communications physics

Article

https://doi.org/10.1038/s42005-024-01830-3

Network mutual information measures for graph similarity

Helcio Felippe[1], Federico Battiston[1] & Alec Kirkley[2,3,4]

A wide range of tasks in network analysis, such as clustering network populations or identifying anomalies in temporal graph streams, require a measure of the similarity between two graphs. To provide a meaningful data summary for downstream scientific analyses, the graph similarity measures used for these tasks must be principled, interpretable, and capable of distinguishing meaningful overlapping network structure from statistical noise at different scales of interest. Here we derive a family of graph mutual information measures that satisfy these criteria and are constructed using only fundamental information theoretic principles. Our measures capture the information shared among networks according to different encodings of their structural information, with our mesoscale mutual information measure allowing for network comparison under any specified network coarse-graining. We test our measures in a range of applications on real and synthetic network data, finding that they effectively highlight intuitive aspects of network similarity across scales in a variety of systems.

Network similarity and distance measures are widely applied across science and engineering disciplines for understanding the shared structure among multiple graphs[1–6]. Common graph-level analysis tasks such as network population clustering[7], network regression[8], and network classification[9] require as input a network similarity measure and are highly sensitive to this choice, leading to the construction of a vast number of graph similarity measures to accommodate different application needs[10–12]. In the exploratory data analysis setting, the goal of performing graph similarity or distance calculations is often to identify some meaningful summary of a set of graphs or network layers—for example, clusters of graphs or a subset of anomalous graphs[13,14]. In this case, it is essential that the graph similarity or distance measure used is principled, interpretable, and capable of distinguishing meaningful shared structure among the input graphs from statistical noise. Existing measures based on collections of network summary statistics or feature embeddings[15,16] are often hard to interpret, and there is no clear criterion for which features to include in the similarity/distance calculation. Methods based on graph spectra[17,18] have a clear connection to community structure and random walk dynamics on graphs but are also challenging to interpret given the localization tradeoff between graph space and frequency space[19]. In addition, many of these methods that do not rely on external labels do not explicitly aim to capture statistically meaningful structural distinctions between graphs, meaning that they can reflect a high amount of similarity between two graphs that is purely due to constraints imposed by global properties such as graph density or degree distributions. In this context, information theory can provide a principled framework for deciding the extent to which two networks are similar by allowing us to compute mutual information between two graphs that quantifies the amount of information they share under a particular encoding scheme. Mutual information measures resulting from a minimum description length (MDL) approach have been used to compare partitions of objects in a diverse range of applications[20–24], becoming the standard measure for comparing partitions of networks in the community detection setting[25].

An often overlooked aspect of graph similarity measures is what scale is highlighted in the network comparison calculation. Most existing graph similarity measures can be broadly categorized based on the structural level —micro, meso, or macro—at which they highlight network similarity[16]. For unsupervised tasks involving graph similarity measures, any of the three network scales may be of interest. For example, when comparing snapshots of a longitudinal social network to examine its temporal evolution, different analyses may require a network similarity measure that focuses on microscale structural overlap among individuals' ego networks, a measure that captures the evolving mesoscale community structure of the network, or a measure that tracks macroscale summary statistics of the network such as its total density. Existing graph similarity measures have largely been constructed to highlight a particular scale of interest—for example, individual edge overlap at the microscale is the focus of traditional graph edit distances[26], and global path structure in the network is highlighted by betweenness or closeness centrality-based network similarity measures[15]. Spectral distance measures can simultaneously highlight different scales within the network due to the presence of high and low-frequency

[1]Department of Network and Data Science, Central European University, Vienna, Austria. [2]Institute of Data Science, University of Hong Kong, Hong Kong SAR, China. [3]Department of Urban Planning and Design, University of Hong Kong, Hong Kong SAR, China. [4]Urban Systems Institute, University of Hong Kong, Hong Kong SAR, China. e-mail: battistonf@ceu.edu; alec.w.kirkley@gmail.com

eigenmodes[17,18], but it is unclear exactly to what extent each of the three scales contributes to the computed similarity. There is currently a lack of graph similarity measures that have a clear dependence on the scale of interest, which is critical for interpretability and flexibility in applications.

In ref. 27, an information-theoretic measure for comparing networks is presented that is based on identifying shared substructures (e.g., stars and cliques) that minimize the description length of a pair of networks. This approach is quite elegant and flexible but is limited in practice to subgraph similarities and can become computationally burdensome with the inclusion of larger structural vocabularies. It also does not allow for the comparison of graphs at the meso- or macroscale while ignoring edge-level details, as is possible with the mesoscale mutual information measure presented in this paper. A pairwise graph mutual information measure is proposed in ref. 28, but this measure requires a continuous embedding of the input graphs which may produce distortions of its topological structure. It also does not consider the problem of adjusting the scale at which the graph structure is compared. In ref. 29, an MDL approach for clustering populations of node-aligned networks is presented, which is equivalent to maximum a posteriori Bayesian inference with the model[30], up to a choice of prior. This approach considers encoding sample graphs based on their cluster's representative "mode" network. The encoding for a cluster of networks resembles the encoding used here for the standard (non-degree-corrected) conditional entropy between graphs, but it does not accommodate the comparison of networks in a symmetric pairwise manner, and also does not address the issue of scale. In general, principled approaches to measure graph similarity in the node-aligned setting that we consider here have primarily focused on analyzing entire populations of graphs[31,32], leaving a relative scarcity of methods for principled pairwise network comparison.

In this paper, we construct a family of mutual information measures for computing graph similarity at different scales (see Fig. 1). We do this by considering multiple encodings of node-aligned graphs that exploit different aspects of shared network structure—edges, node neighborhoods, and mixing patterns among arbitrary partitions of nodes in a network—and using these encodings to quantify the amount of shared information between two input graphs. Our measures are principled, interpretable, fast to compute, and naturally highlight the significant shared structure between two graphs beyond the overlap expected due to global structural constraints such as edge density and degree distributions. To validate our methods, we apply them in a range of experiments, finding that they intuitively capture structural perturbations in synthetic networks and identify meaningful shared structures across layers in multilayer networks.

Methods

To measure the similarity between two graphs G_1 and G_2, we can appeal to information theory and compute the amount of information shared between G_1 and G_2, also known as their mutual information[33]. Due to nice properties such as non-negativity, boundedness, and symmetry, mutual information measures and their normalized variants have been widely used across unsupervised learning tasks within and outside of network science[22,23,34].

A key component in the development of a mutual information measure is the specification of an *encoding scheme*, which specifies how one will represent data configurations in a codebook that is used to communicate the data to a receiver. Encoding schemes that focus on different structural properties in a graph will naturally tend to result in different estimates of the information shared between the graphs, thus different values of the mutual information. Here we describe three encoding schemes that aim to highlight shared structure among graphs at different scales, and discuss how to use them to form three mutual information measures that capture different aspects of similarity among networks.

Network mutual information

Let G_1 and G_2 be graphs on the same set of N-labeled nodes. We will focus our attention on the case where these graphs are undirected with no self- or multi-edges, but discuss later on how our measures can be straightforwardly extended when we relax these assumptions. For convenience, we will represent G_1 and G_2 with sets of undirected, unweighted edges of sizes E_1 and E_2 (the number of unique edges in G_1 and G_2), respectively. In principle, our measure can handle any two graphs G_1 and G_2 that have the same number of nodes N—regardless of whether these node labels are aligned. However, in the case of unaligned node labels one must perform graph alignment prior to using the algorithm in order to obtain meaningful results, and this is a highly nontrivial task that will have a substantial impact on the

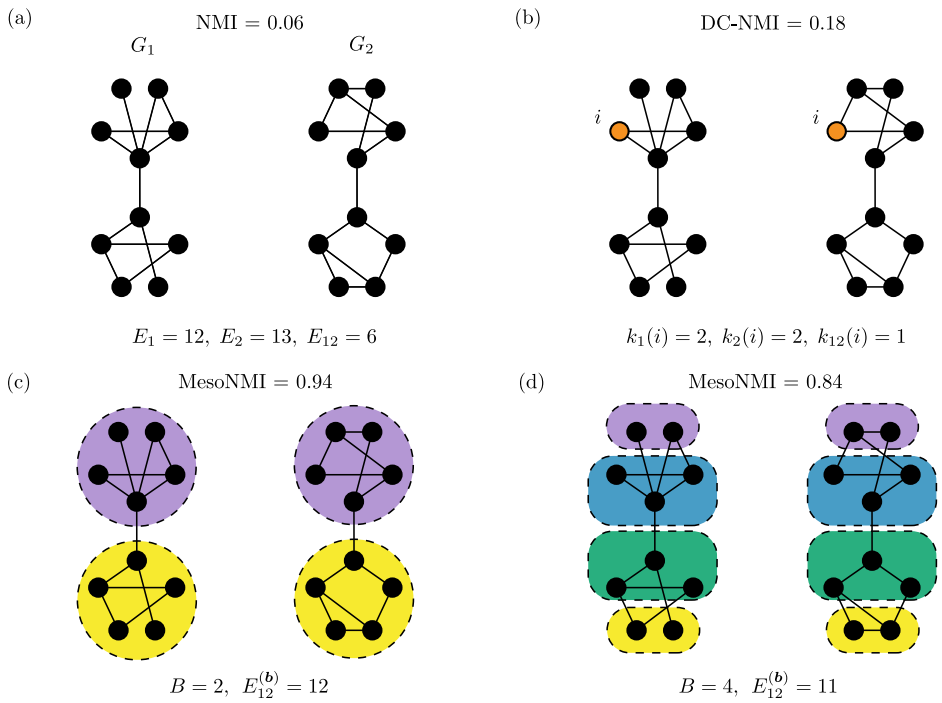

Fig. 1 | Family of proposed network mutual information measures for graph similarity. **a** Standard normalized mutual information (NMI, Eq. (10)) between networks G_1 and G_2, with shared node labels indicated by node positions. Due to little overlap in the edge positions, this NMI measure returns a score $\mathrm{NMI}(G_1; G_2) \approx 0$. **b** Degree-corrected normalized mutual information (DC-NMI, Eq. (16)) between graphs G_1 and G_2. Due to little overlap in node neighborhoods, we also see a low value for $\text{DC-NMI}(G_1; G_2)$. **c** Mesoscale normalized mutual information (MesoNMI, Eq. (30)) between networks G_1 and G_2 with respect to the indicated node partition $\boldsymbol{b}$ into $B = 2$ groups (colored, circled sets of nodes). Since the mesoscale structure of these networks is quite similar, as indicated by the edge overlap $E_{12}^{(\boldsymbol{b})} = 12$, we have a high similarity value $\mathrm{MesoNMI}^{(\boldsymbol{b})}(G_1; G_2) \approx 1$. **d** MesoNMI between the two networks but with respect to a different partition $\boldsymbol{b}$ with $B = 4$ groups (colored, circled sets of nodes). Here we still see a relatively high MesoNMI value, indicating substantial shared structure at this smaller scale. For reference, the Jaccard index among the edge sets in this case is $|G_1 \cap G_2|/|G_1 \cup G_2| = 0.315$—a much higher value than the NMI in (**a**)—indicating that the edge overlap is not much different than expected based on the network densities.

similarity value. Therefore, our measures will primarily be of interest for comparing node-aligned networks generated from cross-sectional studies with identical constraints across subjects (e.g., brain networks among a set of patients[35]), longitudinal studies (e.g., multiple observations of a social network among the same set of students[36]), and experimental studies with repeated measurements[37]. We will let $\mathcal{G}$ be the set of $\binom{N}{2}$ possible edges on these N nodes, so that $G_1, G_2 \subseteq \mathcal{G}$.

Analogous to constructing mutual information among labelings[20], we will derive our network mutual information measure by first considering the information required to transmit the network G_2 to a receiver using a particular encoding scheme. Then we will see by how much we can decrease our information cost if we transmit G_1 and exploit the graphs' shared structure to transmit G_2. The information savings we incur is the mutual information between G_1 and G_2 under the specified encoding.

We will assume that the receiver knows the number of nodes N and edges $E_2 = |G_2|$ in the second graph. (Transmitting this quantity will require comparatively negligible information, so we can ignore it anyway). There are $\binom{N}{2}$ possible unique undirected edges in $\mathcal{G}$ that one can construct using N nodes, so there are $\binom{\binom{N}{2}}{E_2}$ possible networks G_2 consistent with the constraints known by the receiver, among which we must specify a single one to transmit the graph structure G_2. The base-two logarithm of this quantity is therefore the approximate number of bits we will need to encode all these possibilities using binary strings if we choose a simple encoding that does not involve transmitting any intermediate summary information about G_2 to the receiver. (We will describe two possibilities for such an encoding later on.) We call this quantity of information the *entropy* $\mathrm{H}(G_2)$ of the graph G_2 under this simple encoding scheme, and it is given by

$$\mathrm{H}(G_2) = \log\binom{\binom{N}{2}}{E_2}. \tag{1}$$

For simplicity of presentation, we will assume all logarithms are base-two for the remainder of the paper.

We can simplify Eq. (1) into a more recognizable form using the Stirling approximation $\log x! \approx x\log x - x/\ln(2)$, giving

$$\mathrm{H}(G_2) \approx \binom{N}{2}\mathrm{H}_b(p_2), \tag{2}$$

where $p_2 = E_2/\binom{N}{2}$ is the fraction of all possible edges occupied in the graph and

$$\mathrm{H}_b(p) = -p\log p - (1-p)\log(1-p) \tag{3}$$

is the binary entropy function. Equation (2) tells us that it takes approximately $\mathrm{H}_b(p_2)$ bits of information to specify the existence or non-existence of an edge for each of the $\binom{N}{2}$ possible edge slots, given that the receiver knows there will be E_2 total edges.

Now, consider the case where the receiver already knows G_1 prior to us transmitting G_2. In this case, the information required to transmit G_2 should be reduced, since we can exploit the information shared between G_1 and G_2 to reduce the size of our encoding. More specifically, this reduction of information is possible when the receiver knows both G_1 and how G_2 differs from G_1, since this additional constraint further reduces the number of possibilities for G_2 that need to be encoded. In the standard formulation of mutual information between labelings, the discrepancies between the labelings are encoded in a *contingency table*, which counts the number of instances in which an object is classified into one cluster under the first labeling and another cluster in the second labeling. Typically one ignores the amount of information required to transmit the contingency table between two labelings to the receiver, as its information content vanishes in the limit of large labelings[20].

We can consider the labeling associated with the edge set G_i as a length-$\binom{N}{2}$ binary vector whose indices represent all possible edges in $\mathcal{G}$, and that contains a 1 for all entries corresponding to the edges that are present in G_i. (Equivalently, one can just consider flattening the upper triangle of the adjacency matrix representation of G_i into a vector). In this case, the contingency table comparing the labels in G_1 and G_2 takes a particularly simple form. For any possible edge (i, j) with $1 \le i, j \le N$, we can either have that:

1. $(i,j) \in G_1$ and $(i,j) \in G_2$. We call this subset of edges *true positives* to indicate that they are contained in both sets G_1 and G_2. The set of true positives is given by $G_1 \cap G_2$, whose size we denote with $E_{12} = |G_1 \cap G_2|$.
2. $(i,j) \in G_1$ and $(i,j) \notin G_2$. We call this subset of edges *false negatives* to indicate that they are contained in the set G_1 but not the set G_2. The set of false negatives is given by $G_1 \backslash G_2$, whose size is $E_1 - E_{12}$.
3. $(i,j) \notin G_1$ and $(i,j) \in G_2$. We call this subset of edges *false positives* to indicate that they are not contained in the set G_1 but are contained in the set G_2. The set of false positives is given by $G_2 \backslash G_1$, whose size is $E_2 - E_{12}$.
4. $(i,j) \notin G_1$ and $(i,j) \notin G_2$. We call this subset of edges *true negatives* to indicate that they are not contained in either of G_1 or G_2. The set of true negatives is given by $\mathcal{G} - (G_1 \cup G_2)$, whose size is $\binom{N}{2} - E_1 - E_2 + E_{12}$.These true/false positives/negatives are the four values of the contingency table between the binary labelings associated with G_1 and G_2.

Given that the receiver knows G_1, E_2, and the contingency table—or, equivalently, the number of true positives E_{12} since this is the only linearly independent unknown—we can compute the logarithm of the number of possible configurations that G_2 can take under these constraints as the *conditional entropy* $\mathrm{H}(G_2|G_1)$. This is given by

$$\mathrm{H}(G_2|G_1) = \log\binom{E_1}{E_{12}}\binom{\binom{N}{2} - E_1}{E_2 - E_{12}}. \tag{4}$$

The first term in the product counts the number of ways to choose the E_{12} true positives from the set of edges in G_1, and the second term counts the number of ways to choose the $E_2 - E_{12}$ false positives from the remaining edges not contained in G_1. The two actions together, which fully specify G_2, can be taken independently and so the total number of combinations available is given by the product of the two binomial coefficients.

We can now quantify the amount of information shared between the graphs G_1 and G_2 as the reduction in the entropy of G_2 that results from knowing G_1 and the contingency table. We call this the *mutual information* between the graphs, and it can be represented mathematically as

$$\mathrm{MI}(G_1, G_2) = \mathrm{H}(G_2) - \mathrm{H}(G_2|G_1). \tag{5}$$

Grouping terms and applying Stirling's approximation, we arrive at a simple, manifestly symmetric form for the graph mutual information:

$$\mathrm{MI}(G_1, G_2) \approx \binom{N}{2}[\mathrm{H}_b(p_1) + \mathrm{H}_b(p_2) - \mathrm{H}_s(\mathbf{P}_{12})], \tag{6}$$

where

$$\mathbf{P}_{12} = \{p_{12}, p_1 - p_{12}, p_2 - p_{12}, 1 - p_1 - p_2 + p_{12}\} \tag{7}$$

is the (normalized) contingency table between the labelings corresponding to G_1 and G_2—encoding the fraction of all $\binom{N}{2}$ possible edge slots that are true positives, false negatives, false positives, and true negatives respectively —and H_s is the Shannon entropy

$$\mathrm{H}_s(\boldsymbol{p}) = -\sum_i p_i \log p_i. \tag{8}$$

Typically, mutual information measures are written in terms of bits per symbol rather than total bits (the units of Eq. (6)). Dividing out the prefactor of $\binom{N}{2}$, we get the more familiar form of the mutual information

$$\mathrm{I}(\mathrm{G}_1;\mathrm{G}_2) = \mathrm{H}_\mathrm{b}(\mathrm{p}_1) + \mathrm{H}_\mathrm{b}(\mathrm{p}_2) - \mathrm{H}_\mathrm{s}(\mathbf{P}_{12}), \tag{9}$$

which just corresponds to the Shannon mutual information[33] between the binary vectors encoding the edge positions in G_1 and G_2.

The graph mutual information of Eq. (9) takes the form of the standard Shannon mutual information and therefore is bounded in the interval $0 \le \mathrm{I}(G_1; G_2) \le \mathrm{H}_b(p_1), \mathrm{H}_b(p_2)$, allowing for the normalized expression

$$\mathrm{NMI}(\mathrm{G}_1;\mathrm{G}_2) = 2 \times \frac{\mathrm{I}(\mathrm{G}_1;\mathrm{G}_2)}{\mathrm{H}_\mathrm{b}(\mathrm{p}_1) + \mathrm{H}_\mathrm{b}(\mathrm{p}_2)}. \tag{10}$$

Equation (10) maps the graph mutual information to the interval [0, 1] to allow for easier interpretation across systems of different sizes (there are many other options for normalizing the mutual information, which have their own benefits and drawbacks[22]).

We note that, as with any mutual information measure, Eq. (9) is invariant to label permutations in the binary representation of the graphs: $\mathrm{I}(\mathrm{G}_1;\mathrm{G}_2) = \mathrm{I}(\mathrm{G}_1;\overline{\mathrm{G}_2}) = \mathrm{I}(\overline{\mathrm{G}_1};\mathrm{G}_2) = \mathrm{I}(\overline{\mathrm{G}_1};\overline{\mathrm{G}_2})$, where $\overline{G_i} = \mathcal{G} - G_i$ is the graph complement of G_i. This follows intuitively, since specifying the positions of edges is equivalent to specifying the positions of nonedges in terms of information content. In practice, however, this symmetry will rarely ever have an effect on results, since we are nearly always in the sparse regime where $p_1, p_2 < 0.5$, so the mutual information is monotonic in the overlap p_{12} for fixed p_1, p_2.

Computing the graph mutual information measure in Eq. (9) is very fast in practice, only requiring us to find the sizes of the edge sets G_1, G_2, and $G_1 \cap G_2$. The total complexity of these calculations is equal to the complexity of constructing the sets themselves, so poses no additional computational burden.

We can adapt the mutual information of Eq. (9) to networks with directed and/or self-edges by simply changing the constant $\binom{N}{2}$ for the size of $\mathcal{G}$ from which we must pick the edges in our graph. For directed graphs with self-edges, directed graphs without self-edges, and undirected graphs with self-edges, we can set this constant to be N^2, $N(N-1)$, and $\binom{N}{2} + N$, respectively. One can also adapt our network mutual information framework to multigraphs by using the formulation we present in the "Mesoscale network mutual information" section and putting each node in their own group in the input partition.

One can also derive from Eq. (9) a variation of information measure[38] between graphs, thus

$$\mathrm{VI}(\mathrm{G}_1;\mathrm{G}_2) = \mathrm{H}_\mathrm{b}(\mathrm{p}_1) + \mathrm{H}_\mathrm{b}(\mathrm{p}_2) - 2 \times \mathrm{I}(\mathrm{G}_1;\mathrm{G}_2). \tag{11}$$

Equation (11) has an advantage over the graph mutual information for various tasks such as network embedding[39] due to its pseudometric property, but we will not explore its applications here.

Degree-corrected network mutual information

The measure presented in the "Network mutual information" section is based on an encoding scheme in which the receiver's knowledge of the overlap ("true positive" count) E_{12} between the graphs G_1 and G_2 constrains the number of possibilities for G_2 once G_1 is known. This allows for a reduction in the information required to specify G_2 after G_1 is sent, giving a mutual information measure. However, one can also consider modifying the encoding process to exploit other shared structure between the networks prior to the transmission of G_2, which results in a mutual information that captures a different notion of structural similarity between the graphs.

One such modification is "degree-correction", analogous to the degree correction of the stochastic block model (SBM) for community detection[40] in which the degree sequence is specified ahead of time to provide additional compression of network data through its modular structure. Here, instead of specifying the global edge overlap E_{12} between G_1 and G_2, we can specify how much overlap there is between the individual neighborhoods $\partial_i^{(1)}$ and $\partial_i^{(2)}$ of each node i in G_1 and G_2, respectively. In order to do this, a two-step process is required: firstly, we must list the degree sequence $\boldsymbol{k}_i = \{k_i(1), \ldots, k_i(N)\}$ of each graph i. Second, we must specify how the total edge overlap E_{12} is distributed among the N node neighborhoods. These two steps have information content that scales like $O(E_1 + E_2)$, which for sparse graphs can be neglected when we normalize by $\binom{N}{2}$ to get the bits per symbol value of the mutual information. We therefore ignore this intermediate information cost, which we will see leads to a nice clean expression for the degree-corrected graph mutual information.

Given knowledge of the degrees $\boldsymbol{k}_2$ and how the E_{12} overlapping edges are distributed across the node neighborhoods (i.e., the rows of G_2's adjacency matrix representation), we can compute the conditional entropy between G_1 and G_2 as

$$\mathrm{H}_{\mathrm{deg}}(G_2|G_1) = \sum_{i=1}^{N} \log\binom{k_1(i)}{k_{12}(i)}\binom{N - k_1(i)}{k_2(i) - k_{12}(i)}, \tag{12}$$

where $k_{12}(i) = |\partial_i^{(1)} \cap \partial_i^{(2)}|$ is the number of true positives (e.g., overlapping edges with G_1) attached to node i. The first binomial coefficient in the summand counts the number of possible configurations for the overlapping edges (i.e., those shared with G_1) attached to node i in G_2. Meanwhile, the second term counts the number of possible configurations for the non-overlapping edges (i.e., those not shared with G_1) attached to node i in G_2.

The analogous entropy expression for G_2 if the receiver does not have knowledge of any of the shared structure with G_1—but does know the degrees $\boldsymbol{k}_2$, which are independent of G_1—is given by

$$\mathrm{H}_{\mathrm{deg}}(G_2) = \sum_{i=1}^{N} \log\binom{N}{k_2(i)}. \tag{13}$$

This is just the amount of information required to specify the nodes attached to i given its degree. Technically both Eqs. (12) and (13) are only upper bounds on the conditional entropy and entropy for undirected graphs under this encoding scheme, since only one edge direction must be known to specify each edge. They are, however, exact for directed graphs in which the degree sequences $\boldsymbol{k}_i$ can be chosen to be either the in- or out-degrees.

By comparing Eqs. (13) and (12) with Eqs. (1) and (4), we can immediately identify the mutual information value for this degree-corrected

scheme as the average of node-level mutual information values, thus

$$\mathrm{I}_{\mathrm{deg}}(G_1;G_2)=\frac{1}{N}\sum_{i=1}^{N}[\mathrm{H}_b(p_1(i))+\mathrm{H}_b(p_2(i))-\mathrm{H}_s(\mathbf{P}_{12}(i))], \tag{14}$$

where $p_1(i)=k_1(i)/(N-1)$, $p_2(i)=k_2(i)/(N-1)$, and

$$\begin{aligned}\mathbf{P}_{12}(i)=\{&p_{12}(i),p_1(i)-p_{12}(i),p_2(i)-p_{12}(i),\\&1-p_1(i)-p_2(i)+p_{12}(i)\},\end{aligned} \tag{15}$$

with $p_{12}(i)=k_{12}(i)/(N-1)$. We normalize by $N-1$ for graphs without self-edges since a node can connect to at most $N-1$ nodes excluding itself. For networks with self-edges, one can change the normalization to N to account for the possibility of a node that is completely connected to all other nodes, including itself.

A degree-corrected NMI measure can be constructed for Eq. (14) by noting that $\frac{1}{2}[\mathrm{H}_b(p_1(i))+\mathrm{H}_b(p_2(i))]$ is an upper bound for each summand $\mathrm{H}_b(p_1(i))+\mathrm{H}_b(p_2(i))-\mathrm{H}_s(\mathbf{P}_{12}(i))$ in Eq. (14), each of which is a mutual information measure at the node neighborhood level. We can therefore replace the summand in Eq. (14) with this upper bound to obtain an upper bound on the total degree-corrected mutual information. The resulting normalized mutual information measure is given by

$$\mathrm{DC-NMI}(\mathrm{G_1};\mathrm{G_2})=2\times\frac{\mathrm{I}_{\mathrm{deg}}(\mathrm{G_1};\mathrm{G_2})}{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}}[\mathrm{H_b}(\mathrm{p_1(i)})+\mathrm{H_b}(\mathrm{p_2(i)})]}, \tag{16}$$

which will be bounded to [0, 1]. As before, we can also construct a degree-corrected variation of information measure for networks, thus

$$\begin{aligned}\mathrm{VI}_{\mathrm{deg}}(G_1;G_2)=&\frac{1}{N}\sum_{i=1}^{N}[\mathrm{H}_b(p_1(i))+\mathrm{H}_b(p_2(i))]\\&-2\times\mathrm{I}_{\mathrm{deg}}(G_1;G_2).\end{aligned} \tag{17}$$

One can show that the above expression is also a pseudometric like the usual variation of information.

Similar to the non-degree-corrected mutual information measure of Eq. (10), the degree-corrected mutual information measures presented here can be computed with a time complexity that is linear in the size of the edge sets G_1 and G_2 being compared. Once the set intersection $G_1\cap G_2$ has been computed, we just need to iterate through these shared edges again and increment $k_{12}(i)$ whenever node i is present in the edge.

In contrast with the measures of the "Network mutual information" section, which are unaffected by where edges differ between the two graphs, the degree-corrected measures presented here will ascribe higher similarity to graphs whose edge discrepancies are concentrated on relatively few nodes. In this sense, the DC-NMI of Eq. (16) focuses on node-level structural similarity, while the NMI of Eq. (10) focuses on edge-level similarity. We will see from experiments in the Results section that this distinction becomes important for graphs with heterogeneous degrees since differences in network structure can naturally become concentrated on high-degree nodes, particularly in the case of random edge rewiring.

Mesoscale network mutual information

While the measures derived in the "Network mutual information" and "Degree-corrected network mutual information" sections will capture the small-scale information shared between two graphs G_1 and G_2 (e.g., the overlap of specific edges and node neighborhoods), they will fail to capture mesoscale structure such as communities or core-periphery structure unless the exact positions of the edges within these larger-scale structures happen to overlap. Indeed, if G_1 and G_2 are both sparse networks generated from an SBM[41], there is a very low probability that they share a substantial fraction of edges despite having very similar mesoscale divisions into communities since each community is itself a sparse random graph whose edge positions are uncorrelated. In this case, the measures we have discussed will likely ascribe little similarity to the two graphs despite their identical mesoscale community structure. To obtain meaningful graph similarity results at different scales of interest—particularly at the mesoscale where community structure is present—it is therefore important to consider network mutual information formulations that take larger-scale structure into account while ignoring small-scale details. Here we present one possible formulation of such a mutual information measure between graphs which we call the *mesoscale graph mutual information*.

Consider the same problem setting as in the "Network mutual information" section, where we have undirected, unweighted, node-aligned graphs G_1 and G_2 with N nodes and E_1, E_2 edges, respectively. Assume again that N, E_1, E_2 are already known by the receiver. This time we will always allow G_1 and G_2 to potentially have self- or multi-edges since this will allow for easier computations, but in principle the calculation can be extended to graphs without self- or multi-edges with more detailed combinatorics. If G_1 and G_2 share mesoscale structure (e.g., in the form of groups of highly connected nodes), but not necessarily microscale structure (in the form of overlapping edges), then the encoding schemes of the previous sections will be very inefficient, since there is little shared information at the microscale we can exploit to reduce the entropy of our transmission. However, if we instead only aim for a mesoscale description of each network—such as the edge counts within and between communities in a given partition of the graphs—then we can formulate a mutual information measure that captures the shared mesoscale information between G_1 and G_2 while ignoring their microscale differences.

Here we consider partitioning the nodes in both graphs G_1 and G_2 with the same (non-overlapping) partition $\boldsymbol{b}$ with B groups. We will denote with $n_r=\sum_{i=1}^{N}\delta_{b_i,r}$ the number of nodes with community label r in the partition $\boldsymbol{b}$. The partition $\boldsymbol{b}$ can be thought of as a specific coarse-graining of the networks, and tuning B allows us to interpolate between the microscale ($B\sim\mathrm{O}(N)$) and the macroscale ($B\sim\mathrm{O}(1)$) to capture similarities at the scale of interest. A reasonable choice for $\boldsymbol{b}$ once the scale B is chosen is a community partition of one of the networks into B groups, since this represents a meaningful coarse-graining of the network at this scale. However, in principle any choice for $\boldsymbol{b}$ is possible, and we will show in "Results" that often one can choose meaningful coarse-grainings $\boldsymbol{b}$ based on node metadata relevant to a particular application. We will let $\boldsymbol{b}$ be known to the receiver—its corresponding entropy term would vanish in our final mutual information expression anyway.

Under the node partition $\boldsymbol{b}$, each network G_i can be described by a coarse-grained representation $\tilde{G}_i^{(\boldsymbol{b})}$, defined as the multiset of E_i elements where element (r,s) with $r\le s$ has a multiplicity $m_i(r,s)$ equal to the number of edges in G_i containing one node in group r and one node in group s. (This is equivalent to the mixing matrix representation of community-community ties used in the microcanonical SBM[42]). Defining the scale of the full network G_i to be order O(1), the representation $\tilde{G}_i^{(\boldsymbol{b})}$ captures the aggregate structure of G_i at a scale of order $\mathrm{O}(B^{-1})$. In the extreme case with $B=N$ (the partition $\boldsymbol{b}$ puts each node into its own group), we have $\tilde{G}_i^{(\boldsymbol{b})}=G_i$ and we are capturing network similarity at the scale $\mathrm{O}(N^{-1})$. The measure we present in the case $B=N$ can thus be used as mutual information between multigraphs G_1 and G_2.

The mesoscale mutual information measure $\mathrm{I}_{\boldsymbol{b}}(G_1;G_2)$ that we will derive aims to capture the amount of shared information at the scale $\mathrm{O}(B^{-1})$ between the graphs G_1 and G_2 by computing the mutual information between the multisets $\tilde{G}_1^{(\boldsymbol{b})}$ and $\tilde{G}_2^{(\boldsymbol{b})}$ for a chosen partition $\boldsymbol{b}$. (As discussed, $\boldsymbol{b}$ can be derived by using network structure itself or by using external metadata). Instead of formulating the mutual information from the perspective of conditional entropy, we will motivate the mesoscale mutual information from a *joint* transmission process in which we transmit both $\tilde{G}_1^{(\boldsymbol{b})}$ and $\tilde{G}_2^{(\boldsymbol{b})}$—first individually, then together using their shared information. Formulations of mutual information measures using conditional

and joint entropies are equivalent, but in our case the latter allows for a more straightforward exposition.

We can first consider transmitting each individual multiset, $\tilde{G}_i^{(\boldsymbol{b})}$, separately to a receiver. Using the same calculation procedure as with the entropy measures above, we can find these individual entropies to be

$$\mathrm{H}(\tilde{G}_i^{(\boldsymbol{b})}) = \log\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_i \right)\right), \tag{18}$$

where

$$\left(\binom{n}{k}\right) = \binom{n+k-1}{k} \tag{19}$$

is the multiset coefficient. In Eq. (18), $\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_i \right)\right)$ is the number of multisets $\tilde{G}_i^{(\boldsymbol{b})}$ with E_i elements one can construct from a set of objects with cardinality $\binom{B}{2} + B$, that is, the number of independent combinations (r, s) with $1 \le r \le s \le B$. The logarithm of this quantity is thus the entropy of our encoding for specifying $\tilde{G}_i^{(\boldsymbol{b})}$ given the constraints known by the receiver, and transmitting the two multisets individually therefore requires $\mathrm{H}(\tilde{G}_1^{(\boldsymbol{b})}) + \mathrm{H}(\tilde{G}_2^{(\boldsymbol{b})})$ bits of information.

An important property of the multiset coefficient that we will use when constructing our measure is that it is subadditive when transformed by a logarithm to get an entropy. In other words, for any k, l we have

$$\log\left(\binom{n}{k}\right) + \log\left(\binom{n}{l}\right) \ge \log\left(\binom{n}{k+l}\right). \tag{20}$$

To prove this, define $X^{(n,k)}$ as the set of non-negative integer vectors of length n whose values sum to k. The multiset coefficient counts the number of unique vectors in $X^{(n,k)}$, i.e., $|X^{(n,k)}| = \left(\binom{n}{k}\right)$. We can construct a map $f : X^{(n,k)} \times X^{(n,l)} \to X^{(n,k+l)}$ given by $f(x, y) = x + y$. The map f is surjective, since any $z \in X^{(n,k+l)}$ has at least one pair $(x, y) \in X^{(n,k)} \times X^{(n,l)}$ such that $f(x, y) = z$. Therefore,

$$\left(\binom{n}{k}\right)\left(\binom{n}{l}\right) = |X^{(n,k)} \times X^{(n,l)}| \ge |X^{(n,k+l)}| = \left(\binom{n}{k+l}\right), \quad \text{and}$$

taking the logarithm of both sides gives the result in Eq. (20).

Now we can consider transmitting $\tilde{G}_1^{(\boldsymbol{b})}$ and $\tilde{G}_2^{(\boldsymbol{b})}$ together, exploiting the shared information between them to reduce the total entropy of the transmission. Using the generalization of the set intersection to multisets[43], we can define the true positives $E_{12}^{(\boldsymbol{b})}$ in this context as

$$\begin{aligned} E_{12}^{(\boldsymbol{b})} &= |\tilde{G}_1^{(\boldsymbol{b})} \cap \tilde{G}_2^{(\boldsymbol{b})}| \\ &= \sum_{r \le s} \min(m_1(r,s), m_2(r,s)), \end{aligned} \tag{21}$$

where the index r iterates over $1, \ldots, B$, and the index s iterates over $r, \ldots, B$. Equation (21) tells us the total number of common group pairs (r, s) (allowing duplicates) between the multisets $\tilde{G}_1^{(\boldsymbol{b})}$ and $\tilde{G}_2^{(\boldsymbol{b})}$. The information required to transmit $E_{12}^{(\boldsymbol{b})}$ is of $\mathrm{O}(\log(\mathrm{E}_1 + \mathrm{E}_2))$ and can be ignored. With $E_{12}^{(\boldsymbol{b})}$ known, there are $\log\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_{12}^{(\boldsymbol{b})} \right)\right)$ ways to distribute the overlapping pairs among $\tilde{G}_1^{(\boldsymbol{b})}$ and $\tilde{G}_2^{(\boldsymbol{b})}$. Once the overlapping pairs are known, we must specify $E_i - E_{12}$ remaining pairs for each of the multisets $\tilde{G}_i^{(\boldsymbol{b})}$. Putting this all together gives a total joint information of

$$\mathrm{H}' = \log\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_{12}^{(\boldsymbol{b})} \right)\right)\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_1 - E_{12}^{(\boldsymbol{b})} \right)\right)\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_2 - E_{12}^{(\boldsymbol{b})} \right)\right). \tag{22}$$

In principle, with the joint entropy of Eq. (22) one can construct a mesoscale mutual information measure of $\mathrm{H}(\tilde{G}_1^{(\boldsymbol{b})}) + \mathrm{H}(\tilde{G}_2^{(\boldsymbol{b})}) - \mathrm{H}'$. While this has some desirable properties, it is not strictly increasing with the overlap $E_{12}^{(\boldsymbol{b})}$ due to low overlap values placing substantial constraints on the multisets $\tilde{G}_1^{(\boldsymbol{b})}$ and $\tilde{G}_2^{(\boldsymbol{b})}$ in some cases, providing an efficient reduction in the entropy and a high mutual information. Intuitively, one would expect that a higher overlap $E_{12}^{(\boldsymbol{b})}$ among the multisets $\tilde{G}_1^{(\boldsymbol{b})}$ and $\tilde{G}_2^{(\boldsymbol{b})}$ should result in a higher similarity value, and according to this criterion the mutual information $\mathrm{H}(\tilde{G}_1^{(\boldsymbol{b})}) + \mathrm{H}(\tilde{G}_2^{(\boldsymbol{b})}) - \mathrm{H}'$ is unsuitable as a candidate for a similarity measure. Although the graph mutual information of Eq. (9) technically also has a similar limitation due to invariance under graph complements, as discussed in the "Network mutual information" section this is irrelevant in the sparse regime where $E_i < \binom{N}{2}/2$.

Instead of Eq. (22), it turns out that a lower bound on the joint entropy can be used to produce a useful mutual information similarity measure that does exhibit monotonicity with the overlap $E_{12}^{(\boldsymbol{b})}$ as desired. Using the subadditivity of the log-multiset coefficients in Eq. (20), we have the bound

$$\mathrm{H}' \ge \log\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_1 + E_2 - E_{12}^{(\boldsymbol{b})} \right)\right). \tag{23}$$

Using this bound on the joint entropy

$$\mathrm{H}(\tilde{G}_1^{(\boldsymbol{b})}, \tilde{G}_2^{(\boldsymbol{b})}) = \log\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_1 + E_2 - E_{12}^{(\boldsymbol{b})} \right)\right), \tag{24}$$

we can now compute the mesoscale mutual information of graphs G_1 and G_2 with respect to partition $\boldsymbol{b}$ as

$$\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}(G_1; G_2) = \mathrm{H}\left(\tilde{G}_1^{(\boldsymbol{b})}\right) + \mathrm{H}\left(\tilde{G}_2^{(\boldsymbol{b})}\right) - \mathrm{H}\left(\tilde{G}_1^{(\boldsymbol{b})}, \tilde{G}_2^{(\boldsymbol{b})}\right). \tag{25}$$

For fixed E_1, E_2, we have that for any $E_{12}^{(\boldsymbol{b})} \in [1, \min(E_1, E_2)]$ the mesoscale mutual information of Eq. (25) satisfies

$$\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}\left(E_{12}^{(\boldsymbol{b})}\right) - \mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}\left(E_{12}^{(\boldsymbol{b})} - 1\right) = \log \frac{\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_1 + E_2 - E_{12}^{(\boldsymbol{b})} + 1 \right)\right)}{\left(\left(\begin{pmatrix} B \\ 2 \end{pmatrix} + B \\ E_1 + E_2 - E_{12}^{(\boldsymbol{b})} \right)\right)} \ge 0, \tag{26}$$

since the multiset coefficient $\left(\binom{n}{k}\right)$ is a strictly increasing function of k. Here we considered the mutual information as a function of only the overlap $E_{12}^{(\boldsymbol{b})}$ since this is the only remaining free parameter when E_1, E_2 are fixed. Equation (26) implies that, when all else is constant, a greater overlap $E_{12}^{(\boldsymbol{b})}$

among $\tilde{G}_1^{(\boldsymbol{b})}$ and $\tilde{G}_2^{(\boldsymbol{b})}$ gives a greater mesoscale mutual information $\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}(G_1;G_2)$, which is consistent with what we expect from a similarity measure.

From the previous argument, we can also see that

$$\begin{aligned}\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})} &\geq \mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}\left(E_{12}^{(\boldsymbol{b})}=0\right)\\ &= \log\frac{\left(\left(\binom{B}{2}+B\right)\right)\left(\left(\binom{B}{2}+B\right)\right)}{\left(\left(\binom{B}{2}+B\right)\right)}\\ &\geq 0,\end{aligned} \tag{27}$$

or in other words, the mesoscale mutual information is bounded below by 0. This follows from the subadditivity of the logarithm of the multiset coefficient.

We can also use the inequality of Eq. (26) to establish an upper bound for the mesoscale mutual information. Without loss of generality, label the graphs G_i so that $E_1 \leq E_2$. We then have that the maximum overlap is $E_{12}^{(\boldsymbol{b})} = E_1$, and so

$$\begin{aligned}\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})} &\leq \mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}\left(E_{12}^{(\boldsymbol{b})}=E_1\right)\\ &= \log\frac{\left(\left(\binom{B}{2}+B\right)\right)\left(\left(\binom{B}{2}+B\right)\right)}{\left(\left(\binom{B}{2}+B\right)\right)}\\ &= \mathrm{H}\left(\tilde{G}_1^{(\boldsymbol{b})}\right)\\ &\leq \frac{\mathrm{H}\left(\tilde{G}_1^{(\boldsymbol{b})}\right)+\mathrm{H}\left(\tilde{G}_2^{(\boldsymbol{b})}\right)}{2},\end{aligned} \tag{28}$$

where the last inequality uses the fact that the multiset coefficient $\left(\binom{n}{k}\right)$ is an increasing function of k and that $E_1 \leq E_2$.

Using the bounds of Eqs. (27) and (28), we can construct a normalized mesoscale mutual information $\mathrm{NMI}_{\mathrm{meso}}^{(\boldsymbol{b})}$ as follows

$$\mathrm{NMI}_{\mathrm{meso}}^{(\boldsymbol{b})}(G_1;G_2) = 2\times\frac{\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}(G_1;G_2)}{\mathrm{H}\left(\tilde{G}_1^{(\boldsymbol{b})}\right)+\mathrm{H}\left(\tilde{G}_2^{(\boldsymbol{b})}\right)}, \tag{29}$$

which mirrors the form of the graph NMI of Eq. (10). In practice, it is also useful to consider an alternative normalization with a tighter lower bound, thus

$$\mathrm{MesoNMI}^{(\boldsymbol{b})}(G_1;G_2) = \frac{\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}(G_1;G_2)-\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}\left(E_{12}^{(\boldsymbol{b})}=0\right)}{\frac{\mathrm{H}\left(\tilde{G}_1^{(\boldsymbol{b})}\right)+\mathrm{H}\left(\tilde{G}_2^{(\boldsymbol{b})}\right)}{2}-\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}\left(E_{12}^{(\boldsymbol{b})}=0\right)}, \tag{30}$$

where

$$\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}\left(E_{12}^{(\boldsymbol{b})}=0\right) = \log\frac{\left(\left(\binom{B}{2}+B\right)\right)\left(\left(\binom{B}{2}+B\right)\right)}{\left(\left(\binom{B}{2}+B\right)\right)}. \tag{31}$$

Both Eqs. (29) and (30) will fall in the range $[0,1]$, with $\mathrm{NMI}_{\mathrm{meso}}^{(\boldsymbol{b})}(G_1;G_2)=1$ if and only if $G_1 = G_2$. In our experiments, we will use the mesoscale NMI measure defined in Eq. (30).

The mesoscale mutual information above requires the user to choose a common partition $\boldsymbol{b}$ of the networks G_1 and G_2, which effectively sets the scale of interest for comparing the similarity of the two graphs. One can choose the partition $\boldsymbol{b}$ in a number of ways, a meaningful choice being a community partition of either G_1 or G_2 that decomposes the graph into densely connected groups of nodes with sparser connections between groups. In principle, one can also maximize or minimize $\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}(G_1;G_2)$ over all possible partitions $\boldsymbol{b}$ with B groups, which can identify the divisions for which the networks are most/least similar at the chosen scale. Similarly, one could sample over partitions $\boldsymbol{b}$ at a given scale B and examine the full distribution of values $\mathrm{I}_{\mathrm{meso}}^{(\boldsymbol{b})}(G_1;G_2)$ to get a multifaceted assessment of the similarity between G_1 and G_2 at the chosen scale. Finally, as we do in the experiments here, one can use node metadata to define the partition $\boldsymbol{b}$ by grouping nodes with similar characteristics.

All of the proposed measures highlight significant shared structure because they are constructed to compute the amount of shared information between two graphs G_1 and G_2 with respect to a particular property of interest—individual edges for the NMI of Eq. (10), node neighborhoods for the DC-NMI of Eq. (16), and edge densities within groupings of nodes for the MesoNMI of Eq. (30). If there is redundant information in the two graphs that can be compressed under an encoding that exploits a particular property X of interest, this indicates that there is meaningful shared structure in the two graphs with respect to X. The same property makes the proposed measures easily interpretable: If one achieves a high NMI between two graphs, the networks have a meaningful share of overlapping edges; if one achieves a high DC-NMI between two graphs, the networks have a meaningful amount overlap among individual node neighborhoods; and if one achieves a high MesoNMI between two graphs, the networks have a meaningful overlap in the "super-edges" of the coarse-grained multigraph corresponding to the partition of interest. Finally, the proposed methods are fast: The total runtime complexity of each of the three proposed measures is $O(E_1+E_2)$, which is the same complexity as creating the edge sets G_1 and G_2 in the first place.

Figure 1 shows the application of all measures presented in this section to a pair of example networks.

Results

In this section, we apply our graph similarity measures in a variety of experiments with synthetic and empirical networks. First, we illustrate the differences between similarity measures by attacking synthetic networks at different scales. We find that the different encoding schemes indeed produce graph similarity measures that highlight shared structure at different scales and are affected differently by these structural perturbations. Then, we apply our measures to a case study with an empirical multilayer network of global trade patterns, finding that our measures capture meaningful shared structure among the network layers representing the movement of different goods.

Similarity scores among perturbations of synthetic networks

We examine the extent to which each of our similarity measures captures different structural deviations by perturbing synthetic reference networks with different noise (or "attack") strategies. In each simulation, we first generate a reference network from a random graph model—here we used random networks generated from the Erdös–Rényi (ER) model[44,45], Barabási–Albert (BA) model[46], and the stochastic block model (SBM) with equally sized groups and mixing levels that only depend on whether the nodes are in the same group or different groups[41,47]. (This is also known as the planted partition model). These different reference graphs allow us to examine the effects of global network structure and degree heterogeneity on the similarity scores. We then perturb the reference graph by attacking it with one of two types of moves: (1) node attacks (Fig. 2a), which take all the

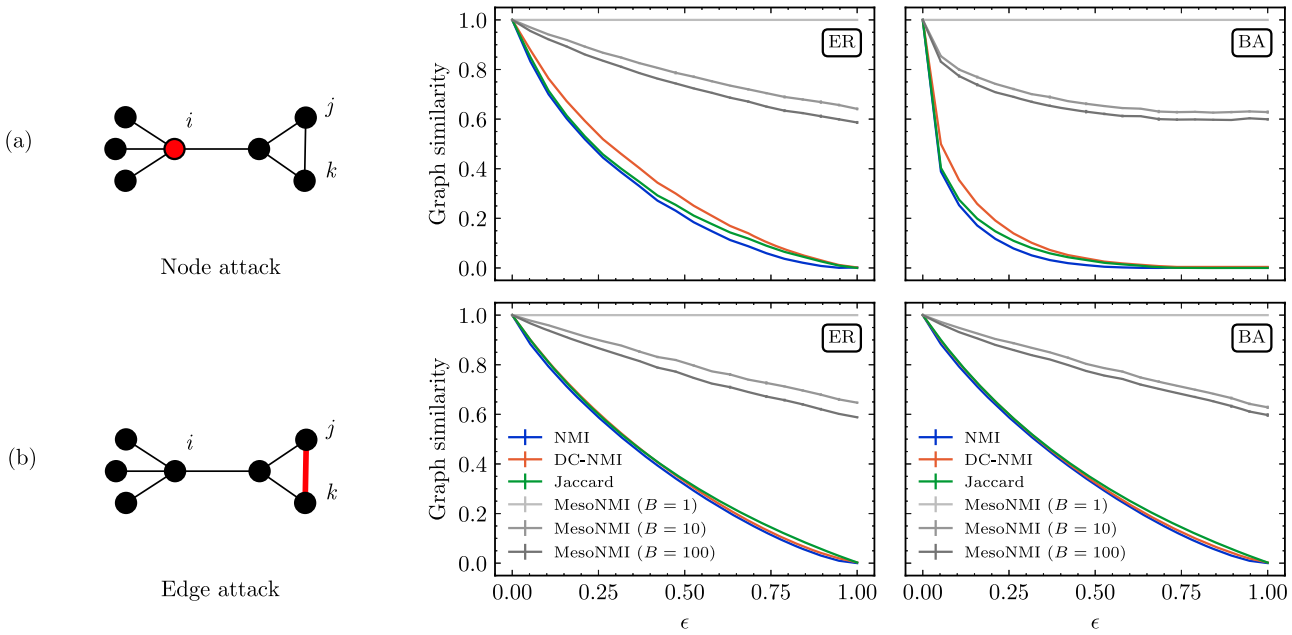

Fig. 2 | Graph similarity measures for networks under node and edge attacks. **a** Graph similarity as a function of the fraction of nodes attacked ϵ for random networks, where nodes are attacked in decreasing order of degree. Graph similarity is measured with the NMI, DC-NMI, and MesoNMI for $B \in \{1, 10, 100\}$ to capture multiple scales of similarity. The Jaccard index $|G_1 \cap G_2|/|G_1 \cup G_2|$ is included for comparison. The MesoNMI partitions $\boldsymbol{b}$ are computed with a standard stochastic block model (SBM) with fixed group sizes $B \in \{1, 10, 100\}$ on the initial (un-attacked) graph. Simulations are averaged over 10 realizations of the initial graph from the Erdös–Rényi model (ER, top left panel) and Barabási–Albert model (BA, top right panel) with $N = 1000$ nodes and average degree $\langle k \rangle = 10$ (error bars indicate three standard errors and are vanishingly small). **b** Graph similarity as a function of the fraction of edges randomly rewired, for the same synthetic networks. Subtle differences in the decay rates of different similarity measures are reflective of intuitive properties of these measures, as discussed in "Results".

edges incident to a node and rewire them uniformly at random to neighbors not currently connected to the attacked node; and (2) edge attacks (Fig. 2b), which take an edge (i, j) and place it between a different pair of nodes (i', j'), chosen uniformly at random from all nodes except i, j. We run our simulations by attacking nodes in decreasing order of degree and edges in a random order, measuring the extent to which the original network has been perturbed with an attack fraction ϵ indicating the fraction of nodes or edges that have been attacked.

It is worth noting that, for large N, the expected fraction of shared edges p_{12} is well-approximated by $p_{12} \approx p_1 p_2$, and one can show that the resulting value of the NMI measure of Eq. (10) is zero in this case. A similar argument can be made for the DC-NMI, and this is seen numerically for $\epsilon = 1$ in Fig. 2b.

Figure 2a, b shows graph similarity values for a number of measures versus the fraction ϵ of node and edge attacks, respectively. We include the Jaccard similarity $|G_1 \cap G_2|/|G_1 \cup G_2|$ for reference. In these experiments, the partitions $\boldsymbol{b}$ required to compute the MesoNMI measures from Eq. (30) are computed by fitting an SBM to the original (unperturbed) network with the indicated number of groups B. Each curve is the average over 10 simulations, each starting from a different initial graph. All networks in the experiment were of size $N = 1000$ and had an average degree of $\langle k \rangle = 10$. Unless stated otherwise, error bars indicate three standard errors. The reference networks were generated from an ER model (left panels) and a BA one (right panels), with the goal of capturing the effects of degree heterogeneity on how each similarity measure is penalized under node and edge attacks.

In all panels, we see steeper decays in the similarity values of the microscale measures—NMI, DC-NMI, and Jaccard index—than the MesoNMI measures (except at $B = 1$, which is trivially equal to 1 for all ϵ since the number of edges is constant throughout the attack process). We can also see a greater (albeit still modest) differentiation between the microscale measures for the node attacks than the edge attacks. The standard NMI measure and Jaccard index are unchanged between the two attack strategies, since it will be penalized equally no matter how edges are replaced. However, the DC-NMI is less penalized by the node attacks than the NMI or the Jaccard index. This is because the rewiring of a hub node i primarily impacts the DC-NMI (Eq. (16)) via the change in this node's new neighborhood overlap $p_{12}(i) = 0$, while i's old and new neighbors j may see only slight changes to their values $p_{12}(j)$ as these nodes often will have other neighbors that are unchanged after the attack. In contrast, all the edges (i, j) replaced due to the hub attack—which may constitute a substantial fraction of all edges in the network—will penalize equally the total edge overlap $|G_1 \cap G_2|$ considered by the NMI and Jaccard measures.

We can also see from Fig. 2a that degree heterogeneity has a substantial effect on the sensitivity of the similarity measures for the node attacks, since attacking nodes in decreasing order of degree results in more edges being rewired at a given attack fraction ϵ for heterogeneous (BA) than homogeneous (ER) degree distributions. However, this effect is not present in Fig. 2b for the edge attacks due to all edges providing a roughly equal contribution to each similarity measure.

The MesoNMI measures in all panels present fairly uniform patterns, with the decay in similarity becoming more severe as we increase the number of groups B in the partition $\boldsymbol{b}$ of the reference network. This is consistent with the attacks being performed at the microscale (e.g., on nodes and edges) rather than at the mesoscale: under the random rewiring of individual edges, the mixing structure within and between groups of the node partition remains less affected since these perturbations may produce new edges that run between the same pairs of groups. This effect is more pronounced as we decrease the number of groups B, since with few groups it is more likely that rewiring an edge (i, j) to a new edge (i', j') will result in the same pair of groups $b_{i'}, b_{j'}$ appearing on the ends of the edge. This robustness to microscale attacks also manifests in less sensitivity to the degree heterogeneity of the original graph, as can be seen in the top right panel.

Figure 3 shows an edge attack experiment, except we use an initial graph drawn from an SBM with two equal groups and tunable mixing level μ fixing the fraction of edges that run within groups. The value of $\mu = 0.5$ corresponds to no mixing preference between the two groups (i.e., the network is an ER random graph), while $\mu = 0$ corresponds to a completely disassortative structure in which all edges are between groups, and $\mu = 1$

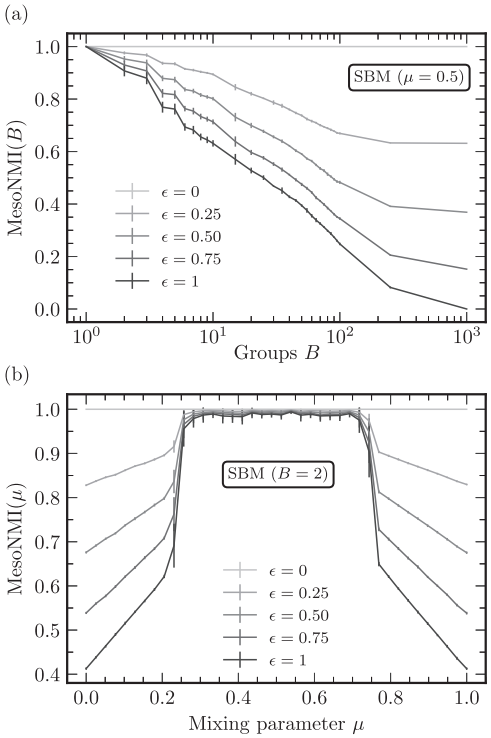

Fig. 3 | Mesoscale mutual information between stochastic block model (SBM) networks. Edge attack simulations were performed on networks generated from an SBM with two groups of 500 nodes, average degree of 10, and mixing level $\mu \in [0, 1]$ fixing the fraction of edges running between nodes of the same group identity. **a** MesoNMI values for different edge attack fractions ϵ (indicated by curves of shades of gray) as a function of the number of groups B of the node partition $\boldsymbol{b}$ used for the MesoNMI calculation, for SBM networks with no mixing preference ($\mu = 0.5$, equivalent to Erdös–Rényi random graphs). The MesoNMI is more sensitive to edge-level attacks as the scale of interest for comparison gets smaller (i.e., the number of groups B gets larger). **b** MesoNMI as a function of the mixing level μ of the initial graph being attacked, with the partition $\boldsymbol{b}$ used for the MesoNMI calculations being fixed as the initial graph's planted partition into $B = 2$ groups. As the mixing level moves away from $\mu = 0.5$, we see a stronger dependence of the MesoNMI on the attack level ϵ due to the rewiring of inter-community ($\mu = 0$) or intra-community ($\mu = 1$) edges to produce an equitable mixture of these two edge types in expectation.

corresponds to a completely assortative structure in which all edges are among nodes that have the same group affiliation. As before, we fix the number of nodes to be $N = 1000$ and the average degree of $\langle k \rangle = 10$. For the MesoNMI measures, we calculate the partition $\boldsymbol{b}$ in the same way as in Fig. 2, computing the SBM-optimal partition on the reference network with the desired value of B being fixed. Although the reference SBMs actually only have two groups in their planted structure, setting B to different values for the attack experiments allows us to examine the similarity of the perturbed networks with this reference graph at different scales using the MesoNMI measure.

In Fig. 3a, we plot the MesoNMI versus the number of groups B for different attack fractions ϵ (shades of gray) for an SBM with no mixing structure ($\mu = 0.5$, i.e., the reference network is an ER random graph). Simulations for the first interval $B \in [1, 10]$ were averaged over 100 random initial networks, while for the tail end we used 10 simulations in total. We find (as expected) that the MesoNMI decreases monotonically as a function of ϵ for all values of B, with $\epsilon = 0$ trivially providing no change to the similarity scores. We also find, as in Fig. 2, that for a fixed value of the attack fraction ϵ the MesoNMI values decrease monotonically as a function of B, indicating that the microscale edge attacks are not felt as intensely by the measures that aggregate structural information at larger scales.

In Fig. 3b, we plot the MesoNMI versus the mixing parameter μ used to generate the reference SBM graph, also at different attack fractions ϵ. In this case, the partition $\boldsymbol{b}$ used to compute the MesoNMI is the original planted partition of the reference network into two groups (simulations were averaged over 50 random initial networks). We find that, for a given attack fraction ϵ, the sensitivity in the MesoNMI is strongly dependent on the mixing level μ, with values decreasing as we move away from $\mu = 0.5$. This is because edge attacks will affect a highly (dis)assortative partition more than one with little mixing preference, since edges are more likely to be rewired to new group affiliations if they are highly concentrated among nodes with certain pairs of group affiliations. We also observe an interesting phase transition-like behavior in the MesoNMI at roughly $\mu = 0.25$ and $\mu = 0.75$, which may have qualitative similarities with the detectability transition of the SBM[48] in which community structure is suddenly statistically indistinguishable from random edge placement at a certain level of mixing.

These experiments altogether verify our intuition about how each measure values certain deviations in network structure, as well as confirm that our mesoscale mutual information measure is truly capturing variation at coarser scales than the NMI and DC-NMI measures. We supplement these tests with further experiments using other random graph ensembles in Supplementary Fig. 1.

FAO trade network case study

To examine how the proposed measures can be used to extract meaningful summaries of shared structure in sets of networks, we apply our measures to the multilayer FAO network of global trade patterns[7,49], where nodes represent $N = 214$ countries, an edge (i, j) represents the trade of a good between countries i and j, and each layer (of which there are 364) represents a different good being traded. We kept only the countries with gross domestic product (GDP) accessible through the World Bank indicators[50] of 2010 (the same year the tradings took place), and so the final network studied was reduced to 194 nodes. The network was binarized and converted to an undirected representation, facilitating a straightforward comparison with the network Jensen Shannon divergence (JSD) measure proposed in ref. 7 and evaluated on the same dataset. (However, as discussed in the "Mesoscale network mutual information" section, one can apply all of our measures to the original directed network representation, and apply the MesoNMI measures to the weighted representation by considering it as a multigraph). The network JSD measure of De Domenico et al.[7] is computed using the spectrum of the combinatorial Laplacian, and should capture both small- and large-scale structural similarities between networks as with other spectral measures[17,18].

We applied the NMI (Eq. (10)), DC-NMI (Eq. (16)), MesoNMI (Eq. (30)), and the network JSD[7]—which was converted to a similarity measure $1 - \text{JSD}$ to facilitate a direct comparison with our measures—to compare all pairs of layers in the FAO trade network and perform hierarchical clustering analysis as in ref. 7. For the MesoNMI, we used a number of partitions $\boldsymbol{b}$ reflecting meaningful divisions of the countries at multiple scales: (1) a partition into $B = 2$ groups representing the Global North and Global South in accordance to the United Nations' Finance Center for South-South Cooperation[51]; (2) a partition into $B = 6$ groups representing the continents in the dataset; (3) a partition into $B = 50$ equally sized groups of countries after ordering by GDP—this is to facilitate analysis at an intermediate network scale; and (4) a partition of the network into $B = 194$ groups each of size one—this is to compare and contrast the multigraph-based encoding of the MesoNMI with the graph-based encoding of the standard NMI, which can give substantially different results in practice. A diagram of our preprocessing and analysis for this section is shown in Supplementary Fig. 2.

In Fig. 4, we show the results of these experiments. In particular, Fig. 4a shows the pairwise similarity matrices for a few of the measures, which convey a similar block structure for the network layer comparisons. However, as the network scale of interest increases (NMI to DC-NMI, to MesoNMI to JSD) we can observe increasing similarity values, reflecting greater overall similarity among the networks when viewing them at larger scales. The MesoNMI measures find a much greater variability in the similarity values across network layers, as indicated by the range of shades present in the heatmap, while the NMI, DC-NMI, and JSD measures infer a tighter range of similarity scores among the layers.

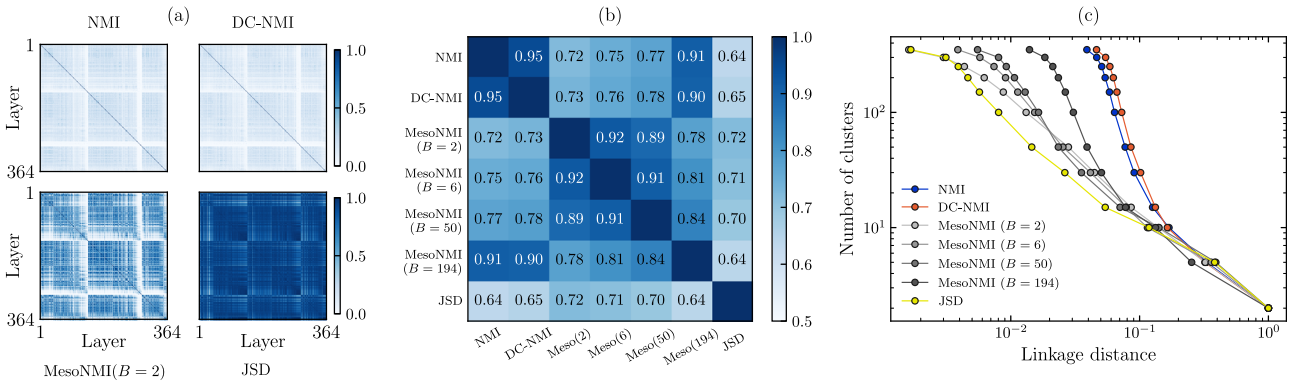

Fig. 4 | Comparison of graph similarity values among layers of the FAO trade network. a Pairwise similarity matrices among layers of the FAO trade network[7], each layer representing the global trade patterns among countries for a particular good. The MesoNMI was computed with respect to a partition of the country nodes in each layer according to a Global North–South dichotomy[51]. The network Jensen Shannon divergence (JSD) measure of De Domenico et al.[7] is transformed into a similarity measure using 1 − JSD and included for comparison. All matrices indicate a similar block structure to the layer similarities, and as the network scale of interest increases (NMI to DC-NMI, to MesoNMI to JSD) we find systematically higher similarity values, with the MesoNMI having the greatest discriminative power. **b** Rank-biased overlap (RBO)[52] between the pairwise distances calculated using each pair of similarity measures. For example, the (NMI, DC-NMI) entry of this matrix is the RBO between the entires of the top two panels in (**a**). As the scale of interest decreases, we find greater RBO between the corresponding pairwise distance matrices. **c** Number of clusters versus the corresponding Ward linkage distance for a hierarchical clustering of the layers[7]. There are discrepancies in the hierarchical cluster structure inferred using the measures, with measures operating at similar scales having similar linkage patterns.

We then compare these similarity matrices to examine how similar the ranks of the entries are across pairs of measures. We use the rank-biased overlap (RBO)[52], a measure indicating the overlap in the rankings of values from two lists. The RBO provides an alternative measure to the Spearman rank correlation that more heavily weights the contributions from the highest ranked items in both lists, which in our case allows us to more heavily favor pairs of measures that assign high similarity scores to common pairs of layers. In Fig. 4b, we plot the results of applying the RBO to the flattened similarity matrices of each pair of measures studied. We observe a very clear trend in which the RBO of a pair of measures is inversely related to the difference in their scale of interest. For example, the NMI and DC-NMI results have a decreasing RBO with the MesoNMI results as we move from $B = 194$ to $B = 50$, to $B = 6$ to $B = 2$, and the similarity among the MesoNMI results exhibits the same trend. Meanwhile, the highest RBO between any of our measures and the JSD is 0.72 (for MesoNMI($B = 2$)), which is in turn equal to the lowest of the RBO values among any of our measures (NMI and MesoNMI($B = 2$)). This may reflect the fact that the JSD is capturing both small- and large-scale similarities among the networks, while our measures are targeting particular scales of interest. In Supplementary Fig. 3a, we repeat the experiment in Fig. 4b using Spearman rank correlation instead of RBO, finding the same qualitative results.

We then examine the behavior of all measures in a hierarchical clustering context. We consider the rows of the similarity matrices in Fig. 4a as node embeddings that reflect the positions of each layer relative to the other layers. We then cluster the layers hierarchically using the Ward linkage criterion, as done in ref. 7 for the JSD distance measure. In this formulation, the pair of clusters whose layers have the least discrepancy in their similarity patterns with other clusters will be iteratively merged until there is only one remaining cluster of layers—the full multilayer network. We show the linkage results from this experiment in Fig. 4c. We can see that, as in Fig. 4b, the measures are ordered roughly by their scale of interest, with measures operating at similar scales having similar linkage patterns. Measures operating at small scales (e.g., NMI and DC-NMI) have a relatively high linkage threshold at which they begin to merge clusters, while measures operating at coarser scales (e.g., JSD and MesoNMI($B = 2$)) have a relatively low linkage threshold at which they begin to merge clusters. At large numbers of clusters, this pattern is reflective of the increasing similarity scores seen across scales in Fig. 4a. However, we see many of the linkage patterns for the coarser measures cross over those for the small-scale measures at roughly ten clusters. This indicates that these measures will give qualitatively distinct clustering patterns beyond those resulting simply from a systematic difference in the magnitude of their similarity scores.

To further examine the discrepancies in the clustering structure induced by each similarity measure, we test the extent to which the clusters obtained through hierarchical clustering reflect similarities in the products being traded in each layer. We assign a product type to each layer using the following 12 categories: {Proteins, Grains, Dairy, Fruits, Vegetables, Sweets, Drinks, Spices, Animals, Raw Materials, Tobacco, Other}[29]. We call these the "ground truth" product types for simplicity but emphasize that, since we are performing unsupervised learning, we are not trying to exactly capture the distinctions among product types with our method (a task for which supervised techniques are better suited).

We compare the clustering outputs from each measure with these ground truth categories using two different measures. First, we examine the extent to which each measure finds greater similarity among product layers within the same category versus other categories. To do this, we compute the ratio of the average similarity scores among layers within the same ground truth category ($\langle \mathrm{Sim}^{(\mathrm{with})} \rangle$) and layers within different ground truth categories ($\langle \mathrm{Sim}^{(\mathrm{betw})} \rangle$). A lower ratio indicates that the measure is better discriminating among the ground truth categories, since it finds higher similarities among within-category layers than between-category layers. The results for this experiment are shown in the left column of Table 1. We can see that the NMI and DC-NMI tend to best discriminate among the product categories, while the JSD is the poorest discriminator among the categories, with similarity scores that are nearly equal between and within categories. We also compute the adjusted mutual information (AMI)[22] between the partitions of the layers induced by each similarity measure with the ground truth layer partition according to product category. Since there are many strategies for determining where to cut each measure's corresponding clustering dendrogram to find its associated partition of the layers, we cut each measure's dendrogram at 12 clusters—the same as the number of clusters in the ground truth partition—to facilitate a fair comparison across measures. We plot the results in the right column of Table 1. We find that the NMI and DC-NMI find clusters that are the most similar to the ground truth according to the AMI, consistent with their capability to distinguish these categories in the previous experiment. Similarly, we find poor performance for the JSD and MesoNMI measures on this dataset. At face value, these results indicate that the similarity among product networks within the same category manifests at the microscale rather than the meso- or macro-scales. However, these results may vary depending on preprocessing choices for the networks (e.g., network pruning), which can affect the similarity values. In

Table 1 | Correspondence between similarity scores and product types of FAO trade network layers

	$\frac{\langle Sim^{(betw)}\rangle}{\langle Sim^{(with)}\rangle}$	$\frac{\sigma(Sim^{(betw)})}{\sigma(Sim^{(with)})}$	AMI
NMI	0.810	0.842	0.175
DC-NMI	0.805	0.817	0.223
MesoNMI ($B = 2$)	0.935	1.007	0.031
MesoNMI ($B = 6$)	0.929	0.998	0.055
MesoNMI ($B = 50$)	0.922	0.994	0.041
MesoNMI ($B = 194$)	0.864	0.924	0.138
JSD	0.982	1.100	0.040

The layers of the FAO trade network were each assigned to one of 12 "ground truth" product categories following the classification in ref. 29. The clusters for each method were compared with the ground truth categories using the ratio of the average between-category similarity and the average within-category similarity, with lower values indicating more tightly knit categories according to the similarity method of interest. The relative standard deviations of these similarity scores are also included for additional context. Then, the similarity matrices computed in Fig. 4 were clustered using Ward linkage as in ref. 7, with the number of clusters fixed at 12 to ensure a fair comparison across methods. The adjusted mutual information (AMI)[22] between the ground truth layer partition and the inferred layer partition was computed for each method. At face value, both tests indicate that the microscale measures are capturing more of the similarity among layers in the same category than the measures focusing on the meso- and macro-scales, suggesting that scattered individual edges may carry more information about shared global trade patterns than larger-scale network structure.

Supplementary Fig. 3b, we compute the AMI between the inferred clusters and the ground truth categories for all levels of each measure's hierarchical linkage dendrogram, finding roughly the same ordering of the measures as for the cut at 12 clusters. In Supplementary Fig. 4, we also plot the distribution of similarity scores within and between ground truth product groups to visualize the full distributions of scores rather than just the averages reported in Table 1. Finally, in Supplementary Fig. 5 we plot the distribution of ground truth categories within the inferred layer clusters for a few of the measures discussed for a more detailed picture.

In Supplementary Figs. 6 and 7, we apply our measures to two additional multilayer network datasets representing collaboration patterns among scientists within different fields of physics and routes for different airline companies, respectively[53].

Conclusion

In this paper, we have proposed a family of mutual information measures for computing the similarity between a pair of node-aligned networks. We adapt the encodings used to construct these measures to accommodate structural similarity at multiple scales within the network. By applying the proposed measures in a range of tasks, we find that the proposed measures are able to consistently capture meaningful notions of network similarity at the desired scale under perturbations to synthetic networks, as well as capture heterogeneity among the layers in real multilayer network datasets arising in the study of global trade, scientific collaboration, and transportation.

There are a number of ways in which our measures can be extended in future work. First, we can enable analyses in a broader range of contexts by lifting the restriction of the method to node-aligned graphs (at the potential cost of increased computational burden), refining our encoding for weighted graphs to enable meaningful analyses beyond the multigraph representation, and/or extending our pairwise encodings to populations of networks (e.g., as in ref. 29). The MesoNMI measure we propose does not consider the impact of degree heterogeneity on similarity, and so could also in principle be extended to a degree-corrected version as done with the standard graph NMI measure. One can additionally apply new encodings under our framework to capture similarity with respect to the presence of complex local structures such as motifs or other subgraphs while allowing for network of different sizes, similar to the calculations performed in ref. 27. Lastly, by adapting the combinatorial calculations to a new set of constraints the methods here presented can in principle be extended to more general discrete structures, such as hypergraphs or simplicial complexes, to account for higher-order interactions[54]. There are also many avenues of exploration to examine additional applications of our methods for downstream tasks such as network regression or anomaly detection and to a wider range of network datasets such as connectomics data[32].

Data availability

The data used in this paper are available at https://github.com/hfelippe/network-MI with the https://doi.org/10.5281/zenodo.13145602.

Code availability

Code implementing the measures presented in this paper is available at https://github.com/hfelippe/network-MI with the https://doi.org/10.5281/zenodo.13145602.

Received: 14 May 2024; Accepted: 4 October 2024;
Published online: 14 October 2024

References

1. Sharan, R. & Ideker, T. Modeling cellular machinery through biological network comparison. *Nat. Biotechnol.* **24**, 427 (2006).
2. Nikolova, N. & Jaworska, J. Approaches to measure chemical similarity—a review. *QSAR Comb. Sci.* **22**, 1006 (2003).
3. Wang, J. & Dong, Y. Measurement of text similarity: a survey. *Information* **11**, 421 (2020).
4. Zeng, Z., Tung, A. K., Wang, J., Feng, J. & Zhou, L. Comparing stars: on approximating graph edit distance. *Proc. VLDB Endow.* **2**, 25 (2009).
5. Guo, X., Hu, J., Chen, J., Deng, F. & Lam, T. L. Semantic histogram based graph matching for real-time multi-robot global localization in large scale environment. *IEEE Robot. Autom. Lett.* **6**, 8349 (2021).
6. Kriege, N. M., Johansson, F. D. & Morris, C. A survey on graph kernels. *Appl. Netw. Sci.* **5**, 1 (2020).
7. De Domenico, M., Nicosia, V., Arenas, A. & Latora, V. Structural reducibility of multilayer networks. *Nat. Commun.* **6**, 6864 (2015).
8. Ok, S. A graph similarity for deep learning. In. *Adv. Neural Inf. Process. Syst.* **33**, 1 (2020).
9. Attar, N. & Aliakbary, S. Classification of complex networks based on similarity of topological network features. *Chaos* **27**, 091102 (2017).
10. Vishwanathan, S. V. N., Schraudolph, N. N., Kondor, R. & Borgwardt, K. M. Graph kernels. *J. Mach. Learn. Res.* **11**, 1201 (2010).
11. Wills, P. & Meyer, F. G. Metrics for graph comparison: a practitioner's guide. *PLoS ONE* **15**, e0228728 (2020).
12. Hartle, H. et al. Network comparison and the within-ensemble graph distance. *Proc. R. Soc. A* **476**, 20190744 (2020).
13. Kyosev, I., Paun, I., Moshfeghi, Y. & Ntarmos, N. Measuring distances among graphs en route to graph clustering. In *2020 IEEE International Conference on Big Data (Big Data)*, 3632 (IEEE, 2020).
14. Ranshous, S. et al. Anomaly detection in dynamic networks: a survey. *Wiley Interdiscip. Rev.: Comput. Stat.* **7**, 223 (2015).
15. Roy, M., Schmid, S. & Tredan, G. Modeling and measuring graph similarity: the case for centrality distance. In *Proceedings of the 10th ACM International Workshop on Foundations of Mobile Computing*, 47 (ACM, 2014).
16. Soundarajan, S., Eliassi-Rad, T. & Gallagher, B. A guide to selecting a network similarity method. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, 1037 (SIAM, 2014).
17. Apolloni, N. N. W. B. An introduction to spectral distances in networks. In *Neural Nets WIRN10: Proc. 20th Ital. Workshop Neural Nets* 226, 227 (IOS Press, 2011).
18. Wilson, R. C. & Zhu, P. A study of graph spectra for comparing graphs and trees. *Pattern Recognit.* **41**, 2833 (2008).
19. Hammond, D. K., Vandergheynst, P. & Gribonval, R. Wavelets on graphs via spectral graph theory. *Appl. Comput. Harmonic Anal.* **30**, 129 (2011).
20. Newman, M. E. J., Cantwell, G. T. & Young, J.-G. Improved mutual information measure for clustering, classification, and community detection. *Phys. Rev. E* **101**, 042304 (2020).

21. McDaid, A. F., Greene, D. & Hurley, N. Normalized mutual information to evaluate overlapping community finding algorithms. Preprint at https://arxiv.org/abs/1110.2515 (2011).
22. Vinh, N. X., Epps, J. & Bailey, J. Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **11**, 2837 (2010).
23. Kirkley, A. Spatial regionalization based on optimal information compression. *Commun. Phys.* **5**, 249 (2022).
24. Kirkley, A. Inference of dynamic hypergraph representations in temporal interaction data. *Phys. Rev. E* **109**, 054306 (2024).
25. Fortunato, S. & Hric, D. Community detection in networks: a user guide. *Phys. Rep.* **659**, 1 (2016).
26. Gao, X., Xiao, B., Tao, D. & Li, X. A survey of graph edit distance. *Pattern Anal. Appl.* **13**, 113 (2010).
27. Coupette, C. and Vreeken, J. Graph similarity description: how are these graphs similar? In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 185 (ACM, 2021).
28. Escolano, F., Hancock, E. R., Lozano, M. A. & Curado, M. The mutual information between graphs. *Pattern Recognit. Lett.* **87**, 12 (2017).
29. Kirkley, A., Rojas, A., Rosvall, M. & Young, J.-G. Compressing network populations with modal networks reveal structural diversity. *Commun. Phys.* **6**, 148 (2023).
30. Young, J.-G., Kirkley, A. & Newman, M. E. J. Clustering of heterogeneous populations of networks. *Phys. Rev. E* **105**, 014312 (2022).
31. Lunagómez, S., Olhede, S. C. & Wolfe, P. J. Modeling network populations via graph distances. *J. Am. Stat. Assoc.* **116**, 2023 (2020).
32. Coupette, C., Dalleiger, S. & Vreeken, J. Differentially describing groups of graphs. *Proc. AAAI Conf. Artif. Intell.* **36**, 3959 (2022).
33. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (John Wiley & Sons, 2012).
34. Lancichinetti, A., Fortunato, S. & Radicchi, F. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* **78**, 046110 (2008).
35. Sporns, O. *Networks of the Brain* (MIT Press, 2010).
36. Eagle, N. & Pentland, A. Reality mining: sensing complex social systems. *Personal. Ubiquitous Comput.* **10**, 255 (2006).
37. Newman, M. E. Network structure from rich but noisy data. *Nat. Phys.* **14**, 542 (2018).
38. Meilă, M. Comparing clusterings by the variation of information. In *Learning Theory and Kernel Machines* (ed. Goos, G.) 173–187 (Springer, 2003).
39. Good, B. H., de Montjoye, Y.-A. & Clauset, A. Performance of modularity maximization in practical contexts. *Phys. Rev. E* **81**, 046106 (2010).
40. Peixoto, T. P. Bayesian stochastic blockmodeling. In *Advances in Network Clustering and Blockmodeling* (eds Doreian, P., Batagelj, V. & Ferligoj, A.) 289–332 (Wiley, New York, 2019).
41. Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: first steps. *Soc. Netw.* **5**, 109 (1983).
42. Peixoto, T. P. Entropy of stochastic blockmodel ensembles. *Phys. Rev. E* **85**, 056122 (2012).
43. Hein, J. L. *Discrete Mathematics* (Jones & Bartlett Learning, 2003).
44. Erdős, P. & Rényi, A. On random graphs. *Publ. Math.* **6**, 290 (1959).
45. Gilbert, E. N. Random graphs. *Ann. Math. Stat.* **30**, 1191 (1959).
46. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509 (1999).
47. Karrer, B. & Newman, M. E. J. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).
48. Decelle, A., Krzakala, F., Moore, C. & Zdeborová, L. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84**, 066106 (2011).
49. Tantardini, M., Ieva, F., Tajoli, L. & Piccardi, C. Comparing methods for comparing networks. *Sci. Rep.* **9**, 17557 (2019).
50. World Bank. World development indicators. https://databank.worldbank.org/source/world-development-indicators (2024).
51. Finance Center for South-South Cooperation, Global South Countries. http://www.fc-ssc.org/en/partnership_program/south_south_countries (2024).
52. Webber, W., Moffat, A. & Zobel, J. A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst.* **28**, 1 (2010).
53. Nicosia, V. & Latora, V. Measuring and modeling correlations in multiplex networks. *Phys. Rev. E* **92**, 032805 (2015).
54. Battiston, F. et al. Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1 (2020).

Acknowledgements

A.K. was supported by an HKU Urban Systems Fellowship Grant and the Hong Kong Research Grants Council under Grant no. ECS–27302523. F.B. acknowledges support from the Air Force Office of Scientific Research under award number FA8655-22-1-7025.

Author contributions

H.F.: conceptualization, software, validation, formal analysis, investigation, data curation, writing—original draft, and visualization; F.B.: conceptualization, writing—review and editing, supervision, and funding acquisition; A.K.: conceptualization, methodology, software, writing—original draft, writing—review and editing, and supervision. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests

Federico Battiston is an Editorial Board Member for *Communications Physics*, but was not involved in the editorial review of, or the decision to publish this article. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-024-01830-3.

Correspondence and requests for materials should be addressed to Federico Battiston or Alec Kirkley.

Peer review information *Communications Physics* thanks Brennan Klein, Corinna Coupette and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.