$$

$$- \sum_{c \neq \alpha, \beta} (\pi_\alpha P_{ac} + \pi_\beta P_{\beta c}) \log \frac{\pi_\alpha P_{ac} + \pi_\beta P_{\beta c}}{\pi_\alpha + \pi_\beta}$$

$$- (\pi_\alpha + \beta) \frac{P_{\alpha a} + P_{\beta a}}{\pi_\alpha + \pi_\beta} + \pi_\alpha (P_{\beta a} + P_{\alpha a}) \frac{P_{\alpha a} + P_{\beta a}}{\pi_\alpha + \pi_\beta}$$

$$+ \sum_{c \neq \alpha, \beta} \pi_c (P_{ca} \log P_{ca} + P_{\beta c} \log P_{\beta c})$$

$$+ \pi_\alpha \sum_{c \neq \alpha, \beta} P_{ac} \log P_{ac} + \pi_\beta \sum_{c \neq \alpha, \beta} P_{\beta c} \log P_{\beta c}$$

$$+ \pi_\alpha (P_{\alpha a} \log P_{\alpha a} + P_{\beta a} \log P_{\beta a})$$

$$- \pi_\beta (P_{\beta a} \log P_{\beta a} + P_{\alpha a} \log P_{\alpha a}). \tag{24}$$

Although Eq. 24 appears more complicated than Eq. 20, we remark that it only requires summing over the different clusters $c$ rather than all pairs of clusters $c$ and $c'$. Therefore, by computing the change in information rate rather than the information rate itself, we reduce the number of computations from $O(n^2)$ to $O(n)$.

**Heuristics for cluster selection.** In the naive implementation of the clustering algorithm, one searches through all $\binom{N}{2} = O(n^2)$ pairs of clusters at each iteration to find the pair whose combination yields the largest decrease in the information rate. Here, we instead propose two heuristics for selecting a subset of $m$ pairs of clusters at each iteration, thereby reducing the number of pairs from $O(n^2)$ to $m$. For all results here and in the main text, we consider $m = 100$ pairs of clusters at each iteration of the clustering algorithm.

The first heuristic is motivated by the observation that optimal clusterings tend to include one large cluster with high-degree nodes (Fig. 2E in the main text). In undirected networks, the stationary distribution over clusters is proportional to the cluster degrees, such that $\pi_\alpha = k_\alpha / 2E$, where $k_c = \sum_i k_i$ is the degree of cluster $c$ and $E = \frac{1}{2} \sum_{ij} G_{ij}$ is the number of edges in the network. We
Therefore select the \( m \) pairs of clusters \( c \) and \( c' \) with the largest combined stationary probabilities \( \pi_c + \pi_{c'} \); for undirected networks, this process is equivalent to selecting the pairs of clusters with the largest combined degrees \( k_c + k_{c'} \).

The second heuristic is motivated by the observation that the one large cluster tends to include nodes that are tightly connected to one another (Fig. 2D in the main text). In undirected networks, the joint transition probability from one cluster to another \( c' \) is proportional to the number of edges between the clusters, such that
\[
\pi_c P_{cc'} = G_{cc'}/2E,
\]
where \( G_{cc'} = \sum_{i \in c, j \in c'} G_{ij} \) is the number of edges connecting nodes in \( c \) to nodes in \( c' \). We therefore select the \( m \) pairs of clusters \( c \) and \( c' \) with the largest joint transition probabilities \( \pi_c P_{cc'} + \pi_{c'} P_{c'c} \); for undirected networks, this process is equivalent to selecting the pairs of clusters with the largest number of connecting edges \( G_{cc'} + G_{c'c} \).

To evaluate the performance of these two heuristics, we compare the rate-distortion curves for the real networks in Supporting Table 1 computed using (i) the brute-force approach that attempts to combine all pairs of clusters, (ii) the stationary distribution heuristic, and (iii) the joint transition probability heuristic. Since each algorithm computes an upper bound on the rate-distortion curve \( R(S) \) for a given network, whichever algorithm returns a smaller upper bound will have achieved a more accurate estimate of the true rate-distortion curve \( R(S) \). Consider, for example, the social network of bottlenose dolphins (50). For scales \( S \leq 0.8 \), we find that the stationary distribution heuristic provides a lower (and thus more accurate) upper bound on the rate-distortion curve \( R(S) \) than both the brute-force algorithm and the joint transition probability heuristic (Supporting Fig. 1A). By contrast, for scales \( S \geq 0.8 \), the joint transition probability heuristic provides the most accurate estimate of the rate-distortion curve (Supporting Fig. 1A). Notably, even though both heuristics only consider a limited set of cluster pairs, they both produce rate-distortion estimates that are comparable in accuracy to, if not more accurate than, the brute-force approach that searches through all pairs at each iteration.

To compare the accuracy of the different algorithms (that is, to determine which algorithm provides a lower upper bound \( \bar{R}(S) \) across all scales) we can compare their compressibility estimates,
\[
\bar{C} = H(x) - \frac{1}{N} \sum_S \bar{R}(S). \tag{25}
\]
If one algorithm produces a lower (that is, more accurate) rate-distortion estimate \( \bar{R}(S) \) on average across all scales \( S \), then one will arrive at a larger lower bound on the compressibility \( \bar{C} \). In Supporting Fig. 1B, we see that the compressibility computed using the stationary distribution heuristic \( \bar{C}_{\text{stationary}} \) is almost always as large as (if not larger than) that computed using all pairs of clusters \( \bar{C}_{\text{all}} \) for the real networks in Supporting Table 1. Similarly, in Supporting Fig. 1C we see the joint transition probability heuristic provides compressibility estimates \( \bar{C}_{\text{transition}} \) that are nearly as accurate as \( \bar{C}_{\text{all}} \) across most of the real networks. In fact, the brute-force compressibility estimates \( \bar{C}_{\text{all}} \) are significantly larger (that is, more accurate) than both of the heuristic estimates \( \bar{C}_{\text{stationary}} \) and \( \bar{C}_{\text{transition}} \) for only 10 of the 69 real networks.

Together, these results demonstrate that the two cluster-selection heuristics provide rate-distortion estimates that are comparable, if not more accurate, than the brute-force clustering algorithm. Moreover, these heuristics reduce the search for an optimal pair of clusters at each iteration from \( O(n^2) = O(N^2) \) pairs to \( m \) pairs. In combination with the speed-up in Sec. 4, this reduces the total runtime of the clustering algorithm from \( O(N^3) \) to \( O(mN^2) \), thereby allowing applications to networks of reasonable size.

5. Directed networks
In the main text, in order to develop analytic predictions for the structure of optimal clusterings and the impact of network structure on compressibility, we focused on the special case of undirected networks. Specifically, we predicted (and confirmed) that optimal compressions would tend to maximize the number of edges within the one large cluster and minimize the number of edges on the boundary and outside of the cluster (Fig. 2D–E in the main text). In turn, these findings led to the predictions that network compressibility (Eq. 25) should increase with both modular structure and heavily-tailed degrees, which we confirmed in undirected real and model networks (Fig. 4 in the main text). Here we demonstrate that these central results from the main text are not limited to undirected networks, but in fact generalize to directed networks.

Structure of optimal compressions. Our clustering algorithm described in the main text is general, applying to any weighted, directed network. We can therefore use the algorithm to compute optimal compressions and rate-distortion curves \( R(S) \) for the directed real networks listed in Supporting Table 1. We remark that among the 69 real networks studied in the main text, 38 have directed versions (see Supporting Table 1), which we analyze here.

As was the case for undirected networks (Fig. 2B in the main text), we find that optimal compressions in directed networks tend to form one large cluster with the maximum possible size \( N-n+1 = SN \), where \( S = 1 - \frac{n-1}{N} \) is the scale of description, and \( n \) = 1 minimal
Supporting Fig. 2. Structure of optimal clusterings in directed networks. (A) Size of the largest cluster in a compression, normalized by the size of the network $N$, as a function of the scale $S$ for the directed real networks in Supporting Table 1. The median over real networks (solid line) matches the largest possible normalized cluster size, $(N−n)/(N−S)$, (B) Illustration of edges within the one large cluster (blue), outside the cluster (red), crossing the boundary out of the cluster (purple), and crossing the boundary into the cluster (green). (C) Fraction of the $k_{\text{out}}$ edges emanating from the large cluster that either connect to nodes outside the cluster $1 − G_{cc}/k_{\text{out}}^\text{med}$ (blue) or remain within the cluster $G_{cc}/k_{\text{out}}^\text{med}$ (blue) as a function of the scale $S$. (D) Fraction of the $k_{\text{in}}^\text{out}$ edges incident to the large cluster that emanate either from nodes outside the cluster $1 − G_{cc}/k_{\text{in}}^\text{med}$ (green) or within the cluster $G_{cc}/k_{\text{in}}^\text{med}$ (blue) as a function of the scale $S$. (E-F) Average out-degree (E) and in-degree (F) of nodes inside (blue) and outside (red) the large cluster, normalized by the average degree of the network, as a function of the scale $S$. In panels (C-F), solid lines and shaded regions represent averages and one-standard-deviation error bars, respectively, over the directed real networks in Supporting Table 1, and dashed lines correspond to clusters with nodes selected at random.

clusters containing one node each (Supporting Fig. 2A). To analyze the structure of the large cluster, in the main text we divided the edges in an undirected network into three categories (Fig. 2C in the main text): those within the cluster, those outside of the cluster, and those on the cluster boundary. For directed networks, we can further divide boundary edges into two categories—those connecting from inside to outside the cluster and those connecting from outside to inside the cluster—thereby resulting in a total of four edge types (Supporting Fig. 2B).

For undirected networks, in the main text we demonstrated that the one large cluster had tight within-cluster connectivity (maximizing the number of edges inside the cluster) and sparse connectivity to the rest of the network (minimizing the number of boundary edges; Fig. 2D in the main text). Indeed, this result generalizes to directed networks, with the one large cluster favoring within-cluster edges over both outgoing (Supporting Fig. 2C) and incoming (Supporting Fig. 2D) boundary edges. In the main text, we also demonstrated that the one large cluster sought to include high-degree nodes and exclude low-degree nodes in undirected networks (Fig. 2E in the main text). Here, we confirm that this result also applies to directed networks, with the one large cluster containing nodes with both larger out-degrees (Supporting Fig. 2E) and in-degrees (Supporting Fig. 2F) than the rest of the network. Together, these results establish that our predictions about the structure of optimal compressions (see Sec. 2) generalize to directed networks.

Impact of network structure on compressibility. We are now prepared to study the compressibility of directed networks. In the main text, we demonstrated that the compressibility of undirected networks increases with the logarithm of the average degree (Fig. 3E in the main text). Moreover, based on the structure of optimal clusterings (Fig. 2 in the main text), we hypothesized (and confirmed) that the compressibility of undirected networks increases with both transitivity and degree heterogeneity (Fig. 4 in the main text). Here, we demonstrate that each of these results about the impact of network structure on compressibility generalizes to directed networks.

We begin by analyzing the dependence of network compressibility on average degree. For a directed network with adjacency matrix $G$, where $G_{ij} = 1$ if there is a directed edge from node $i$ to node $j$, the out-(in-)degree of node $i$ is given by $k_{i\text{out}} = \sum_j G_{ij}$ ($k_{i\text{in}} = \sum_j G_{ji}$). Despite each node possibly having different out- and in-degrees, the average out- and in-degrees in a network are equal, since

$$\langle k_{\text{out}} \rangle = 1/N \sum_i k_{i\text{out}} = 1/N \sum_i G_{ij} = 1/N \sum_j k_{j\text{in}} = \langle k_{\text{in}} \rangle .$$

Therefore, even for directed networks, we can simply discuss the average degree. In Supporting Fig. 3A, we demonstrate that the compressibility $C$ (Eq. 25) of real directed networks grows logarithmically with the average degree, confirming that our prediction for undirected networks (Fig. 3D-E in the main text) extends to directed networks.

We now consider the impact of transitivity on the compressibility of directed networks. As in the main text, we quantify transitivity using the average clustering coefficient. For undirected networks, the clustering coefficient of a node $i$ is the number of edges connecting the neighbors of $i$ divided by the $\binom{\ell}{2} = k_i(k_i-1)/2$ possible connections, and the average clustering coefficient is given by averaging
over the nodes in a network. Although the clustering coefficient was originally defined for undirected networks, this definition has since been extended to directed networks.\(^{(51)}\) In a directed network, the clustering coefficient of a node \(i\) is the number of connected pairs among its \(k^\text{out}_i + k^\text{in}_i\) out- and in-neighbors divided by the \(\binom{k^\text{out}_i + k^\text{in}_i}{2}\) possible connections. Just as we found for undirected networks (Fig. 4A–E in the main text), the compressibility of directed networks is significantly correlated with the average clustering coefficient (Supporting Fig. 3B), thereby indicating that transitivity and the presence of tightly-knit communities serve to make networks more compressible.

Finally, we study the impact of heterogeneous (or heavy-tailed) degrees on the compressibility of directed networks. For directed networks, there are two definitions of degree heterogeneity: that of out-degrees \(\langle k^\text{out}_i - k^\text{in}_j \rangle / k^\text{out}_i\) and that of in-degrees \(\langle k^\text{in}_i - k^\text{in}_j \rangle / k^\text{in}_i\). We find that the compressibility of directed networks is significantly correlated with both the out-degree (Supporting Fig. 3C) and in-degree (Supporting Fig. 3D) heterogeneities. Thus, even in directed networks, heavy-tailed degree distributions with well-connected hubs serve to increase network compressibility. Together, these results demonstrate that the central conclusions from the main text generalize to directed networks; namely, that strong transitivity and degree heterogeneity – the two defining features of hierarchical organization\(^{(24)}\) – increase the compressibility of complex networks.

### 6. Robustness to clustering heuristics

In the main text, we computed rate-distortion curves \(\hat{R}(S)\) using our clustering algorithm with the pair-selection heuristics described in Sec. 4. We found that network compressibility increased with average degree (Fig. 3D–E in the main text), transitivity (quantified by average clustering coefficient; Fig. 4D–E in the main text), and degree heterogeneity (Fig. 4F–J in the main text). Here, we confirm that these results are not simply due to our choices of clustering heuristics. To do so, we recompute the compressibilities \(C\) of the networks in Supporting Table 1 using the brute-force implementation of our clustering algorithm that searches through all pairs of clusters at each iteration (see Sec. 4). We remark that, in order to employ the brute-force implementation, for each network of size \(N > 10^3\) we analyze 50 randomly-sampled subnetworks of 100 nodes each. Using the brute-force algorithm, we confirm that the compressibility increases with average degree (Supporting Fig. 4A), transitivity (Supporting Fig. 4B), and degree heterogeneity (Supporting Fig. 4C). These observations demonstrate that the central results from the main text are robust to our choice of pair-selection heuristic in the clustering algorithm.
7. Network datasets and processing

The real-world networks analyzed in the main text are listed and briefly described in Supporting Table 1. The web, citation, animal, semantic, social, protein, flight, and road networks are gathered from online network repositories. The language and music networks were generated by the authors previously (see below). The brain networks are generated from structural and functional data gathered and analyzed previously (see below).

For the language networks, we developed code to (i) remove punctuation and white space, (ii) filter words by their part of speech, and (iii) record the transitions between the filtered words. Here we focus on networks of transitions between nouns, noting that the same methods can be used to record transitions between other parts of speech. The raw text was gathered from Project Gutenberg (gutenberg.org/wiki/Main_Page).

For the music networks, we read in audio files in MIDI format using the midisynth function in MATLAB (R2018a). For each song, we split the notes by their channel, which represents the different instruments. For each channel, we created a network of note transitions. We then create a transition network representing the entire song by aggregating the transitions between notes across the different channels. The MIDI files were gathered from mididownload.com and from kunstderfuge.com. Our code and data are available upon request from the corresponding author.

For the brain networks, we study the structural and functional connectivity of five randomly-selected subjects from the Human Connectome Project. For all subjects, the brain is divided into 100 predefined cortical regions. The structural connectivity networks reflect physical white-matter tracts between brain regions measured using diffusion tensor imaging (DTI). We threshold the structural networks to only include connections between regions that are stronger than the mean. The functional connectivity networks reflect Pearson correlations between regional brain activity. Specifically, the activity is defined by blood-oxygen-level-dependent (BOLD) functional magnetic resonance imaging (fMRI) signals. To ensure that the functional networks have approximately the same edge density as the structural networks, we threshold the functional networks to only include connections stronger than the mean plus one standard deviation.

1. Arun E Sizemore, Elisabeth A Kanuza, Chad Giust, and Danielle S Bassett. Knowledge gaps in the early growth of semantic feature networks. Nat. Hum. Behav., 2(9):682, 2018.
2. Alexei Vázquez, Romualdo Pastor-Satorras, and Alessandro Vespignani. Large-scale topological and dynamical properties of the Internet. Phys. Rev. E, 65(6):066130, 2002.
3. Xiao Fan Liu, K Tse Chi, and Michael Small. Complex network structure of musical compositions: Algorithmic generation of appealing music. Physica A, 389(1):126–132, 2010.
4. Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. Proc. Natl. Acad. Sci., 99(12):7821–7826, 2002.
5. Eleanor R Brush, David C Krakauer, and Jessica C Flack. Conflicts of interest improve collective computation of adaptive social structures. Sci. Adv., 4(1):e1600331, 2018.
6. Virpi Kalakoski and Pertti Saariluoma. Taxi drivers’ exceptional memory of street names. Mem. Cogn., 28(4):634–638, 2001.
7. Christopher W Lynn and Danielle S Bassett. How humans learn and represent networks. In press, Proc. Natl. Acad. Sci. (arxiv.org/abs/1909.07186), 2020.
8. Anne-Claude Gavin, Patrick Aloy, Paola Grandi, Roland Krause, Markus Boesche, Martina Marzioch, Christina Rau, Lars Juh Jensen, Sonja Bastuck, Birgit Dömpelfeld, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature, 440(7084):631–636, 2006.
9. Christopher W Lynn and Danielle S Bassett. The physics of brain network structure, function, and control. Nat. Rev. Phys., 1, 2019.
10. Petra E Verstes, Timothy Rittman, Kirstie J Whitaker, Rafaela Romero-Garcia, Frantisek Vála, Manfred G Kitzbichler, Konrad Wagstyl, Peter Fonagy, Raymond J Dolan, Peter B Jones, et al. Gene transcription profiles associated with inter-modal hubs and connection distance in human functional magnetic resonance imaging networks. Philos. Trans. R. Soc. B, 371(1705):20150362, 2016.
11. Kirstie J Whitaker, Petra E Verstes, Rafaela Romero-Garcia, Frantisek Vála, Michael Moutoussis, Gitte Prabhu, Nikolai Weiskopf, Martina C Callaghan, Konrad Wagstyl, Timothy Rittman, et al. Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. Proc. Natl. Acad. Sci., 113(32):9125–9130, 2016.
12. Claude E. Shannon. A mathematical theory of communication. Bell Syst. Tech. J., 27(3):379–423, 1948.
13. Thomas M Cover and Joy A Thomas. Elements of information theory. John Wiley & Sons, 2012.
14. Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci., 105(4):1118–1123, 2008.
15. Martin Rosvall and Carl T Bergstrom. An information-theoretic framework for resolving community structure in complex networks. Proc. Natl. Acad. Sci., 104(18):7327–7331, 2007.
16. Noam Slonim, Gurinder Singh Atwal, Galper Trakol, and William Bialek. Information-based clustering. Proc. Natl. Acad. Sci., 102(51):18297–18302, 2005.
17. Jure Leskovec, Kevin J Lang, Anita B Dasgupta, and Michael W Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. Internet Math., 6(1):29–123, 2009.
18. Wayne W Zachary. An information flow model for conflict and fission in small groups. J. Anthropol. Res., 33(3):452–473, 1977.
19. Evan Archer, Il Memming Park, and Jonathan W Pillow. Recent advances in graph partitioning. In Algorithm Engineering, pages 117–154. Springer, 2016.
20. Mark EJ Newman and Gasim Reiner. Estimating the number of communities in a network. Phys. Rev. Lett., 117(7):078301, 2016.
21. Austin H Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. Science, 355(6325):163–166, 2016.
22. Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. Science, 286(5439):509–512, 1999.
23. Ernesto Ravazz and Albert-László Barabási. Hierarchical organization in complex networks. Phys. Rev. E, 67(2):026112, 2003.
24. Marco Tomassini and Leslie Luthis. Empirical analysis of the evolution of a scientific collaboration network. Physica A, 385(2):750–764, 2007.
25. Ernesto Ravazz and. Detecting hierarchical modularity in biological networks. In Computational Systems Biology, pages 145–160. Springer, 2009.
26. Moneerad M Farhood, David J Heeger, Gary Marcus, Uri Hassan, and Yuia Lerner. The neural processing of hierarchical structure in music and speech at different timescales. Front.
| Type       | Name                                      | N     | E     | Description                                               | Reference |
|------------|-------------------------------------------|-------|-------|-----------------------------------------------------------|-----------|
| **Language** |                                           |       |       |                                                           |           |
|            | Shakespeare: Combined works*              | 11,234| 97,892| Noun transitions.                                        | (54)      |
|            | Homer: Iliad*                             | 3,556 | 23,608| Noun transitions.                                        | (55)      |
|            | Plato: Republic*                          | 2,271 | 9,796 | Noun transitions.                                        | (56)      |
|            | Jane Austen: Pride and Prejudice*         | 1,994 | 12,120| Noun transitions.                                        | (57)      |
|            | William Blake: Songs of Innocence...*     | 370   | 781   | Noun transitions.                                        | (58)      |
|            | Miguel de Cervantes: Don Quijote*         | 6,090 | 43,682| Noun transitions.                                        | (59)      |
|            | Walt Whitman: Leaves of Grass*            | 4,791 | 16,526| Noun transitions.                                        | (60)      |
| **Music**  |                                           |       |       |                                                           |           |
|            | Michael Jackson: Thriller*                | 67    | 446   | Note transitions.                                        | (61)      |
|            | Beatles: Hard Day's Night*                | 41    | 212   | Note transitions.                                        | (62)      |
|            | Queen: Bohemian Rhapsody*                 | 71    | 961   | Note transitions.                                        | (63)      |
|            | Toto: Africa*                             | 39    | 163   | Note transitions.                                        | (64)      |
|            | Mozart: Sonata No 11*                    | 55    | 354   | Note transitions.                                        | (65)      |
|            | Beethoven: Sonata No 23*                  | 69    | 900   | Note transitions.                                        | (66)      |
|            | Chopin: Nocturne Op 9-2*                  | 59    | 303   | Note transitions.                                        | (67)      |
|            | Bach: Clavier Fugue 13*                   | 40    | 143   | Note transitions.                                        | (68)      |
|            | Brahms: Ballade Op 10-1*                  | 69    | 670   | Note transitions.                                        | (69)      |
| **Web**    |                                           |       |       |                                                           |           |
|            | Google internal*                          | 12,354| 142,296| Hyperlinks between internal Google cites.                  | (70, 71) |
|            | Education                                 | 2,622 | 6,065 | Hyperlinks between education webpages.                    | (72, 73) |
|            | EPA                                       | 2,232 | 6,876 | Hyperlinks between pages linking to www.epa.gov.          | (73, 74) |
|            | Indochina                                 | 9,638 | 45,886| Hyperlinks between pages in Indochina.                    | (73, 75) |
|            | 2004 Election blogs*                      | 793   | 13,484| Hyperlinks between blogs on US politics.                   | (73, 76) |
|            | Spam                                      | 3,796 | 36,404| Hyperlinks between spam pages.                            | (73, 77) |
|            | WebBase                                    | 6,843 | 16,374| Hyperlinks gathered by web crawler.                       | (73, 75) |
| **Citation**|                                           |       |       |                                                           |           |
|            | arXiv Hep-Ph*                             | 12,711| 139,500| Citations in Hep-Ph section of the arXiv.                  | (71, 78) |
|            | arXiv Hep-Th*                             | 7,464 | 115,932| Citations in Hep-Th section of the arXiv.                  | (71, 78) |
|            | arXiv*                                    | 3,991 | 16,621| Citations between scientific papers.                      | (71, 79) |
|            | DBLP*                                     | 240   | 858   | Citations between scientific papers.                      | (71, 80) |
| **Animal** |                                           |       |       |                                                           |           |
|            | Dolphins                                  | 62    | 159   | Social relationships.                                    | (50, 71) |
|            | Little Rock Lake*                         | 183   | 2,434 | Food web of animal consumption.                          | (71, 81) |
|            | Macaques                                 | 62    | 325   | Dominance relationships.                                 | (71, 82) |
|            | Sheep                                     | 24    | 91    | Dominance relationships.                                 | (71, 83) |
|            | Wetlands*                                 | 128   | 2,075 | Food web of animal consumption.                          | (71, 84) |
| **Semantic**|                                           |       |       |                                                           |           |
|            | Game of Thrones                          | 796   | 2,823 | Character co-occurrences.                                | (71)      |
|            | Algebra                                   | 278   | 3,553 | Concept co-occurrences.                                  | (86)      |
|            | Bible                                     | 1,707 | 9,059 | Pronoun co-occurrences.                                  | (71)      |
|            | Les Miserables                            | 77    | 254   | Character co-occurrences.                                | (71)      |
|            | Edinburgh Thesaurus*                      | 7,754 | 226,518| Word similarities in human experiments.                   | (87, 88) |
|            | Roget Thesaurus*                          | 904   | 3,447 | Linked semantic categories.                               | (88, 89) |
|            | Glossary terms                            | 60    | 114   | Words used in definitions of other words.                  | (88)      |
|            | FOLDOC*                                   | 13,274| 90,736| Same as above (computing terms).                          | (88, 90) |
|            | ODLIS*                                    | 1,802 | 12,378| Same as above (information science terms).                | (88, 91) |
| **Brain**  |                                           |       |       |                                                           |           |
|            | Structural connectivity 1*                | 100   | 806   | Structural connections between brain regions.             | (52)      |
|            | Structural connectivity 2*                | 100   | 858   | Structural connections between brain regions.             | (52)      |
|            | Structural connectivity 3*                | 100   | 865   | Structural connections between brain regions.             | (52)      |
|            | Structural connectivity 4*                | 100   | 916   | Structural connections between brain regions.             | (52)      |
|            | Structural connectivity 5*                | 100   | 906   | Structural connections between brain regions.             | (52)      |
|            | Functional connectivity 1                 | 100   | 811   | Functional correlations between brain regions.            | (52)      |
|            | Functional connectivity 2                 | 100   | 800   | Functional correlations between brain regions.            | (52)      |
|            | Functional connectivity 3                 | 100   | 830   | Functional correlations between brain regions.            | (52)      |
|            | Functional connectivity 4                 | 100   | 854   | Functional correlations between brain regions.            | (52)      |
|            | Functional connectivity 5                 | 100   | 799   | Functional correlations between brain regions.            | (52)      |
| **Social** |                                           |       |       |                                                           |           |
|            | Zachary's karate club                     | 34    | 78    | Interactions between karate club members.                 | (18, 71) |
|            | Facebook                                  | 13,130| 75,562| Subset of the Facebook network.                           | (71, 92) |
|            | arXiv Astr-Ph                            | 17,903| 196,972| Coauthorships in Astr-Ph section of arXiv.                | (71, 78) |
|            | arXiv Hep-Th                             | 22,721| 2,672,975| Coauthorships in Hep-Th section of arXiv.                | (71, 78) |
|            | Adolescent health*                        | 2,155 | 8,970 | Friendships between students.                             | (71, 93) |
|            | Highschool*                               | 67    | 267   | Friendships between highschool students.                  | (71, 94) |
|            | Jazz                                      | 198   | 2,742 | Collaborations between jazz musicians.                    | (71, 95) |
| **Protein**|                                           |       |       |                                                           |           |
|            | C. elegans                               | 453   | 2,025 | Interactions between metabolites.                         | (71, 96) |
|            | Human (Figeys)                            | 2,239 | 6,452 | Protein interactions.                                    | (71, 97) |
|            | Human (Stelzl)                            | 1,706 | 6,207 | Protein interactions.                                    | (71, 98) |
|            | Yeast                                     | 1,870 | 2,277 | Protein interactions.                                    | (71, 99) |
| **Flight** |                                           |       |       |                                                           |           |
|            | US Flights*                               | 1,574 | 28,236| Flights between US airports.                              | (71)      |
|            | OpenFlights*                              | 2,939 | 30,501| Flights between world cities.                             | (71, 100) |
| Type | Name | N | E | Description | Reference |
|------|------|---|---|-------------|-----------|
| Road | EuroRoad | 1,174 | 1,417 | Network of roads between European cities. | (71, 101) |
|      | Chicago* | 12,982 | 39,018 | Network of roads in Chicago. | (71, 102) |
|      | New York City | 264,346 | 730,100 | Network of roads in New York. | (71) |
|      | Bay Area | 321,270 | 794,830 | Network of roads in San Francisco area. | (71) |

Supporting Table 1. Real networks analyzed in the main text. For each network we list its type; name and whether it has a directed version (denoted by *); number of nodes N; number of edges E; brief description; and reference.
