Multiscale mixing patterns in networks

Leto Peel,1,2,‡ Jean-Charles Delvenne,1,3,† and Renaud Lambiotte2,4,‡

1ICTEAM, Université catholique de Louvain, Louvain-la-Neuve, B-1348, Belgium
2naXys, Université de Namur, Namur, B-5000, Belgium
3CORE, Université catholique de Louvain, Louvain-la-Neuve, B-1348, Belgium
4Mathematical Institute, University of Oxford, Oxford, UK

Assortative mixing in networks is the tendency for nodes with the same attributes, or metadata, to link to each other. It is a property often found in social networks manifesting as a higher tendency of links occurring between people with the same age, race, or political belief. Quantifying the level of assortativity or disassortativity (the preference of linking to nodes with different attributes) can shed light on the factors involved in the formation of links and contagion processes in complex networks. It is common practice to measure the level of assortativity according to the assortativity coefficient, or modularity in the case of discrete-valued metadata. This global value is the average level of assortativity across the network and may not be a representative statistic when mixing patterns are heterogeneous. For example, a social network spanning the globe may exhibit local differences in mixing patterns as a consequence of differences in cultural norms. Here, we introduce an approach to localise this global measure so that we can describe the assortativity, across multiple scales, at the node level. Consequently we are able to capture and qualitatively evaluate the distribution of mixing patterns in the network. We find that for many real-world networks the distribution of assortativity is skewed, overdispersed and multimodal. Our method provides a clearer lens through which we can more closely examine mixing patterns in networks.

Networks are used as a common representation for a wide variety of complex systems, spanning social [13], biological [4, 5] and technological [6, 7] domains. Nodes are used to represent entities or components of the system and links between them are used to indicate pairwise interactions. The link formation processes in these systems are still largely unknown, but the broad variety of observed structures suggest that they are diverse. One approach to investigate the underlying generative processes is to examine the correlation, or assortative mixing, of node attributes (or “metadata”) across edges. This analysis allows us to make generalisations about whether we are more likely to observe links between nodes with the same characteristics (assortativity) or between those with different ones (disassortativity). Social networks frequently contain positive correlations of attribute values across connections [8]. These correlations occur as a result of the complimentary processes of selection (or “homophily”) and influence (or “contagion”) [9]. For example, assortativity has frequently been observed with respect to age, race and social status [10], as well as behavioural patterns such as smoking and drinking habits [11, 12]. Examples of disassortative networks include heterosexual dating networks (gender), ecological food webs (metabolic category), and technological and biological networks (node degree) [13].

The standard approach to quantifying the level of assortativity in a network is by calculating the assortativity coefficient [13]. Such a summary statistic is useful to capture the average mixing pattern across the whole network. However, such a generalisation is only really meaningful if it is representative of the population of nodes in the network. That is, if the assortativity of most individuals is concentrated around the mean, since we lose information about the diversity of mixing patterns present in the network. When networks are heterogeneous and contain diverse mixing patterns, the global measure may not provide an accurate description. Furthermore it does not provide a means for quantifying the diversity or identifying anomalous or outlier patterns of interaction.

Quantifying diversity and measuring how mixing may vary across a network becomes a particularly pertinent issue with modern advances in technology that have enabled us to capture, store and process massive-scale networks. Previously social interaction data was collected via time-consuming manual processes of conducting surveys or observations. For practical reasons these were often limited to a specific organisation or group [1, 2, 14, 15]. Summarising the pattern of assortative mixing as a single value may be reasonable for these small-scale networks that tend to focus on a single social dimension (e.g., a specific working environment or common interest). Now, technology such as online social media platforms allow for the automatic collection of increasingly larger amounts of social interaction data. For instance, the largest connected component of the Facebook network was previously reported to account for approximately 10% of the global population [16]. Within these vast multi-dimensional social networks the occurrence of heterogeneous patterns of mixing is more likely, perhaps due to differences in cultural backgrounds. Figure 1 shows, using the methods we will introduce, an example of this variation in mixing on a subset of nodes in the Facebook social network [17]. A high variation in mixing patterns also presents increased opportunity for
FIG. 1. Local assortativity of gender in a sample of Facebook friendships [17]. Different regions of the graph exhibit strikingly different patterns, suggesting that a single variable, e.g. global assortativity, would provide a poor description of the system.

the global assortativity to be a poor representation the entire population. To address this issue, we develop a node-centric measure of the assortativity within a local neighbourhood. Varying the size of the neighbourhood allows us to interpolate from the mixing pattern between an individual node and its neighbours to the global assortativity coefficient. In a number of real-world networks we find that the global assortativity is not representative of the collective patterns of mixing.

I. MIXING IN NETWORKS

Currently the standard approach to measure the propensity of links to occur between similar nodes is to use the assortativity coefficient introduced by Newman [13]. Here we will focus on undirected networks and discrete-valued node attributes, but assortativity and the methods we propose naturally extend to directed networks and scalar attributes (see SI).

The global assortativity coefficient \( r_{\text{global}} \) for discrete-valued attributes compares the proportion of links connecting nodes of with same attribute value, or type, relative to the proportion expected if the edges in the network were randomly rewired. The difference between these proportions is commonly known as modularity \( Q \), a measure frequently used in the task of community detection [18]. The assortativity coefficient is a normalised such that \( r_{\text{global}} = 1 \) if all edges only connect nodes of the same type (i.e., maximum modularity \( Q_{\text{max}} \)) and \( r_{\text{global}} = 0 \) if the number of edges is equal to the expected number for a randomly rewired network in which the total number of edges incident on each type of node is held constant. Mathematically, we calculate the global assortativity \( r_{\text{global}} \) by [13]

\[
r_{\text{global}} = \frac{Q}{Q_{\text{max}}} = \frac{\sum g e_{gh} - \sum g a_{gh}^2}{1 - \sum g a_{gh}^2},
\]

where \( e_{gh} \) is the proportion of edges in the network that connect vertices with type \( y_i = g \) to vertices with type \( y_j = h \):

\[
e_{gh} = \frac{1}{2m} \sum_{ij} A_{ij} \delta_{y_i,g} \delta_{y_j,h},
\]

and \( a_g = \sum_h e_{gh} \) represents the total proportion of edges incident on nodes of type \( g \). The normalisation constant \( Q_{\text{max}} = 1 - \sum_g a_{gh}^2 \) ensures that the assortativity coefficient lies in the range \(-1 \leq r \leq 1\) and that the maximum value of 1 is attainable.

A. Local patterns of mixing

The global assortativity is a summary statistic that describes the pattern of mixing on average across the whole network. But as with all summary statistics there may be cases where it provides a poor representation of the network as a whole, for instance, if there are localised heterogeneous patterns of mixing across the network. Figure 2 illustrates an analogy to Anscombe’s quartet of bivariate datasets with identical correlation coefficients [19]. Each of the five networks in the top row have the same number of nodes \((n = 40)\) and edges \((m = 160)\) and have been constructed to have the same \( r_{\text{global}} \) with respect to a binary attribute, indicated by a cross \((c)\) or a diamond \((d)\). All five networks have \( m_{cc} + m_{dd} = 80 \) edges between nodes of the same type and \( m_{cd} = 80 \) edges between nodes of different types, such that each has \( r_{\text{global}} = 0 \). Local patterns of mixing are formed by splitting each of the types \( \{c, d\} \) further into two equally sized subgroups \( \{c_1, c_2, d_1, d_2\} \). The middle row depicts the placement of edges within and between the four subgroups. Distributing edges uniformly between subgroups creates a network with homogeneous mixing [Fig. 2(a)].

We propose a local measure of assortativity \( r(\ell) \) that captures the mixing pattern within the local neighbourhood of a given node \( \ell \). Trivially one could calculate the local assortativity by adjusting [1] to only consider the immediate neighbours of \( \ell \). However, taking this approach can encounter problems. For nodes with low degree, we would be calculating assortativity based only on a small sample, providing a potentially poor estimate of the node’s mixing preference. And when all of \( \ell \)’s neighbours are of the same type, then we would assign \( r(\ell) = \infty \) because \( 1 - \sum a_{gh}^2 = 0 \).

We face similar issues in time-series analysis when we wish to interpret how a signal varies over time. If the observations are subject to random noise then direct analysis of the series may be more descriptive of the noise process than the underlying signal we are interested in.
FIG. 2. Five networks (top) of $n = 40$ nodes and $m = 160$ edges with the same global assortativity $r_{\text{global}} = 0$, but with different local mixing patterns as shown by the distributions of $r_{\text{multi}}$ (bottom).

However, averaging over the whole series may provide the most accurate estimate of the mean, but treats all variation as noise and misses any important trends. A common solution to this problem is to use a local filter such as the exponential weighted moving average. In this way, values further in time from the point of interest are weighted less. We adopt a similar strategy in calculating the local assortativity. To make the connection with time-series analysis concrete, consider that a random walk on a graph can be represented as a time-series of values that correspond to the attribute values of the visited nodes in the walk. For such a time-series we can calculate $e_{gh}$ used in (1) as,

$$e_{gh} = \sum_{ij} \pi_i \frac{A_{ij}}{k_i} \delta_{y_i,g} \delta_{y_j,h} ,$$  

(3)

where $\pi_i$ is the probability of being at node $i$ and is obtained from the stationary distribution of the random walk. For an undirected network and a simple random walk, in which outgoing edges are selected with equal probability, $\pi_i = k_i/2m$ is proportional to node degree and (3) reduces to (2). From this perspective, assortativity is the autocorrelation (with a time lag of 1) of the time-series of node attribute values.

Global assortativity counts all edges in the network equally. To create our local measure of assortativity we instead reweight the edges in the network based on how local they are to $\ell$. We do so by replacing the stationary distribution $\pi$ in (3) with an alternative distribution over the nodes $w(i; \ell)$.

$$e_{gh}(\ell) = \sum_{ij} w(i; \ell) \frac{A_{ij}}{k_i} \delta_{y_i,g} \delta_{y_j,h} ,$$  

(4)

and compare the proportion of links between nodes of the same type in the local neighbourhood to the global value $\nu(\ell) = \sum_g (e_{gg}(\ell) - e_{gg})$. Then we can calculate the local assortativity as the deviation from the global assortativity:

$$r(\ell) = \frac{1}{Q_{\text{max}}} \left( \nu(\ell) + \sum_g (e_{gg}(\ell) - a_g^2) \right)$$  

(5)

$$= \frac{1}{Q_{\text{max}}} \sum_g (e_{gg}(\ell) - a_g^2)$$  

(6)

All that remains is to define a distribution $w(i; \ell)$. We continue from the perspective of measuring the autocorrelation of values encountered during a random walk on the graph. However, this time we make the modification that at each time step with probability $(1 - \alpha)$ we return to the node of interest $\ell$ [Fig. 3(a)]. We use the stationary distribution of this random walk $w_{\alpha}(i; \ell)$ (also known as the personalised PageRank vector) in (4). In the special case of a network consisting of nodes linked in a line, $w_{\alpha}(i; \ell)$ corresponds to an exponential distribution [Fig. 3(b)] and is analogous to the previously mentioned exponential filter commonly used in time-series analysis.

This means that we can calculate a local assortativity $r_{\alpha}(\ell)$ for each node and use $\alpha$ to interpolate from the
FIG. 3. Example of the local assortativity measure for discrete attributes (a) assortativity is calculated (as in [1]) according to the actual proportion of links in the network connecting nodes of the same type relative to the expected proportion of links between nodes of the same type, (b) the nodes in the network are weighted according to a random walk with restart probability of $1 - \alpha$, (c) an example of the local assortativity applied to a simple line network with two types of nodes: yellow or green. The blue bars show the stable distribution ($w(i; \ell)$) of the random walk with restarts at $\ell$ for different values of $\alpha$. Underneath each distribution the nodes in the line network are coloured according to their local assortativity value.

trivial local neighbourhood assortativity ($\alpha = 0$, the random walker never leaves the initial node) to the global assortativity ($\alpha = 1$, the random walker never restarts) $r_1(\ell) = r_{\text{global}}$ [Fig. 3(c)].

B. Choice of $\alpha$

We can use $\alpha$ to interpolate from the global measure at $\alpha = 1$ to the local measure based only on the neighbours of $\ell$ when $\alpha = 0$. As previously mentioned, either extreme can be problematic, $r_1$ is uniform across network, while $r_0$ may be based on a small sample (particularly in the case of low degree nodes) and therefore subject to overfitting. Moreover, both extremes are blind to the possible existence of coherent regions of assortativity inside the network, as $r_1$ considers the network as a whole while $r_0$ considers the local assortativities of the nodes as independent entities.

To circumvent these issues, we consider calculating the assortativity across multiple scales by calculating a “multiscale” distribution $w_{\text{multi}}$ by integrating over all possible values of $\alpha$ [20],

$$w_{\text{multi}}(i; \ell) = \int_0^1 w_\alpha(i; \ell) \, d\alpha , \quad (7)$$

which is effectively the same as treating $\alpha$ as an unknown with a uniform prior distribution (see SI for details). Using this distribution, we can calculate a multiscale measure $r_{\text{multi}}$ that captures the assortativity of a given node across all scales.

As a simple demonstration, we return to Figure 2 in which the distribution of $r_{\text{multi}}$ for each synthetic network is shown in the bottom row. We see under homogeneous mixing [Fig. 2(a)] a unimodal distribution peaked around 0 confirming that the global measure $r_{\text{global}}$ is representative of the mixing patterns in the network. However, when mixing is heterogeneous [Fig. 2(b–e)] we observe multimodal distributions of $r_{\text{multi}}$ that allow us to disambiguate between different local mixing patterns.

II. REAL NETWORKS

Next we use $r_{\text{multi}}$ to evaluate the mixing patterns in some real networks: an ecological network and set of online social networks. In both cases nodes have multiple attributes assigned to them, providing different dimensions of analysis.

A. Weddell Sea Food Web

We first examine a network of ecological consumer interactions between species dwelling in the Weddell Sea [4]. Figure 4 shows the distributions (green) of local assortativity for five different discrete-valued node attributes. For comparison we present a null distribution (black) obtained by randomly re-wiring the edges such that the attribute values, degree sequence and global assortativity are all preserved (see SI). In each case we observe skewed and/or multi-modal distributions. The empirical distributions appear overdispersed compared to the null distributions.

It may be surprising to see that, for some attributes, the null distribution appears to be multi-modal. Closer inspection reveals that the different modes are correlated with the attribute values and that multi-modality arises from the unbalanced distribution of nodes and incident edges across different node types. This effect is particularly pronounced for the attribute “Metabolic Category” for which we observe two distinct peaks in the distribution. The larger peak that occurs around $r_{\text{multi}} \sim 0$
represents all species that belong to the metabolic category plant and accounts for the majority (348/492) of the species in the network, upon which approximately two thirds of the edges are incident. This bias in the distribution of edges across the different node types means that randomly assigned edges are more likely to connect two nodes of the majority class than any other pair of nodes. In fact, it is impossible to assign edges such that nodes in each Metabolic Category exhibit (approximately) the same assortativity as the global value. Specifically, to achieve \( r_{\text{global}} = -0.13 \) it is necessary that more than half of the edges connect species from different metabolic categories. However, this is impossible for the plant category without changing the distribution of edges over categories.

B. Facebook 100

We next consider a set of online social networks collected from the Facebook social media platform at a time when it was only open to 100 US universities [3]. The process of incrementally providing these universities access to the platform, meant that at this point in time very few links existed between each of the universities’ networks, which provides the opportunities to study each of these social systems in a relatively independent manner. One of the original studies on this dataset examined the assortativity of each demographic attribute in each of the networks [3]. This study found some common patterns that occurred in many of the networks, such as a tendency to be assortative by matriculation year and dormitory of residence, with some variation around the magnitude of assortativity for each of the attributes across the different universities.

In this case it makes sense to analyse the universities separately since the university membership played an important and restricting role in the formation of the network. However, a modern version of this dataset might contain a higher density of inter-university links, making it less reasonable to treat them independently; in general partitioning networks based on attributes without careful considerations can be problematic [21].

Figure 5 depicts the distributions of \( r_{\text{multi}} \) for each of the 100 networks according to dormitory. For many of the networks we observe a positively-skewed distribution. The surrounding subplots show details for four universities with approximately the same global assortativity \( r_{\text{global}} \sim 0.13 \), but with qualitatively different distributions of \( r_{\text{multi}} \). Common across these distributions is that the peaks match those of the null distribution and all of the empirical distributions exhibit a positive skew beyond that of the null distribution. Closer inspection reveals that in all four networks the nodes associated with a higher local assortativity belong to a community of nodes more loosely connected to the rest of the network. These nodes also correspond to students in their first year, which suggests that residence plays a more important role in link formation among new students, than it does for students that have attended the university for longer.

Our local measure of assortativity also provides the means to compare the assortativity of multiple attributes and how they covary across the network. Figure 6 compares the local assortativities year of study and place of residence. The central scatter plot shows for each university the correlation of local assortativity for the two attributes against the proportion of nodes which are more assortative by residence than by year of study. The four surrounding sub-plots show the joint distribution of year and dorm local assortativity for specific universities. The yellow points indicate students in their first year. In most universities we observed that first-year students were the most assortative by either year, residence or both. In both Auburn and Pepperdine there is a negative correlation between year and dorm assortativity suggesting that many friendships have formed either through being in the same year or from sharing a dorm.

For Simmons and Rice we observe a positive correlation between the two dorm and year local assortativity. However, in Simmons we see that the first-year students form a separate cluster, while in Rice they are much more interspersed. This difference may relate to how students are placed in university dorms. At Simmons all first year students live on campus and form the majority residents in the few dorms they occupy. Rice houses their new

1 source: [http://www.simmons.edu/student-life/life-at-simmons/housing/residence-halls](http://www.simmons.edu/student-life/life-at-simmons/housing/residence-halls)
FIG. 5. Distributions of the local assortativity by residence (dorm) for each of the schools in the Facebook 100 dataset. Dotted black lines indicate the 10 and 90 percentiles while the solid black lines show the interquartile range. The global assortativity is indicated by the blue square markers. The distributions for four schools (Dartmouth, Wesleyan, Wellesley and Haverford) are shown in detail in the surrounding. Each of them has approximately the same global assortativity ($r_{\text{global}} \sim 0.13$), but the distributions indicate different levels of heterogeneity in the pattern of mixing by residence. While the distributions are different, there exists a common trend that the first year students tend to be more loosely connected to the rest of the network and exhibit the higher values of assortativity (nodes to the right of the dashed cyan line).

FIG. 6. A scatter plot in which each school is a point indicating the correlation of local assortativities by dorm and matriculation year ($x$-axis) and proportion of nodes which are more assortative by dorm than by year ($y$-axis). Joint distributions of the dorm and year assortativities for four of the schools are shown in the surrounding plots.

takes according to a different strategy, by placing them evenly spread across all the available dorms. The fact that students are mixed across years and that the vast majority (almost 78\%\footnote{source: http://campushousing.rice.edu/}) of students reside in university accommodation, offers a possible explanation for why we observe a smooth variation in values of assortativity without a distinction between new students and the rest of the population.
III. DISCUSSION

Characterising the level of assortativity plays an important role in understanding the processes of link formation in complex systems. However, as we have shown here, the global assortativity may not be representative of the variation in mixing patterns across the network as a whole. We have seen that the distribution of mixing in real networks are often skewed, overdispersed and possibly multimodal. In fact, for certain network configurations we have seen that a unimodal distribution may not even be possible.

As network data grow bigger there is a greater possibility for heterogeneous sub-groups to co-exist within the overall population. The presence of these sub-populations adds further to the ongoing discussions of the interplay between node metadata and network structure [21] and suggests that while we may observe a relationship between particular node properties and link formation in part of a network, it does not imply that this relationship exists across the network as a whole. A promising direction for future work might be to identify these sub-groups as a new perspective in the community detection problem [22].

In our heavily data-centric society, domains such as advertising, credit, education, employment and many others are all set to hugely benefit. It is important that we are mindful of the fundamental assumptions we make when analysing these data. One of the most common assumptions made in data analysis is that the observed data are independently and identically distributed; the so-called i.i.d. assumption. In network science the independence assumption is explicitly violated by the network structure and presence of (dis)assortative mixing. Here we have considered a second potential violation of this assumption, where data may not be identically distributed but instead vary over the graph structure.

This heterogeneity not only challenges traditional assumptions of statistics and machine learning, but also offers potential new opportunities as well. Recent results studying prediction tasks on networks with heterogenous mixing patterns demonstrate that with an appropriately constructed learning algorithm it is still possible to make accurate predictions [23] and in some cases even utilise the heterogeneity to further improve performance [24].

Here we have considered the localisation of a global network measure using a network analogy of the exponentially weighted mean frequently used in time series analysis. Our focus has been on local patterns of assortativity, but the same methodology could easily be applied to any global network measure, such as clustering coefficient or mean degree, or to capture the local distribution of node properties.

ACKNOWLEDGEMENTS

We thank Michael Schaub for insightful conversations. This work was supported by IAP (Belgian Scientific Policy Office) [LP, JCD, RL], ARC (Federation Wallonia-Brussels) [LP, JCD, RL], and the MOVE-IN Louvain programme supported by the European Commission (COFUND Marie Curie actions) [LP].

[1] David Krackhardt, “The ties that torture: Simmelian tie analysis in organizations.” Research in the Sociology of Organizations 16, 183–210 (1999).
[2] Emmanuel Lazega, The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership (Oxford University Press on Demand, 2001).
[3] Amanda L Traud, Peter J Mucha, and Mason A Porter, “Social structure of facebook networks,” Physica A: Statistical Mechanics and its Applications 391, 4165–4180 (2012).
[4] Ulrich Brose et al., “Body sizes of consumers and their resources,” Ecology 86, 2545 (2005).
[5] H Jeong, B Tombor, R Albert, ZN Oltvai, and AL Barabasi, “The large-scale organization of metabolic networks,” Nature 407, 651 (2000).
[6] Réka Albert, Hawoong Jeong, and Albert-László Barabási, “Internet: Diameter of the world-wide web,” Nature 401, 130 (1999).
[7] Duncan J Watts and Steven H Strogatz, “Collective dynamics of‘small-world‘networks,” nature 393, 440 (1998).
[8] Miller McPherson, Lynn Smith-Lovin, and James M Cook, “Birds of a feather: Homophily in social networks,” Annual review of sociology 27, 415–444 (2001).
[9] Sinan Aral, Lev Muchnik, and Arun Sundararajan, “Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks.” Proceedings of the National Academy of Sciences 106, 21544–21549 (2009).
[10] James Moody, “Race, school integration, and friendship segregation in america,” American Journal of Sociology 107, 679–716 (2001).
[11] Jere M Cohen, “Sources of peer group homogeneity,” Sociology of education 227–241 (1977).
[12] Denise B Kandel, “Homophily, selection, and socialization in adolescent friendships,” American journal of Sociology 84, 427–436 (1978).
[13] Mark EJ Newman, “Mixing patterns in networks,” Physical Review E 67, 026126 (2003).
[14] Samuel F Sampson, A novitiate in a period of change: An experimental and case study of social relationships, Ph.D. thesis, Cornell University Ithaca, NY (1968).
[15] Wayne W Zachary, “An information flow model for conflict and fission in small groups,” J. Anthropol. Res. , 452–473 (1977).
[16] Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow, “The anatomy of the facebook social graph,” arXiv preprint arXiv:1111.4503 (2011).
[17] Julian J McAuley and Jure Leskovec, “Learning to discover social circles in ego networks,” in Adv. in Neural
Appendix A: Directed networks

We can easily extend the multiscale mixing measure $r_\alpha$ described in the main text to directed networks. The main change is to incorporate two sets of marginals $a$ and $b$ that describe the proportion of edges starting from and ending at each of the attribute types. Then the directed global assortativity of a network with respect to a particular discrete node attribute $y_i$ is:

$$r_{\text{global}} = \frac{\sum_g e_{g}}{1 - \sum_g a_g b_g} ,$$

where $a_g$ and $b_g$ represent the total number of outgoing and incoming links of all nodes of type $g$:

$$a_g = \sum_h e_{gh} , \quad b_g = \sum_g e_{gh} .$$

Then we can update our definition of local assortativity accordingly,

$$r(\ell) = \frac{1}{Q_{\text{max}}} \sum_g (e_{gg}(\ell) - a_g b_g) .$$

Appendix B: Scalar attributes

For scalar attributes we can simply calculate the Pearson’s correlation across edges. Using $x_i$ and $x_j$ to indicate the scalar attribute value of the sender and receiver nodes in edge $A_{ij}$ then we can write the global assortativity as,

$$r_{\text{global}} = \frac{\text{cov}(x_i, x_j)}{\sigma_i \sigma_j} = \frac{\sum_{ij} A_{ij} (x_i - \bar{x}_i) (x_j - \bar{x}_j)}{\sqrt{\sum_i k_i (x_i - \bar{x}_i)^2} \sqrt{\sum_j k_j (x_j - \bar{x}_j)^2}} ,$$

where $\bar{x}_i$ and $\sigma_i$ represent the mean and standard deviation of the attribute values of the sender nodes. In the case of an undirected network, then:

$$r_{\text{global}} = \frac{\sum_{ij} A_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{\sum_i k_i (x_i - \bar{x})^2} ,$$

where $\bar{x} = \frac{1}{2m} \sum_i k_i x_i$ is the mean value of $x$ weighted by node degree $k$. If we standardise the scalar values using the linear transformation $\tilde{x}_i = \frac{x_i - \bar{x}}{\sigma_i}$, then we can simplify this further as,

$$r_{\text{global}} = \sum_{ij} \frac{A_{ij} \tilde{x}_i \tilde{x}_j}{2m} .$$

Then we can calculate the local assortativity $r_\alpha(\ell)$ for scalar variables as,

$$r_\alpha(\ell) = \sum_{ij} w_\alpha(i; \ell) \frac{A_{ij} \tilde{x}_i \tilde{x}_j}{k_i} .$$

Appendix C: Discrete assortativity as a correlation

The assortativity coefficient $r_{\text{global}}$ for discrete attributes can be interpreted as a normalised Pearson’s correlation. To see this, we start by observing that the Pearson’s correlation of two binary variables is equivalent to the Phi coefficient for binary contingency tables [25].
Table I shows a contingency table using the same notation as the directed assortativity, i.e., \( a \) and \( b \) give the marginal proportions and \( e \) gives the joint proportions. Then the Pearson product-moment correlation of these variables is known as \( \phi \), given by:

\[
\phi = \frac{e_{11} - a_1 b_1}{\sqrt{a_1 b_1 a_2 b_2}} \quad (C1)
\]

Note that it is only necessary to calculate this in terms of \( e_{11} \), since \( e_{11} - a_1 b_1 = e_{22} - a_2 b_2 \). We can see this using the identity \( e_{22} = b_2 - a_1 + e_{11} \):

\[
e_{22} - a_2 b_2 = b_2 - a_1 + e_{11} - (1 - a_1)(1 - b_1) \quad (C2)
\]

\[
e_{22} - a_2 b_2 = (1 - b_1) - a_1 + e_{11} - (1 - a_1 - b_1 + a_1 b_1) \quad (C3)
\]

\[
e_{22} - a_2 b_2 = e_{11} - a_1 b_1 \quad (C4)
\]

A well-known issue with \( \phi \) is that the extreme values of \( +1 \) and \( -1 \) are typically unobtainable, which can cause issues with its interpretation. In fact \( \phi = 1 \) can only occur if \( a_1 = b_1 \), e.g., when the network is undirected, while \( \phi = -1 \) can only occur if \( a_1 = b_2 = 0.5 \) [26, 27]. To address this issue, there have been a number of proposed normalisations to ensure the \( \phi = 1 \) is obtainable [28]. One such normalisation is the \( \phi/\phi_{\text{max}} \) proposed by Cureton [29],

\[
\frac{\phi}{\phi_{\text{max}}} = \frac{e_{11} - a_1 b_1}{\beta - a_1 b_1} \quad (C5)
\]

where \( \beta \) is the maximum possible value that \( e_{11} \) can take, i.e., \( \min(a_1 b_1) \). Note for undirected networks

\[
\sqrt{a_1 b_1 a_2 b_2} = \sqrt{a_1^2 a_2^2}
\]

\[
= a_1 a_2 \quad (C6)
\]

\[
= a_1 (1 - a_1) \quad (C7)
\]

\[
= a_1 - a_1^2 \quad (C8)
\]

which equals \( \phi_{\text{max}} \) when \( a_1 \leq a_2 \).

Then we can generalise \( \phi/\phi_{\text{max}} \) from binary to multi-category variables by treating each distinct value as a binary variable and taking their sum. If we set \( \beta = 1 \), then we obtain [A1] and thus we recover Newman’s assortativity [13]. We also note that [A1] also corresponds to Cohen’s \( \kappa \) that is frequently used to assess inter-rater agreement [40].

The normalisation of the assortativity coefficient means that \( r_{\text{min}} \leq r \leq 1 \) and

\[
r_{\text{min}} = -\frac{\sum_g a_g b_g}{1 - \sum_g a_g b_g} \quad , (C10)
\]

which lies in the range \(-1 \leq r_{\text{min}} < 0\).
when normalising to ensure that $\sum_{gh} e_{gh}(\ell) = 1$, a necessary step in calculating the assortativity. Then when we examine the distribution of $r(\ell)$ across all nodes in the network, the resulting distribution will be biased representation. To deal with this issue we calculate each of the local assortativities as normal, but assign each a weight $z_\ell = \sum_{gh} e_{gh}(\ell)$, i.e., the sum of local edge counts before normalisation. The weight $z_\ell$ describes our confidence in the local assortativity estimate from $z_\ell = 0$, indicating no confidence, to $z_\ell = 1$ when all node attributes within the neighbourhood are known. We adjust for these weights when plotting the histograms in the main text.

**Appendix E: Calculating the personalised PageRank vector**

The personalised PageRank vector is the stationary distribution of a random walk with restarts. We calculate it by direct simulation of the random walk process using the power method:

$$w_\alpha(i; \ell)_{s+1} = \alpha \sum_j \frac{A_{ij}}{k_i} w_\alpha(j; \ell)_s + (1 - \alpha) \delta_{i, \ell} , \quad (E1)$$

and at convergence yields a distribution $w(i; \ell)$ with a mode at $\ell$.

**Appendix F: Integrating over $\alpha$**

To integrate over all values of $\alpha$, we take advantage of the fact that we can equivalently write the $\eta$-th approximation the power method in [31] as the $\eta$-th degree truncation of the power series [31]:

$$w_\alpha(i; \ell)_\eta = \delta_{i, \ell} + \sum_{s=1}^\eta \alpha^s \left[ \left( \frac{A_{i\ell}}{k_i} \right)^s - \left( \frac{A_{i\ell}}{k_i} \right)^{(s-1)} \right]. \quad (F1)$$

By taking advantage of the relationship between $\alpha$ and the sequence of approximations computed by the power method, we can calculate the distribution $w_\alpha(i; \ell)$ for a given $\alpha = \alpha_0$ and use the sequence of approximations to calculate the distribution for any other $\alpha$ [20]:

$$w_\alpha(i; \ell)_\eta = \delta_{i, \ell} + \sum_{s=1}^\eta \frac{\alpha^s}{\alpha_0^s} \left( w(i; \ell, \alpha_0)_s - w(i; \ell, \alpha_0)_{s-1} \right). \quad (F2)$$

We can then integrate over all possible values of $\alpha$ [20],

$$w_{\text{multi}}(i; \ell)_\eta = \int_0^1 w_\alpha(i; \ell)_\eta \, d\alpha \quad (F3)$$

$$= \delta_{i, \ell} + \sum_{s=1}^\eta \frac{\left( w_{\alpha_0}(i; \ell)_s - w_{\alpha_0}(i; \ell)_{s-1} \right)}{(s+1)\alpha_0^s}. \quad (F4)$$

**Appendix G: Null model network generation**

We created a null model to generate networks with the same global assortativity as the observed network to compare the distributions of $r_{\text{multi}}$. For a fair comparison, we decided to keep the node degree and metadata label fixed while randomly rewiring the network. We do so using a modified version of the Markov chain Monte Carlo (MCMC) sampling of the configuration model for stub-labelled simple graphs [32]. The modification is to ensure that we sample a graph with (approximately) the same global assortativity as the observed network. We achieve this by adding a rejection sampling step based on the binomial likelihood of observing the number of edges between nodes of the same type $m_{\text{in}} = m \sum_g e_{gg}$ given the proportion of edges required to maintain the global assortativity $\omega_{\text{in}} = \sum_g e_{gg}$:

$$L(G_i) = \log \left( \frac{m}{m_{\text{in}}} \right)^{(\omega_{\text{in}} - 1)} \left( 1 - \omega_{\text{in}} \right)^{m - m_{\text{in}}}. \quad (G1)$$

The modified MCMC algorithm is shown in Algorithm 1.

**Algorithm 1** stub-labeled MCMC

Require: initial simple graph $G_0$, initial temp $t_0$

Ensure: sequence of graphs $G_i$

for $i < \text{number of graphs to sample}$ do

choose two edges at random

randomly choose one of the two possible swaps

if edge swap would create a self-loop or multiedge then

resample current graph: $G_i \leftarrow G_{i-1}$

else if $\text{Unif}(0,1) < \exp \left( \frac{L(G_i) - L(G_{i-1})}{t_i} \right)$ then

swap the chosen edges, producing $G_i$

else

reject $G_i$

end if

$t_{i+1} \leftarrow \text{update}(t_i)$

end for

**Appendix H: Datasets**

1. Weddell Sea Food Web

The food web of the Antarctic Weddell Sea [4] consists of 488 species and 15885 consumer relations. For each of the nodes in this network we have five discrete-valued attributes: Metabolic Category \{Plant, Ectotherm vertebrate, Endotherm vertebrate, Invertebrate\}, Feeding Type \{Carnivorous/necrovorous, Herbivorous/detrivorous, Detrivorous, Omnivorous, Primary\}. For simple graphs sampling from the space of stub-labelled graphs is equivalent to sampling from the space of vertex-labelled graphs [32].
producer, Carnivorous}, FeedingMode \{Pelagic predator, Predator/scavenger, Primary producer, Predator, Deposit-feeder, Grazer, Suspension-feeder\}, Mobility \{1,2,3,4\}, Environment \{Bathydemersal, Land-based, Resource, Pelagic, Benthopelagic, Benthic, Demersal\}. For scalar attributes we use the mean mass of the species, mobility (although discrete, the values are ordinal), and node degree.

2. **Facebook 100**

The Facebook100 dataset \cite{facebook100} contains an anonymised snapshot of the friendship connections among 1,208,316 users affiliated with the first 100 colleges admitted to Facebook. The dataset contains a total of 93,969,074 friendship edges between users of the same college. Each node has a set of discrete-valued social variables: status \{undergraduate, graduate student, summer student, faculty, staff, alumni\}, dorm, major, gender \{male, female\}, and graduation year.