Supporting Information

Quantifying the compressibility of complex networks

Christopher W. Lynn$^{1,2}$ and Danielle S. Bassett$^{3,4,5,6,7,8,*}$

$^{1}$Initiative for the Theoretical Sciences, Graduate Center, City University of New York, New York, NY 10016, USA
$^{2}$Joseph Henry Laboratories of Physics, Princeton University, Princeton, NJ 08544, USA
$^{3}$Department of Bioengineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA
$^{4}$Department of Physics & Astronomy, College of Arts & Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA
$^{5}$Department of Electrical & Systems Engineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA
$^{6}$Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
$^{7}$Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
$^{8}$Santa Fe Institute, Santa Fe, NM 87501, USA
## Contents

1. Introduction .................................................. 3
2. Information content of different edges .................. 3
   2.1 Within versus boundary edges .......................... 4
   2.2 Within versus outside edges ......................... 5
3. Rate-distortion curve for $k$-regular networks ......... 5
4. Speeding up the clustering algorithm .................... 6
   4.1 Change in information rate ............................ 7
   4.2 Heuristics for cluster selection ..................... 8
5. Errors in information bounds .............................. 11
6. Compressing model networks ............................... 13
7. Directed networks ........................................... 16
   7.1 Structure of optimal compressions .................. 16
   7.2 Impact of network structure on compressibility ..... 18
8. Robustness to clustering heuristics ..................... 20
9. Network datasets and processing .......................... 21
References .................................................... 24
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 tractable upper bound on the rate-distortion curve $\bar{R}(S)$ used throughout the main text provides a reasonable approximation to the true rate-distortion curve. In Sec. 6, we study the structure of optimal compressions for the model networks analyzed in the main text. In Sec. 7, we demonstrate that the central results from the main text generalize to directed networks. In Sec. 8, 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. 9, 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 $\bar{I}(x, y)$, 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 $\bar{I}(x, y)$. 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

$$\bar{I}(x, y) = \frac{1}{2E} \left[ \sum_{i \notin c} k_i \log k_i + k_c \log k_c - 2 \sum_{i \notin c} G_{ic} \log G_{ic} - G_{cc} \log G_{cc} \right], \quad (S1)$$
where the sums run over all nodes $i$ not in $c$, $E = \frac{1}{2} \sum_{ij} G_{ij}$ is the number of edges in the network, $k_i = \sum_j G_{ij}$ 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 $G_{cc} = \sum_{ij \in c} G_{ij}$ is the number of edges connecting nodes within $c$ (see Materials and Methods in the main text).

2.1 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

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

$$\approx \frac{1}{2E} (2 \log k_c - 2 \log G_{cc}) ,$$

(S2)

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 \bar{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]$$

$$\approx \frac{1}{2E} (\log k_i + \log k_c - 2 \log G_{ic}) ,$$

(S3)

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 $\Delta \bar{I}_{\text{within}}$ will be less than $\Delta \bar{I}_{\text{boundary}}$ if $k_c / G_{cc}^2 \leq k_i / G_{ic}^2$. Given a random selection of nodes to include in $c$, both the fraction $G_{cc} / k_c$ 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, $SN/N = S$. Because $G_{cc} \gg G_{ic}$, for a randomly-selected cluster $c$ we have $k_c / G_{cc}^2 \approx 1 / SG_{cc} \ll 1 / SG_{ic} \approx k_i / G_{ic}^2$. Thus, the information cost of edges within $c$ ($\Delta \bar{I}_{\text{within}}$) is likely to be lower than the cost of an edge on the boundary of $c$ ($\Delta \bar{I}_{\text{boundary}}$), 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).

2.2 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

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

$$\approx \frac{1}{2E} (\log k_i + \log k_j),$$

(S4)

where the approximation follows from $\log(k_i + 1) \approx \log k_i$ and $\log(k_j + 1) \approx \log k_j$. Comparing Eqs. S2 and S4, we see that $\Delta \bar{I}_{\text{within}}$ will be less than $\Delta \bar{I}_{\text{outside}}$ if $k_c^2 / G_{cc}^2 \leq k_i k_j$. Based on the above result that optimal clusters $c$ tend to include tight within-cluster connectivity, we know that the proportion $G_{cc} / k_c$ is greater than the scale $S$ (see Fig. 2D in the main text). Thus, we will have $k_c^2 / G_{cc}^2 \leq k_i k_j$ (and therefore $\Delta \bar{I}_{\text{within}} \leq \Delta \bar{I}_{\text{outside}}$) 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 $\bar{R}(S)$ 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 $\bar{I}(x, y)$ is given in Eq. S1. 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 $k$, we have

$$\sum_{i \in c} k_i \log k_i = (1 - S)Nk \log k,$$

(S5)
and
\[ k_c \log k_c = SNk \log(SNk). \] (S6)

- Assuming that each edge has a probability \( S \) of connecting to nodes in \( c \), we can approximate
\[ G_{ic} \approx Sk_i = Sk \text{ and } G_{cc} \approx Sk_c = S^2Nk. \] Thus, we can approximate
\[ \sum_{i \in c} G_{ic} \log G_{ic} \approx (1 - S)NSk \log(Sk), \] (S7)
and
\[ G_{cc} \log G_{cc} \approx S^2Nk \log(S^2Nk). \] (S8)

Plugging Eqs. S5-S8 into Eq. S1, and noting that \( 2E = kN \), after some algebra we arrive at an analytic approximation to the rate-distortion curve for a \( k \)-regular network,
\[ \bar{R}(S) \approx (1 - S)^2 \log k + S(1 - S) \log N - S \log S. \] (S9)

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

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

(S10)

where \( \pi_c = \sum_{i \in c} \pi_i \) is the stationary distribution over clusters, and

\[ P_{cc'} = \frac{1}{\pi_c} \sum_{i \in c} \sum_{j \in c'} P_{ij} \]

(S11)

is the conditional probability of transitioning from cluster \( c \) to cluster \( c' \). Since Eq. S10 involves summing over all pairs of clusters \( c \) and \( c' \), computing the information rate requires \( O(n^2) = O(N^2) \) computations. Finally, the algorithm repeats this process for all numbers of clusters \( n = N, \ldots, 1 \), requiring a total of \( O(N^5) \) computations. This \( N^5 \) dependence limits a naïve 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.

4.1 Change in information rate

As discussed above, a naïve implementation of the algorithm would re-compute the information rate (Eq. S10) 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

\[ \bar{I}^{\text{old}} = - \sum_{c \neq \alpha, \beta} \pi_c \sum_{c' \neq \alpha, \beta} P_{cc'} \log P_{cc'} - \sum_{c \neq \alpha, \beta} \pi_c (P_{c\alpha} \log P_{c\alpha} + P_{c\beta} \log P_{c\beta}) \]

\[ - \pi_\alpha \sum_{c \neq \alpha, \beta} P_{\alpha c} \log P_{\alpha c} - \pi_\beta \sum_{c \neq \alpha, \beta} P_{\beta c} \log P_{\beta c} \]

\[ - \pi_\alpha (P_{\alpha\alpha} \log P_{\alpha\alpha} + P_{\alpha\beta} \log P_{\alpha\beta}) - \pi_\beta (P_{\beta\beta} \log P_{\beta\beta} + P_{\beta\alpha} \log P_{\beta\alpha}). \]

(S12)
After combining the clusters $\alpha$ and $\beta$, the new information rate is given by

\[
\bar{I}_{\text{new}} = - \sum_{c \neq \alpha, \beta} \pi_c \sum_{c' \neq \alpha, \beta} P_{cc'} \log P_{cc'} - \sum_{c \neq \alpha, \beta} \pi_c (P_{\alpha\alpha} + P_{\beta\beta}) \log (P_{\alpha\alpha} + P_{\beta\beta}) - (\pi_\alpha + \pi_\beta) \sum_{c \neq \alpha, \beta} \frac{\pi_\alpha P_{\alpha\alpha} + \pi_\beta P_{\beta\beta}}{\pi_\alpha + \pi_\beta} \log \frac{\pi_\alpha P_{\alpha\alpha} + \pi_\beta P_{\beta\beta}}{\pi_\alpha + \pi_\beta} \\
- (\pi_\alpha + \pi_\beta) \pi_\alpha (P_{\alpha\alpha} + P_{\beta\beta}) + \pi_\beta (P_{\beta\beta} + \beta_{3\beta}) \log \frac{\pi_\alpha (P_{\alpha\alpha} + P_{\beta\beta}) + \pi_\beta (P_{\beta\beta} + \beta_{3\beta})}{\pi_\alpha + \pi_\beta}
\]

(S13)

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

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

(S14)

\[
= - \sum_{c \neq \alpha, \beta} \pi_c (P_{\alpha\alpha} + P_{\beta\beta}) \log (P_{\alpha\alpha} + P_{\beta\beta}) - \sum_{c \neq \alpha, \beta} (\pi_\alpha P_{\alpha\alpha} + \pi_\beta P_{\beta\beta}) \log \frac{\pi_\alpha P_{\alpha\alpha} + \pi_\beta P_{\beta\beta}}{\pi_\alpha + \pi_\beta} \\
- (\pi_\alpha (P_{\alpha\alpha} + P_{\beta\beta}) + \pi_\beta (P_{\beta\beta} + \beta_{3\beta}) \log \frac{\pi_\alpha (P_{\alpha\alpha} + P_{\beta\beta}) + \pi_\beta (P_{\beta\beta} + \beta_{3\beta})}{\pi_\alpha + \pi_\beta} \\
+ \sum_{c \neq \alpha, \beta} \pi_c (P_{\alpha\alpha} \log P_{\alpha\alpha} + P_{\beta\beta} \log P_{\beta\beta}) + \pi_\alpha \sum_{c \neq \alpha, \beta} P_{\alpha\alpha} \log P_{\alpha\alpha} + \pi_\beta \sum_{c \neq \alpha, \beta} P_{\beta\beta} \log P_{\beta\beta} \\
+ \pi_\alpha (P_{\alpha\alpha} \log P_{\alpha\alpha} + P_{\beta\beta} \log P_{\beta\beta}) - \pi_\beta (P_{\beta\beta} \log P_{\beta\beta} + \beta_{3\beta} \log P_{\beta\beta})
\]

Although Eq. S14 appears more complicated than Eq. S10, 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 $O(n)$.

### 4.2 Heuristics for cluster selection

In the naïve 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 $\pi_c = k_c / 2E$, where $k_c = \sum_{i \in c} k_i$ 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 $c$ 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 $\bar{R}(S)$ 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. For scales $S \lesssim 0.8$, we find that the stationary distribution heuristic provides a lower (and thus more accurate) upper bound on the rate-distortion curve $\bar{R}(S)$ than both the brute-force algorithm and the joint transition probability heuristic (Supporting Fig. 1A). By contrast, for scales $S \gtrsim 0.8$, the joint transition probability heuristic provides the
Supporting Fig. 1 | Performance of clustering heuristics. (A) Rate-distortion curves for a network of interactions between bottlenose dolphins¹ computed using the clustering algorithm with the brute-force strategy that tests all pairs of clusters at each iteration (black), the top stationary distribution heuristic (blue), and the top joint transition probability heuristic (red). (B) The difference between compressibility computed using the stationary distribution heuristic $C^{\text{stationary}}$ and that computed using all pairs of clusters $C^{\text{all}}$ as a function of the compressibility $C^{\text{all}}$ for the real networks in Supporting Table 1. (C) Difference between compressibility computed using the joint transition probability heuristic $C^{\text{transition}}$ and that computed using all pairs of clusters $C^{\text{all}}$ as a function of the compressibility $C^{\text{all}}$. In panels (B) and (C), for networks of size $N > 100$, data points and error bars represent averages and standard deviations, respectively, over 50 subnetworks of 100 nodes each sampled using random walks beginning at random seed nodes (see Materials and Methods in the main text).

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,

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

(S15)

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 $\overline{C}$. In Supporting Fig. 1B, we see that the compressibility computed using the stationary distribution heuristic $\overline{C}_{\text{stationary}}$ is almost always as large as (if not larger than) that computed using all pairs of clusters $\overline{C}_{\text{all}}$ for the real networks in Supporting Table 1. Similarly, in Supporting Fig. 1C we see the joint transition probability heuristic provides compressibility estimates $\overline{C}_{\text{transition}}$ that are nearly as accurate as $\overline{C}_{\text{all}}$ across most of the real networks. In fact, the brute-force compressibility estimates $\overline{C}_{\text{all}}$ are significantly larger (that is, more accurate) than both of the heuristic estimates $\overline{C}_{\text{stationary}}$ and $\overline{C}_{\text{transition}}$ 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. 4.1, this reduces the total run-time of the clustering algorithm from $O(N^5)$ to $O(mN^2)$, thereby allowing applications to networks of reasonable size.

5 Errors in information bounds

In the main text, rather than computing the information rate $I(x, y)$ of a given clustering directly, we instead consider a tractable upper bound $\overline{I}(x, y)$ (see Materials and Methods in the main text). For a given network, we minimize $\overline{I}(x, y)$ across all scales to arrive at an upper bound on the rate-distortion curve $\overline{R}(S)$, which we then use to compute a lower bound $\overline{C}$ on the compressibility (Eq. S15). The upper bound $\overline{R}(S)$ on the rate-distortion curve and the corresponding lower bound $\overline{C}$ on the compressibility together form the basis of our main results (Figs. 3 and 4 in the main text). Therefore, it is important to verify that these bounds provide reasonable approximations to the true rate-distortion curves $R(S)$ and compressibilities $C$ of the networks analyzed in the main text.

To investigate the accuracy of the upper bound on the rate-distortion curve $\overline{R}(S)$, we consider a tractable lower bound $\overline{R}(S)$ (see Materials and Methods in the main text). Importantly, the true rate-distortion curve $R(S)$ lies between the upper and lower bounds, such that $\overline{R}(S) \geq R(S) \geq C$.\overline
Supporting Fig. 2 | Errors in rate-distortion curves and compressibilities. (A) Upper bounds $\bar{R}(S)$ (blue) and lower bounds $\underline{R}(S)$ (red) on the rate-distortion curves of the real networks in Supporting Table 1, normalized by the entropy $H(x)$ of each network. Lines represent medians over the real networks, and shaded regions reflect interquartile ranges. (B) Difference $\bar{R}(S) - \underline{R}(S)$ between the upper and lower bounds on the rate-distortion curves, normalized by the entropy $H(x)$. The black line and shaded region represent the median and interquartile range over the real networks in Supporting Table 1, respectively, and the green dashed line indicates Zachary’s karate club (see Fig. 2A in the main text). (C) Box plots (over the real networks in Supporting Table 1) of the lower bound on the compressibility $\bar{C}$ (computed using the upper bound on the rate-distortion curve $\bar{R}(S)$; blue), the upper bound on the compressibility $\bar{C}$ (computed using $\underline{R}(S)$; red), and the difference $\bar{C} - \underline{C}$ (grey), all normalized by the entropy $H(x)$. Whiskers have a maximum length of 1.5 times the interquartile range. Outliers are indicated by circles, either black (for road networks) or grey (for protein networks). Green triangles illustrate compressibility values for Zachary’s karate club.

Thus, the error in the upper bound ($\bar{R}(S) - R(S)$) is no larger than the difference between the upper and lower bounds ($\bar{R}(S) - \underline{R}(S)$). Put simply, if the difference $\bar{R}(S) - \underline{R}(S)$ between the bounds is small, then we know that the error $\bar{R}(S) - R(S)$ is even smaller.

For all networks, one can show that the upper and lower bounds are equal (and therefore exact) at both the minimum scale $S = 1/N$ and the maximum scale $S = 1$ (see Materials and Methods in the main text). Moreover, in the main text, we demonstrated for Zachary’s karate club that the upper bound $\bar{R}(S)$ remains close to the lower bound $\underline{R}(S)$ across all intermediate scales (Fig. 2A in the main text). In order to compare the rate-distortion bounds across different networks,
here we normalize $\bar{R}(S)$ and $B(S)$ by the entropy $H(x)$ of each network. Notably, we find that a tight correspondence between the upper and lower bounds is not unique to Zachary’s karate club, but instead is a general feature of the real networks in Supporting Table 1 (Supporting Fig. 2A).

Indeed, the maximum difference $\bar{R}(S) - B(S)$ between the upper and lower bounds tends to only reach about 10% of the entropy of a network (Supporting Fig. 2B). For comparison, the difference between bounds is actually larger in Zachary’s karate club (reaching 15% of the entropy) than the typical network in Supporting Table 1 (Supporting Fig. 2B). These results establish that the upper bound $\bar{R}(S)$ provides a good approximation to the true rate-distortion curve $R(S)$.

We now investigate the accuracy of the lower bound $C$ on the compressibility (Eq. S15). To do so, we consider the upper bound $\bar{C}$ on the compressibility induced by the lower bound $\bar{R}(S)$ on the rate-distortion curve. We find that the lower bound $C$ remains close to the upper bound $\bar{C}$ across almost all networks in Supporting Table 1 (Supporting Fig. 2C). For comparison, Zachary’s karate club exhibits a difference $\bar{C} - C$ that is typical for the networks in Supporting Table 1 (Supporting Fig. 2C, right). Interestingly, among the five outliers with an abnormally large difference $\bar{C} - C$, four are the road networks in Supporting Table 1 (Supporting Fig. 2C, right). As discussed in the main text, the abnormal compression properties of road networks likely reflects the unique physical constraints on their structure. Together, the results of this section verify that the upper bound $\bar{R}(S)$ on the rate-distortion curve and the corresponding lower bound $C$ on the compressibility provide reasonable estimates of the true information properties of the real networks analyzed in the main text.

6 Compressing model networks

In our investigations of network compressibility, we analyzed a number of model networks, including Erdős-Rényi, $k$-regular (Fig. 3B in the main text), stochastic block (Fig. 4A in the main text), and scale-free (Fig. 4F in the main text) networks. Here, we study the structure of optimal compressions in these model networks, as well as networks with hierarchical structure, wherein small tightly-connected groups of nodes connect to form larger, but looser groups.\(^3\)

To recall, based on the information content of different edges (see Sec. 2), we predicted 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. In the main text, we confirmed these predictions for the real networks in Supporting Table 1 (see Fig. 2 in the main text). In Supporting Fig. 3, we verify that these results extend to all of the model networks listed above.

First, across all model networks considered, we find that the edges emanating from the one large cluster tend to connect to nodes within the cluster (Supporting Fig. 3A) and avoid crossing over to nodes outside of the cluster (Supporting Fig. 3B). Interestingly, we observe stark differences in the properties of optimal compressions between different network models. For example, although Erdős-Rényi and $k$-regular networks do not contain large-scale structure, optimal compressions are still able to identify groups of nodes that are slightly more strongly connected to each other than to the rest of the network. Meanwhile, optimal compressions in stochastic block, scale-free, and hierarchical networks identify groups of nodes with much stronger within-group connectivity. In fact, for stochastic block networks, it is clear that optimal compressions identify the built-in modules, with the one large cluster iteratively enveloping each module one by one as the scale increases (Supporting Fig. 3A-B).

Second, for Erdős-Rényi, scale-free, and hierarchical networks, we confirm that optimal compressions select groups of nodes with higher degrees than average (Supporting Fig. 3C) while omitting low-degree nodes (Supporting Fig. 3D). As expected, this effect is much stronger in scale-free and hierarchical networks (which exhibit large heterogeneities in node degree) than in Erdős-Rényi networks (which do not contain large-scale structure). We remark that in $k$-regular networks, there are no differences in node degrees by definition. Similarly, for stochastic block networks, we find a negligible difference between the degrees of nodes inside versus outside the large cluster. This final result once again suggests that optimal compressions in stochastic block networks focus on their modular, rather than heterogeneous, structure. Together, the results of this section demonstrate that the structure of optimal compressions observed in real networks (Fig. 2 in the main text) extends to a range of model networks. Moreover, for hierarchical networks, we es-
Supporting Fig. 3 | Structure of optimal clusterings in model networks. (A-B) Fraction of the $k_c$ edges emanating from the large cluster that connect to nodes either within the cluster $G_{cc}/k_c$ (A) or outside the cluster $1 - G_{cc}/k_c$ (B) as a function of the scale $S$. (C-D) Average degrees of nodes inside (C) and outside (D) the large cluster, normalized by the average degree of the network, as a function of the scale $S$. In panels (C-D), data are not displayed for $k$-regular networks because the normalized degree of all nodes is one by definition. Across all panels, the data for Erdős-Rényi (black), $k$-regular (blue), stochastic block (10 modules; green), and scale-free (red) networks represent averages over 50 randomly-generated networks, each with $N = 10^3$ nodes and average degree $\langle k \rangle = 100$. The hierarchical data (magenta) are computed using a Ravasz-Barabási network with four recursive levels ($N = 625$, $\langle k \rangle = 6.32$). The fraction $f = 0.87$ of within-module edges in the stochastic block networks and the exponent $\gamma = 2.16$ for the scale-free networks are chosen to match the hierarchical network.
establish that optimal compressions cluster together groups of nodes that are more tightly connected than scale-free networks (Supporting Fig. 3A), yet higher in degree than stochastic block networks (Supporting Fig. 3C), thereby capitalizing on both the modular and heterogeneous properties of hierarchical organization.

7 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. Here we demonstrate that the central results from the main text generalize to directed networks.

7.1 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 \( \bar{R}(S) \) 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 clusters containing one node each (Supporting Fig. 4A). 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. 4B).

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
Supporting Fig. 4 | 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, $(N - n + 1)/N = S$. (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 $k_c^{\text{out}}$ edges emanating from the large cluster that either connect to nodes outside the cluster $1 - G_{cc}/k_c^{\text{out}}$ (purple) or remain within the cluster $G_{cc}/k_c^{\text{out}}$ (blue) as a function of the scale $S$. (D) Fraction of the $k_c^{\text{in}}$ edges incident to the large cluster that emanate either from nodes outside the cluster $1 - G_{cc}/k_c^{\text{in}}$ (green) or within the cluster $G_{cc}/k_c^{\text{in}}$ (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.
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. 4C) and incoming (Supporting Fig. 4D) 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. 4E) and in-degrees (Supporting Fig. 4F) 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.

7.2 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 \( G_{ij} \), 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 \( k^\text{out}_i = \sum_j G_{ij} \) (\( k^\text{in}_i = \sum_j G_{ji} \)). Despite each node possibly having different out- and in-degrees, the average out- and in-degrees in a network are equal, since

\[
\langle k^\text{out} \rangle = \frac{1}{N} \sum_i k^\text{out}_i = \frac{1}{N} \sum_{ij} G_{ij} = \frac{1}{N} \sum_j k^\text{in}_j = \langle k^\text{in} \rangle. \tag{S16}
\]

Therefore, even for directed networks, we can simply discuss the average degree. In Supporting Fig. 5A, we demonstrate that the compressibility \( \bar{C} \) (Eq. S15) 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.
Supporting Fig. 5 | Compressibility of directed networks. (A) Compressibility $\bar{C}$ (Eq. S15) versus average degree for the directed real networks in Supporting Table 1. We note that average degree is plotted on a log scale. The dashed line indicates a logarithmic fit. (B) Compressibility $\bar{C}$ versus transitivity (quantified by the average clustering coefficient) for directed real networks (Supporting Table 1) with a linear best fit (dashed line). (C-D) Compressibility $\bar{C}$ versus out-degree heterogeneity (C) and in-degree heterogeneity (D) for directed real networks (Supporting Table 1) with linear best fits (dashed lines). In all panels, for networks of size $N > 10^3$, the data points and error bars represent means and standard deviations over 50 randomly-sampled subnetworks of $10^3$ nodes each (see Materials and Methods in the main text).
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 \( \binom{k_i}{2} = k_i(k_i - 1)/2 \) 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. In a directed network, the clustering coefficient of a node \( i \) is the number of connected pairs among its \( k_i^\text{out} + k_i^\text{in} \) out- and in-neighbors divided by the \( \binom{k_i^\text{out} + k_i^\text{in}}{2} \) 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. 5B), 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 \( \langle k_i^\text{out} - k_j^\text{out} \rangle / \langle k_i^\text{out} \rangle \) and that of in-degrees \( \langle k_i^\text{in} - k_j^\text{in} \rangle / \langle k_i^\text{in} \rangle \). We find that the compressibility of directed networks is significantly correlated with both the out-degree (Supporting Fig. 5C) and in-degree (Supporting Fig. 5D) 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 – increase the compressibility of complex networks.

### 8 Robustness to clustering heuristics

In the main text, we computed rate-distortion curves \( \bar{R}(S) \) using our clustering algorithm with the pair-selection heuristics described in Sec. 4.2. 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. 4I-J in the main text). Here, we confirm that these results are not simply due to our choices of clustering heuristics. To do
Supporting Fig. 6 | Compressibility without clustering heuristics. Compressibility \( C \) versus average degree (\( A \)), transitivity (\( B \)), and degree heterogeneity (\( C \)) for the real networks in Supporting Table 1. Compressibility is computed using the brute-force clustering algorithm that tests all pairs of clusters at each iteration rather than a reduced number of pairs selected through heuristics (see Sec. 4.2). In all panels, for networks of size \( N > 100 \) data points and error bars represent means and standard deviations over 50 randomly-sampled subnetworks of 100 nodes each (see Materials and Methods in the main text).

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. 4.2). We remark that, in order to employ the brute-force implementation, for each network of size \( N > 100 \) 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. 6A), transitivity (Supporting Fig. 6B), and degree heterogeneity (Supporting Fig. 6C). These observations demonstrate that the central results from the main text are robust to our choice of pair-selection heuristic in the clustering algorithm.

9 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).\textsuperscript{6,7}

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.\textsuperscript{5} 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 (\url{gutenberg.org/wiki/Main_Page}).

For the music networks, we read in audio files in MIDI format using the \texttt{readmidi} 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 \url{midiworld.com} and from \url{kunstderfuge.com}. Our code and data are openly available at \url{github.com/ChrisWLynn/Network_compressibility}.

For the brain networks, we study the structural and functional connectivity of five randomly-selected subjects from the Human Connectome Project.\textsuperscript{6} For all subjects, the brain is divided into 100 predefined cortical regions.\textsuperscript{57} 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 imagining (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.
| Type       | Name                                      | $N$  | $E$     | Description                          | Reference |
|------------|-------------------------------------------|------|---------|--------------------------------------|-----------|
| Language   | Shakespeare: Combined works*              | 11,234 | 97,892 | Noun transitions.                    | 8         |
|            | Homer: Iliad*                              | 3,556 | 23,608  | Noun transitions.                    | 9         |
|            | Plato: Republica*                          | 2,271 | 9,796   | Noun transitions.                    | 10        |
|            | Jane Austen: Pride and Prejudice*         | 1,994 | 12,120  | Noun transitions.                    | 11        |
|            | William Blake: Songs of Innocence...*     | 370   | 781     | Noun transitions.                    | 12        |
|            | Miguel de Cervantes: Don Quijote*         | 6,090 | 43,682  | Noun transitions.                    | 13        |
|            | Walt Whitman: Leaves of Grass*            | 4,791 | 16,526  | Noun transitions.                    | 14        |
| Music      | Michael Jackson: Thriller*                | 67    | 446     | Note transitions.                    | 15        |
|            | Beatles: Hard Day’s Night*                | 41    | 212     | Note transitions.                    | 16        |
|            | Queen: Bohemian Rhapsody*                 | 71    | 961     | Note transitions.                    | 17        |
|            | Toto: Africa*                              | 39    | 163     | Note transitions.                    | 18        |
|            | Mozart: Sonata No 11*                     | 55    | 354     | Note transitions.                    | 19        |
|            | Beethoven: Sonata No 23*                  | 69    | 900     | Note transitions.                    | 20        |
|            | Chopin: Nocturne Op 2-2*                  | 59    | 303     | Note transitions.                    | 21        |
|            | Bach: Clavier Fugue 13*                   | 40    | 143     | Note transitions.                    | 22        |
|            | Brahms: Ballade Op 10-1*                  | 69    | 670     | Note transitions.                    | 23        |
| Web        | Google internal*                          | 12,354| 142,296 | Hyperlinks between internal Google cites. | 24,25     |
|            | Education                                 | 2,622 | 6,065   | Hyperlinks between education webpages. | 26,27     |
|            | EPA                                       | 2,232 | 6,876   | Hyperlinks between pages linking to www.epa.gov. | 27,28     |
|            | Indochina                                 | 9,638 | 45,886  | Hyperlinks between pages in Indochina. | 27,29     |
|            | 2004 Election blogs*                      | 793   | 13,484  | Hyperlinks between blogs on US politics. | 28,30     |
|            | Spam                                      | 3,796 | 36,404  | Hyperlinks between spam pages.       | 27,31     |
|            | WebBase                                   | 6,843 | 16,374  | Hyperlinks gathered by web crawler.  | 27,29     |
| Citation   | arXiv Hep-Ph*                             | 12,711| 139,500 | Citations in Hep-Ph section of the arXiv. | 25,32     |
|            | arXiv Hep-Th*                             | 7,464 | 115,932 | Citations in Hep-Th section of the arXiv. | 25,32     |
|            | DBLP*                                     | 3,991 | 16,621  | Citations between scientific papers. | 25,33     |
|            | ODLIS*                                    | 240   | 858     | Citations between scientific papers. | 25,34     |
| Animal     | Dolphins                                  | 62    | 159     | Social relationships.                | 1.25      |
|            | Little Rock Lake                          | 183   | 2,434   | Food web of animal consumption.      | 25,35     |
|            | Macaques                                  | 62    | 325     | Dominance relationships.              | 25,36     |
|            | Sheep                                     | 24    | 91      | Dominance relationships.              | 25,37     |
|            | Wetlands*                                 | 128   | 2,075   | Food web of animal consumption.      | 25,38     |
|            | Zebras                                    | 23    | 105     | Social interactions.                 | 25,39     |
| Semantic   | Game of Thrones                           | 796   | 2,823   | Character co-occurrences.            | 25         |
|            | Algebra                                   | 278   | 3,553   | Concept co-occurrences.              | 40         |
|            | Bible                                     | 1,707 | 9,059   | Pronoun co-occurrences.              | 25         |
|            | Les Miserables                            | 77    | 254     | Character co-occurrences.            | 25         |
|            | Edinburgh Thesaurus*                      | 7,754 | 226,518 | Word similarities in human experiments. | 41,42     |
|            | Rogel Thesaurus*                          | 904   | 3,447   | Linked semantic categories.          | 42,43     |
|            | Glossary terms                            | 60    | 114     | Words used in definitions of other words. | 42         |
|            | FOLDOC*                                   | 13,274| 90,736  | Same as above (computing terms).     | 42,44     |
|            | ODLIS*                                    | 1,802 | 12,378  | Same as above (information science terms). | 42,45     |
| Brain      | Structural connectivity 1*                | 100   | 806     | Structural connections between brain regions. | 6         |
|            | Structural connectivity 2*                | 100   | 858     | Structural connections between brain regions. | 6         |
|            | Structural connectivity 3*                | 100   | 865     | Structural connections between brain regions. | 6         |
|            | Structural connectivity 4*                | 100   | 916     | Structural connections between brain regions. | 6         |
|            | Structural connectivity 5*                | 100   | 906     | Structural connections between brain regions. | 6         |
|            | Functional connectivity 1                 | 100   | 811     | Functional correlations between brain regions. | 6         |
|            | Functional connectivity 2                 | 100   | 800     | Functional correlations between brain regions. | 6         |
|            | Functional connectivity 3                 | 100   | 830     | Functional correlations between brain regions. | 6         |
|            | Functional connectivity 4                 | 100   | 854     | Functional correlations between brain regions. | 6         |
|            | Functional connectivity 5                 | 100   | 799     | Functional correlations between brain regions. | 6         |
| Social     | Zachary’s karate club                     | 34    | 78      | Interactions between karate club members. | 2.25       |
|            | Facebook                                  | 13,130| 75,562  | Subset of the Facebook network.      | 25,46      |
|            | arXiv Astr-Ph                             | 17,903| 196,972 | Coauthorships in Astr-Ph section of arXiv. | 25,32      |
|            | Adolescent health*                        | 2,155 | 8,970   | Friendships between students.         | 25,47      |
|            | Highschool*                               | 67    | 267     | Friendships between highschool students. | 25,48      |
|            | Jazz                                      | 198   | 2,742   | Collaborations between jazz musicians. | 25,49      |
| Protein    | C. elegans                                | 453   | 2,025   | Interactions between metabolites.     | 25,50      |
|            | Human (Figeys)                            | 2,239 | 6,452   | Protein interactions.                 | 25,51      |
|            | Human (Stelzl)                            | 1,706 | 6,207   | Protein interactions.                 | 25,52      |
|            | Yeast                                     | 1,870 | 2,277   | Protein interactions.                 | 25,53      |
| Flight     | US Flights*                               | 1,574 | 28,236  | Flights between US airports.          | 25         |
|            | OpenFlights*                              | 2,939 | 30,501  | Flights between world cities.         | 25,54      |
| Road       | EuroRoad*                                 | 1,174 | 1,417   | Network of roads between European cities. | 25,55      |
|            | Chicago*                                  | 12,982| 39,018  | Network of roads in Chicago.          | 25         |
|            | New York City                            | 264,346 | 730,100 | Network of roads in New York.         | 25         |
|            | Bay Area                                  | 321,270 | 794,830 | Network of roads in San Francisco area. | 25         |

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.
References

1. Lusseau, D. et al. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behav. Ecol. Sociobiol.* **54**, 396–405 (2003).

2. Zachary, W. W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **33**, 452–473 (1977).

3. Ravasz, E. & Barabási, A.-L. Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 026112 (2003).

4. Fagiolo, G. Clustering in complex directed networks. *Phys. Rev. E* **76**, 026107 (2007).

5. Lynn, C. W., Papadopoulos, L., Kahn, A. E. & Bassett, D. S. Human information processing in complex networks. *Nat. Phys.* 1–9 (2020).

6. Van Essen, D. C. et al. The WU-Minn Human Connectome Project: An overview. *Neuroimage* **80**, 62–79 (2013).

7. Bertolero, M. A. & Bassett, D. S. Deep neural networks carve the brain at its joints. *arXiv preprint arXiv:2002.08891* (2020).

8. Shakespeare, W. *The complete works of William Shakespeare* (Wordsworth Editions, 2007).

9. Pope, A., Buckley, T. A. et al. *The Iliad of Homer* (WW Gibbings, 1891).

10. Jowett, B. *The republic of Plato* (Clarendon press, 1888).

11. Austen, J. *Pride and prejudice* (Broadview Press, 2001).

12. Blake, W. *Songs of Innocence and of Experience*, vol. 5 (Princeton University Press, 1998).

13. Cervantes, M. *Don Quixote* (LBA, 2018).

14. Whitman, W. *Leaves of grass* (Oregan Publishing, 2017).

15. Jackson, M. Thriller (1984).
16. The Beatles. A hard day’s night (1964).
17. Queen. Bohemian rhapsody (1975).
18. Toto. Africa (1982).
19. Mozart, W. A. Piano sonata no. 11 (1784).
20. Beethoven, L. v. Piano sonata no. 23 (1807).
21. Chopin, F. Nocturnes, op. 9, no. 2 (1832).
22. Bach, J. S. The well-tempered clavier, book I, no. 13 (1722).
23. Brahms, J. Ballades, op. 10, no. 1 (1854).
24. Palla, G., Farkas, I. J., Pollner, P., Derényi, I. & Vicsek, T. Directed network modules. New J. Phys. 9, 186 (2007).
25. Kunegis, J. Konect: The Koblenz network collection. In Proceedings of the 22nd International Conference on World Wide Web, 1343–1350 (ACM, 2013).
26. Gleich, D., Zhukov, L. & Berkhin, P. Fast parallel pagerank: A linear system approach. In Yahoo! Research Technical Report YRL-2004-038, vol. 13, 22 (2004).
27. Rossi, R. A. & Ahmed, N. K. The network data repository with interactive graph analytics and visualization. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (2015).
28. De Nooy, W., Mrvar, A. & Batagelj, V. Exploratory social network analysis with Pajek, vol. 27 (Cambridge University Press, 2011).
29. Boldi, P., Codenotti, B., Santini, M. & Vigna, S. UbiCrawler: A scalable fully distributed web crawler. Software Pract. Exper. 34, 711–726 (2004).
30. Adamic, L. A. & Glance, N. The political blogosphere and the 2004 US election: Divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, 36–43 (ACM, 2005).

31. Castillo, C., Chellapilla, K. & Denoyer, L. Web spam challenge 2008. In *Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web* (2008).

32. Leskovec, J., Kleinberg, J. & Faloutsos, C. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data* **1**, 2 (2007).

33. Šubelj, L. & Bajec, M. Model of complex networks based on citation dynamics. In *Proceedings of the 22nd international conference on World Wide Web*, 527–530 (ACM, 2013).

34. Ley, M. The dblp computer science bibliography: Evolution, research issues, perspectives. In *International symposium on string processing and information retrieval*, 1–10 (Springer, 2002).

35. Martinez, N. D. Artifacts or attributes? Effects of resolution on the little rock lake food web. *Ecol. Monogr.* **61**, 367–392 (1991).

36. Takahata, Y. Diachronic changes in the dominance relations of adult female Japanese monkeys of the Arashiyama B group. *The monkeys of Arashiyama* 123–139 (1991).

37. Hass, C. C. Social status in female bighorn sheep (*Ovis canadensis*): Expression, development and reproductive correlates. *J. Zool.* **225**, 509–523 (1991).

38. Ulanowicz, R. E. & DeAngelis, D. L. Network analysis of trophic dynamics in south Florida ecosystems. *US Geological Survey Program on the South Florida Ecosystem* 114 (1999).

39. Sundaresan, S. R., Fischhoff, I. R., Dushoff, J. & Rubenstein, D. I. Network metrics reveal differences in social organization between two fission-fusion species, Grevy’s zebra and onager. *Oecologia* **151**, 140–149 (2007).

40. Christianson, N. H., Sizemore Blevins, A. & Bassett, D. S. Architecture and evolution of semantic networks in mathematics texts. *Proc. R. Soc. Lond.* **476**, 20190741 (2020).
41. Kiss, G. R., Armstrong, C., Milroy, R. & Piper, J. An associative thesaurus of English and its computer analysis. In *The computer and literary studies*, 153–165 (Edinburgh University Press, 1973).

42. Batagelj, V. & Mrvar, A. Pajek datasets. <vlado.fmf.uni-lj.si/pub/networks/data/> (2006).

43. Roget, P. M. *Roget’s Thesaurus of English Words and Phrases* (TY Crowell Company, 1911).

44. Howe, D. The free on-line dictionary of computing. <foldoc.org> (1993).

45. Reitz, J. M. *Online dictionary for library and information science* (Libraries Unlimited, 2010).

46. Viswanath, B., Mislove, A., Cha, M. & Gummadi, K. P. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online social networks*, 37–42 (ACM, 2009).

47. Moody, J. Peer influence groups: Identifying dense clusters in large networks. *Soc. Netw. 23*, 261–283 (2001).

48. Coleman, J. S. *et al.* *Introduction to mathematical sociology.* (London Free Press Glencoe, 1964).

49. Gleiser, P. M. & Danon, L. Community structure in jazz. *Adv. Complex Syst. 6*, 565–573 (2003).

50. Duch, J. & Arenas, A. Community detection in complex networks using extremal optimization. *Phys. Rev. E 72*, 027104 (2005).

51. Ewing, R. M. *et al.* Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol. Syst. Biol. 3*, 89 (2007).

52. Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell 122*, 957–968 (2005).

53. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature 411*, 41–42 (2001).

27
54. Opsahl, T., Agneessens, F. & Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Netw.* **32**, 245–251 (2010).

55. Šubelj, L. & Bajec, M. Robust network community detection using balanced propagation. *Eur. Phys. J. B* **81**, 353–362 (2011).

56. Eash, R., Chon, K., Lee, Y. & Boyce, D. Equilibrium traffic assignment on an aggregated highway network for sketch planning. *Transp. Res.* **13**, 243–257 (1979).

57. Thomas Yeo, B. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).