Introducing Graph Cumulants:
What is the Kurtosis of your Social Network?

Lee M. Gunderson[*,1,†] & Gecia Bravo-Hermsdorff[*,2,†]

[*]*Both authors contributed equally to this work*

[1]Department of Astrophysical Sciences, Princeton University, Princeton, NJ, 08544, USA
[2]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, 08544, USA

[†]Correspondence to: geciah@princeton.edu or leeg@princeton.edu

In an increasingly interconnected world, understanding and summarizing the structure of these networks becomes increasingly relevant. However, this task is nontrivial; proposed summary statistics are as diverse as the networks they describe, and a standardized hierarchy has not yet been established. In contrast, vector-valued random variables admit such a description in terms of their cumulants (e.g., mean, (co)variance, skew, kurtosis). Here, we introduce the natural analogue of cumulants for networks, building a hierarchical description based on correlations between an increasing number of connections, seamlessly incorporating additional information, such as directed edges, node attributes, and edge weights. These graph cumulants provide a principled and unifying framework for quantifying the propensity of a network to display any substructure of interest (such as cliques to measure clustering). Moreover, they give rise to a natural hierarchical family of maximum entropy models for networks (i.e., ERGMs) that do not suffer from the "degeneracy problem", a common practical pitfall of other ERGMs.

1

The power of natural science relies on the ability to find abstract representations of complex systems and describe them with meaningful summary statistics. For example, using the pluripotent language of graph theory, the field of network science distills a variety of systems into the entities comprising them (the nodes) and their pairwise connections (the edges). Remarkably many systems, from a wide variety of fields, naturally benefit from such a description, such as electrical circuits (*1*), brains (*2*), food webs (*3*), friendships (*4*), transportation networks (*5*), and the internet (*6*).

As the field developed, real networks were noticed to display several recurring themes (*7, 8, 9*), such as: sparsity (a small fraction of all potential connections actually exist), heterogeneous degree distributions (some nodes have many more connections than average), and high clustering (nodes form tightly connected groups). As a consequence, measures are often tailored to capture such properties (*10, 11, 12*) and network models are often chosen to mimic them (*13, 14, 15, 16, 17, 18*). However, network statistics are often intertwined (*19, 20*), and models with notably different mechanisms (*16, 17, 18, 21, 22, 23*) can reproduce the same commonly observed properties. It is not currently clear how to compare different methods within a single principled framework; a standardized hierarchy of descriptive network statistics and associated models is needed.

Real networks often contain a hierarchy of interconnected scales (*24*). This paper formalizes a novel bottom-up approach, where the elementary units are the edges (as compared to the nodes (*20*)), and the hierarchy is built by considering correlations between an increasing number of them. Relationships between several edges can be expressed as subgraphs. Such substructures are often referred to as network motifs (or anti-motifs) when their appearance (or absence) in a network is deemed to be statistically significant (*25*). For example, triangular substructures appear in a wide range of real networks, indicating a common tendency for clustering. However, networks in different domains appear to display different motif profiles, suggesting

that quantifying propensity for substructures could play a crucial role in understanding network function. For example, the feed-forward ($\triangle$) and the bi-fan ($\square$) substructures are prevalent in protein interaction networks, while the bi-directional two-hop ($\triangle$) substructure is more prominent in transportation networks. Current methods for quantifying a network's propensity for various substructures are generally characterized by the following two choices. First, one needs a network statistic to measure the propensity for the substructures of interest. Second, to determine statistical significance, one needs a null model for the observed network: an ensemble of networks that matches some set of properties deemed to be important.

For the first choice, a common approach is to simply count the number of instances of a substructure (*25, 26, 27*). However, substructure counts can be misleading, and are generally not faithful measures of their propensities (*28, 29*). For example, the number of triangles (or of any other substructure) naturally correlates with the number of larger substructures that contain them as subgraphs. Unfortunately, there is currently no standard measure of the propensity for an arbitrary substructure (*27*); other choices tend to be rather context-dependent or specifically tailored to the substructure of interest (e.g., clustering coefficients), often leading to a proliferation of essentially incomparable statistics. This issue is exacerbated when one considers incorporating additional information, such as directed edges, node attributes, and edge weights.

Regardless of the measure used, in order to assess whether its value in the observed network is statistically significant, one must compare against its distribution in some appropriate null model. This distribution is often obtained by measuring its value in networks explicitly sampled from the chosen null model. Replicating the observed degree sequence (i.e., the configuration model) is a popular choice of null model. In some cases, this node-centric model may be appropriate, such as when the nodes in the observed network have unique identities that should indeed be preserved by the null model (e.g., generating randomized networks of transactions

3

between some specific set of countries). However, the defining symmetry of graphs is that nodes are *a priori* indistinguishable (i.e., node exchangeability), and often one desires a null model that is similar to the observed network, but not necessarily with exactly the "same" nodes (e.g., generating typical networks of interactions between students to model the spread of infections). In addition, the configuration model does not treat all substructures equally. For example, hubs and clustering, both hallmark features of real networks, are associated with a propensity for (k-)stars (i.e., a central node connected to k others) and for (k-)cliques (i.e., fully connected groups of k nodes), respectively. However, the configuration model fully constrains the counts of all k-stars, and is therefore useless to assess their relative propensity in the observed network. Moreover, by design, it does not promote clustering at any level, and is thus inappropriate to study the higher-order organization of clustering. Even if one were to employ modifications to promote triangles, the issue is simply postponed to slightly larger substructures (*30*).

For a real-valued random variable, cumulants provide a hierarchical sequence of summary statistics that efficiently encode its distribution (*31, 32, 33*). Aside from their unique mathematical properties[1], low-order cumulants have intuitive interpretations. The first two orders, the mean and variance, correspond to the center of mass and spread of a distribution, and are taught in nearly every introductory statistics course (*34*). The next two orders have likewise been given unique names (skew and kurtosis), and are useful to describe data that deviate from normality, appearing in a variety of applications, such as finance (*35*), economics (*36*), and psychology (*37*). Generalizations of these notions have proven similarly useful: for example, joint cumulants (e.g., covariance) are the natural choice for measuring correlations in multivariate random variables. Unsurprisingly, cumulants are essentially universally used by the statistics community.

By generalizing the combinatorial definition of cumulants (*38, 39, 40*), we obtain their

---

[1]In particular, their additive nature when applied to sums of independent random variables (see supplementary materials S8).

analogue for networks, which we refer to as *graph cumulants*. After introducing the framework, we demonstrate their usefulness as principled and intuitive measures of the propensity (or aversiveness) of a network to display any substructure of interest, allowing for systematic and meaningful comparisons between networks. We then describe how this framework seamlessly incorporates additional network features, providing examples using real datasets containing directed edges, node attributes, and edge weights. Finally, we show how graph cumulants give rise to a natural hierarchical family of maximum entropy null models from a single observed network.

For ease of exposition, we first introduce our framework for the case of simple graphs, i.e., undirected, unweighted networks with no self-loops or multiple edges.

Graph moments

The cumulants of a real-valued random variable are frequently given in terms of its moments, i.e., the expected value of its powers. A real network can be thought of as a realization of some graph-valued random variable, and its graph cumulants are likewise given in terms of its graph moments.

To motivate our definition of graph moments, consider the measurement that provides the smallest nonzero amount of information about a network G with n nodes. This measurement is a binary query, which yields 1 if an edge exists between a random pair of nodes in G, and 0 otherwise[2]. At first order, we consider repeated observations of a *single* such measurement. We define the first-order graph moment $\mu_{1/}$ (the "mean") as the expected value of this quantity: the counts of edges in G normalized by the counts of edges in the complete graph with n nodes. Hence, the first-order graph moment of a network is simply its edge density.

---

[2]One might consider measuring a random node instead. However, the presence of a node *per se* does not give any new information. On the other hand, querying the degree of a node yields an integer, containing more information than the binary result from querying the existence of a single edge.

For the second-order graph moments, we again consider a binary query, but now using *two simultaneous* such measurements. This query yields 1 if G contains edges between both pairs of nodes, and 0 otherwise. We now must distinguish between two new cases: when the two edges share a node, thereby forming a wedge ($\mu_{2\wedge}$); and when they do not share any node ($\mu_{2\parallel}$). Each case is associated with a different second-order graph moment, which we analogously define as the counts of the associated subgraph in G, normalized by the counts of this subgraph in the complete graph with n nodes. Hence, the second-order graph moments of a network are the densities of substructures with two edges.

Likewise, we define an rth-order graph moment for each of the ways that r edges can relate to each other (see fig. 1 for the subgraphs associated with graph moments up to third order). Again, μ_{rg} is the density of subgraph g, defined as its counts in G normalized by its counts in the complete graph with the same number of nodes as G.
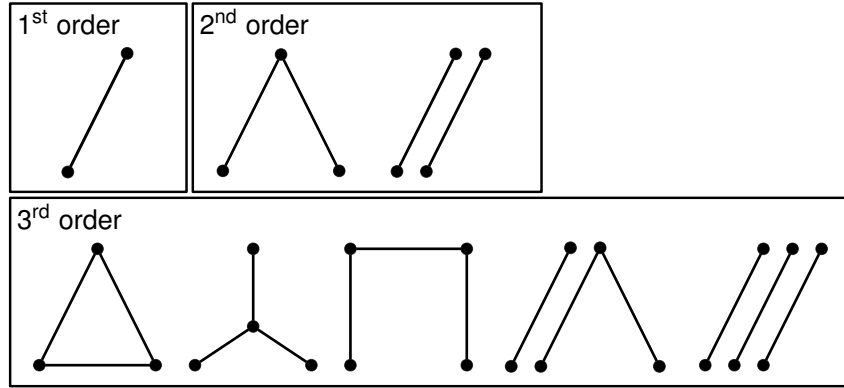
Fig. 1: Subgraphs associated with graph moments. The graph moment μ_{rg} of a network G is defined as the counts of the subgraph g (with r edges), normalized by the counts of this subgraph if connections were present between all pairs of nodes in G. Displayed here are the subgraphs g associated with the graph moments up to third order (for simple graphs). The full set of rth-order moments contains all substructures with exactly r edges (including disconnected subgraphs), corresponding to all the ways that r edges can relate to each other.

The scalability of computing graph moments is determined by the computational complexity

associated with counting connected subgraphs, as these imply the counts of disconnected subgraphs (see supplementary materials S1.1 for details). This is an important fact, as the counts of disconnected subgraphs are generally orders of magnitude larger, and we can leverage numerous methods for efficiently counting connected subgraphs (*41, 42, 43, 44*).

Graph cumulants

As the density of smaller substructures increases, the appearance of larger substructures that contain them as subgraphs will clearly also tend to increase. Hence, we would like to measure the difference between the observed value of a given graph moment and that which would be expected due to graph moments of lower order, so as to quantify the propensity (or aversiveness) for that specific network substructure. Cumulants are the natural statistics with this desired property. For example, the variance quantifies the intuitive notion of the "spread" of a distribution (regardless of its mean), while the skew and the kurtosis reflect, respectively, the asymmetry and contribution of large deviations to this spread.

While often defined via the cumulant generating function (*45, 46*), cumulants have an equivalent combinatorial definition (*38, 39, 40*) (see fig. 2). At order r, it involves the partitions of a set of r elements:

$$\mu_r = \sum_{\pi \in P_r} \prod_{p \in \pi} \kappa_{|p|}, \tag{1}$$

where μ_r is the rth moment, κ_r is the rth cumulant, P_r is the set of all partitions of a set of r unique elements, π is one such partition, p is a subset of a partition π, and $|p|$ is the number of elements in a subset p.

When generalizing this definition to graph moments, the partitioning P_r of the edges must

7

respect their connectivity (see fig. 2, bottom row), i.e.

$$\mu_{rg} = \sum_{\pi \in P_{E(g)}} \prod_{p \in \pi} \kappa_{|p| \, g_p}, \tag{2}$$

where $E(g)$ is the set of the r edges forming subgraph g, $P_{E(g)}$ is the set of partitions of these edges, and g_p is the subgraph formed by the edges in a subset p. These expressions can then be inverted to yield the graph cumulants in terms of graph moments (summarized in supplementary materials S10 and provided in our code).
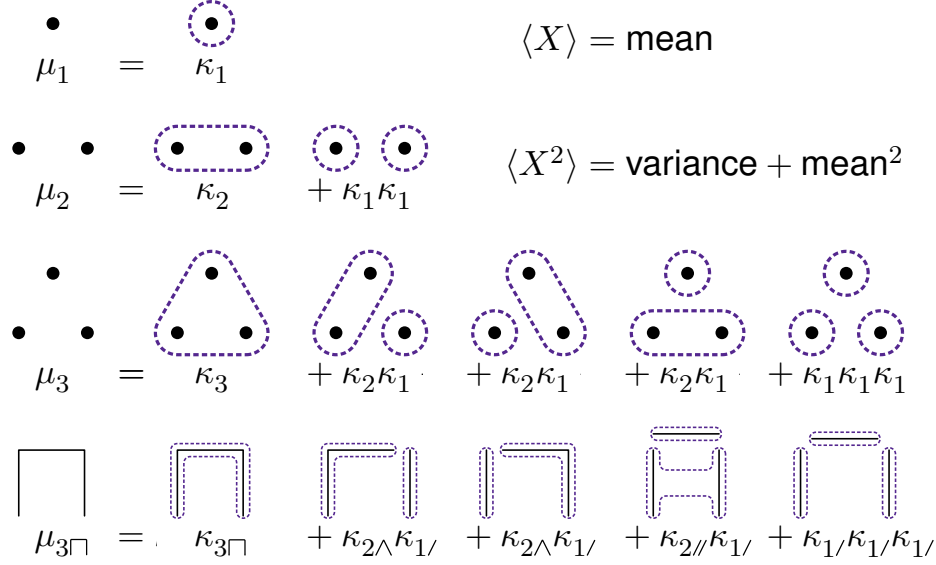
Fig. 2: To expand an rth-order graph moment in terms of graph cumulants, enumerate all partitions of the r edges comprising its subgraph. The top three rows illustrate the combinatorial expansion of the classical moments in terms of cumulants. Analogously, the bottom row shows how to expand $\mu_{3\sqcap}$ in terms of graph cumulants. The last term ($\kappa_{1/}^3$) corresponds to partitioning this subgraph into three subsets, each with a single edge. The first term ($\kappa_{3\sqcap}$) corresponds to "partitioning" this subgraph into a single subset containing all three edges, thus inheriting the connectivity of the entire subgraph. The remaining terms ($\kappa_2\kappa_1$) correspond to partitioning this subgraph into a subset with one edge and a subset with two edges. This can be done in three different ways: in two cases (the two $\kappa_{2\wedge}\kappa_{1/}$ terms), the subset with two edges has those edges sharing a node; and in one case (the $\kappa_{2/\!/}\kappa_{1/}$ term), the subset with two edges has those edges not sharing any node.

Essentially, the defining feature of cumulants is their additive nature when summing independent random variables (*31, 32, 46, 45*). In supplementary materials S8, we show that the graph cumulants of independent graph-valued random variables also have this additive property for a natural notion of summing graphs. Moreover, we remark that the Erdős–Rényi distribution has graph cumulants of zero for all orders $r \geq 2$, similar to the classical cumulants of the normal distribution, which are zero for all orders $r \geq 3$.

Quantifying propensity for substructures: scaled graph cumulants

Cumulants are often scaled, as dimensionless quantities allow for interpretable comparisons. For example, the precision of a measurement is often quantified by the relative standard deviation $(\kappa_2^{1/2}/\kappa_1)$, and the linear correlation between two random variables X and Y is often quantified by the Pearson correlation coefficient $(\mathrm{Cov}(X,Y)/\kappa_2^{1/2}(X)\kappa_2^{1/2}(Y)$, i.e., their second-order joint cumulant divided by the geometric mean of their individual second-order cumulants). Likewise, we define scaled graph cumulants as $\tilde{\kappa}_{rg} \equiv \kappa_{rg}/\kappa_{1/}^r$, and report the "signed" r^{th} root, i.e., the real number with magnitude equal to $\left|\tilde{\kappa}_{rg}\right|^{1/r}$ and with the same sign as $\tilde{\kappa}_{rg}$. These scaled graph cumulants offer principled measures of the propensity (or aversiveness) of a network to display any substructure of interest.

Scaling graph cumulants also allows for meaningful comparisons of the propensity of different networks to exhibit a particular substructure, even when these networks have different sizes and edge densities. To illustrate this point, we consider clustering, a hallmark feature of many real networks (*7, 8, 13*). This notion is frequently understood as the prevalence of triadic closure (*4, 9*), an increased likelihood that two nodes are connected if they have mutual neighbors, i.e., a propensity for triangles. This property is often quantified by the clustering coefficient $C_{\triangle}$, defined as the probability that two neighbors of the same node are themselves connected. While this quantity is easily expressed within our formalism as $C_{\triangle} = \mu_{3\triangle}/\mu_{2\wedge}$, it is neither a graph cumulant nor dimensionless. We propose that the scaled triangle cumulant $\tilde{\kappa}_{3\triangle}$ is a more appropriate measure of clustering in networks, as demonstrated in fig. 3 (for the natural extension to clustering in bipartite networks, see fig. S1).
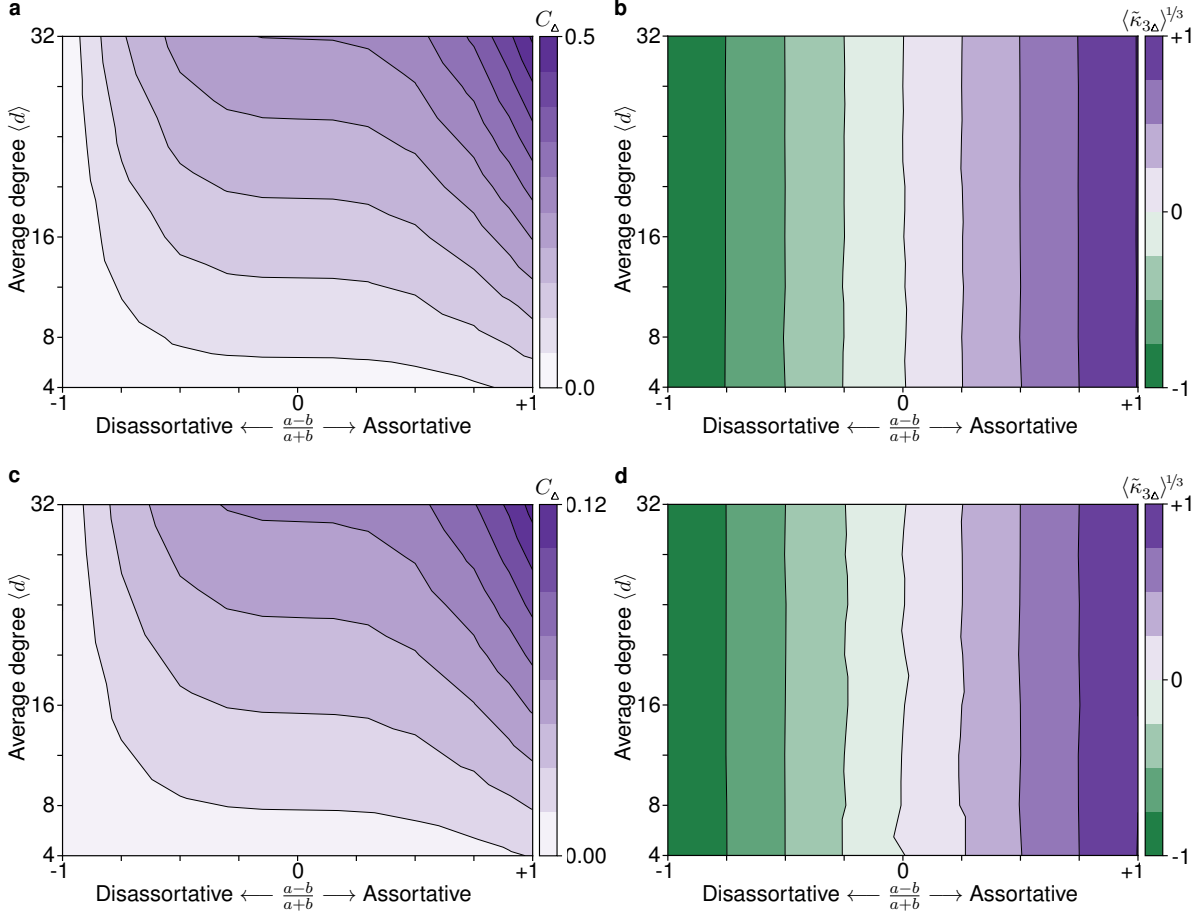
Fig. 3: The scaled triangle cumulant provides a principled measure of triadic closure. We sampled networks from the symmetric stochastic block model on n nodes with 2 communities, $\mathrm{SSBM}(n, 2, a, b)$, for a range of a and b, for both 128 nodes (top) and 512 nodes (bottom). The horizontal axis indicates the level of assortativity: 0 corresponds to Erdős–Rényi (ER) graphs (no community structure), $+1$ to two disjoint ER graphs (only connections within the two communities), and -1 to random bipartite graphs (only connections between the two communities, thus, no triangles). The vertical axis indicates the edge density, in terms of average degree. For each set of parameters, we compute the average of both the global clustering coefficient $C_\triangle$ (left) and scaled triangle cumulant $\tilde{\kappa}_{r\triangle}$ (right), displaying the signed fourth root for $\tilde{\kappa}_{3\triangle}$. While both measures increase with assortativity (as desired), the clustering coefficient also varies with network size and edge density, whereas the scaled triangle cumulant is invariant to such differences.

Graph cumulants for networks with additional features

While it is possible to treat most networks as undirected, unweighted graphs, real networks frequently contain more information (e.g., node attributes). Our formalism naturally incorporates many such augmentations. As before, there is a graph moment of order r for each of the unique substructures (now endowed with the additional features) containing r edges. The conversion to graph cumulants likewise respects the additional features. We now discuss the specific cases of directed edges, node attributes, and edge weights. In supplementary materials S2, we illustrate their ability to quantify the propensity for different substructures in a variety of real networks with these additional features. In supplementary materials S10, we provide expressions for computing some of these augmented graph moments and graph cumulants. For clarity, we consider each of these features individually; combining them is relatively straightforward.

Directed edges

When analyzing a directed network, the graph moments must incorporate the orientation of the edges. While the first-order graph moment ($\mu_{1\prime}$) still simply considers the number of directed edges (as edge orientation only carries meaning when considered with respect to the orientation of some other edge), there are now five second-order directed graph moments. The wedge configuration (two edges sharing one node) is now associated with three moments: one with both edges oriented towards the central node ($\mu_{2\wedge}$), one with both edges oriented away from it ($\mu_{2\wedge}$), and one with an edge towards and the other away ($\mu_{2\wedge}$). The relative orientation of two edges that do not share any node cannot be determined, and therefore is still associated with a single moment ($\mu_{2\#}$). Finally, the configuration of two reciprocal edges (i.e., two nodes that are connected by edges in both directions) is associated with the fifth second-order moment ($\mu_{2\emptyset}$). The appropriate normalization is with respect to the counts in the complete directed graph, i.e.,

that which has every pair of nodes connected by edges in both directions. Fig. S3 illustrates how incorporating the directed nature of protein interaction networks reveals additional structure.

Node attributes

Often, nodes of a network have intrinsic attributes (e.g., demographics for social networks). The graph moments of such networks are defined by endowing the subgraphs with same attributes. For example, consider the case of a network in which every node has one of two possible "flavors": "charm" and "strange". There are now three first-order graph moments: an edge between two "charm" nodes (μ_{1}), an edge between two "strange" nodes (μ_{1}), and an edge between one of each (μ_{1}). To compute the moments, we normalize by their counts in the complete graph on the same set of nodes: here, $\binom{n_\bullet}{2}$, $\binom{n_\bullet}{2}$, and $n_\bullet n_\bullet$, respectively. Fig. S2 considers a (binary) gendered network of primary school students (*47*), illustrating how incorporating node attributes elucidates the correlations between node type and their connectivity patterns.

A common special case of networks with node attributes are bipartite networks: nodes have one of two types (e.g., authors and publications (*48*), plants and pollinators (*49*)) and edges are only allowed between nodes of different type. As certain subgraphs are now unrealizable, bipartite networks have only one first-order moment (an edge connecting a "charm" to a "strange", μ_{1}), and two second-order wedge moments: a "charm" node connected to two "strange" nodes ($\mu_{2\wedge}$), and a "strange" node connected to two "charm" nodes ($\mu_{2\wedge}$).

Edge weights

To compute graph moments for weighted networks, subgraphs should be counted with multiplicity equal to the product of their edge weights (*24*) (see supplementary materials S8 for a detailed justification). The normalization is the same as in the unweighted case, i.e., divide

the (weighted) count of the relevant subgraph by the count of this subgraph in the unweighted complete graph with the same number of nodes. Note that, unlike unweighted networks, graph moments may be greater than one for weighted networks. In fig. S4, we analyze a weighted network of social interactions (*50*), illustrating how allowing for variable connection strength can increase statistical significance and even change the resulting interpretations.

Inference from a single network

Unbiased estimators of graph cumulants

Thus far, we have not made a distinction between the graph moments of an observed network G and those of the distribution $\mathcal{G}$ from which this network was sampled. This is because, in a sense, they are the same: $\langle \mu_{rg}(G) \rangle = \mu_{rg}(\mathcal{G})$ (where the angled brackets $\langle \cdot \rangle$ denote expectation with respect to the distribution $\mathcal{G}$), a property known as "inherited on the average" (*51*).

However, for cumulants, this distinction is important. Cumulants of a distribution are defined by first computing the moments of the distribution, then converting them to cumulants (as opposed to computing the cumulants of the individual samples, then taking the expectation of those quantities). Due to the fact that the cumulants are nonlinear functions of the moments, they are not necessarily preserved in expectation, i.e., not inherited on the average (*51, 52*). For example, consider the problem of estimating the variance of a distribution over the real numbers given n observations from it. Simply using the variance of these observations gives an estimate whose expectation is less than the variance of the underlying distribution, and one should multiply it by the well-known correction factor of $\frac{n}{n-1}$. The generalizations of this correction factor for higher-order cumulants are known as the k-statistics (*52*); given a finite number of observations, they are the minimum-variance unbiased estimators of the cumulants of the underlying distribution (*53, 54, 55, 56, 57, 58*).

In many applications, one wishes to estimate a probability distribution over networks $\mathcal{G}$ after

14

observing only a single network G. Just as with classical cumulants, applying equation 2 to this network also yields biased estimates of the graph cumulants of the underlying distribution, i.e., $\langle \kappa_{rg}(G) \rangle \neq \kappa_{rg}(\mathcal{G})$. In supplementary materials S3, we describe a procedure to obtain unbiased estimators of graph cumulants $\hat{\kappa}_{rg}$, as well as the variance of these estimators $\mathrm{Var}(\hat{\kappa}_{rg})$ (supplementary materials S6.2). In particular, for simple graphs, we provide the expressions for these unbiased estimators up to third order, as well as the variance for first order. Obtaining a complete list of the expressions for the unbiased estimators of graph cumulants and their variance would provide a powerful tool for network science (see supplementary materials S6.1 for a detailed discussion). This would allow for principled statistical tests of the propensity (or aversiveness) for any substructure without explicitly sampling from some null model, a procedure that is in general quite computationally expensive and often a main obstacle in the analysis of real networks (*20, 59, 60, 61, 62*).

However, sometimes one indeed requires samples from the underlying distribution, such as when assessing the statistical significance of some property that cannot be easily expressed in terms of the statistics defining this distribution. We now discuss how these unbiased graph cumulants can be used to obtain a principled hierarchical family of network models (see supplementary materials S4 and S5 for more details).

A principled hierarchical family of network models

A ubiquitous problem, arising in many forms, is that of estimating a probability distribution based on partial knowledge. Often, one desires the distribution to have some set of properties, but the problem is typically still highly unconstrained. In such cases, the maximum entropy principle (*63, 64, 65*) provides a natural solution: of the distributions that satisfy these desired properties, choose the one that assumes the least amount of additional information, i.e., that which maximizes the entropy. For example, when modeling real-valued data, one often uses a

15

Gaussian, the maximum entropy distribution (over the reals) with prescribed mean and variance.

The analogous maximum entropy distributions for networks are known as exponential random graph models (ERGMs). These models are used to analyze a wide variety of real networks (*66, 67, 68, 69, 70*). Although it is possible to define an ERGM by prescribing any set of realizable properties, attention is often given to the counts of edges and of other context-dependent substructures (such as wedges or triangles for social networks (*4,60,71,72*)). While individual networks sampled from this distribution do not necessarily have the same counts (of edges and of specified substructures), the average values of these counts are required to match those of the observed network.

Unfortunately, ERGMs of this type often result in pathological distributions, exhibiting strong multimodality, such that typical samples have properties far from those of the observed network they were intended to model. For example, the networks sampled from such an ERGM might be either very dense or very sparse, despite the fact that averaging over these networks yields the same edge count as that of the observed network. This phenomenon is known as the "degeneracy problem", and much effort has gone into understanding it (*73, 72, 74*). While some remedies have been proposed (*75, 76, 77*), a principled and systematic method for alleviating degeneracy has thus far remained elusive.

Based on our framework, we propose a hierarchical family of ERGMs that are immune to the degeneracy problem (see fig. 4 for an example and supplementary materials S5 for details). Our hierarchy is based on correlations between an increasing number of individual connections: at r^{th} order, the proposed ERGM is specified by *all* the graph cumulants of order at most r. Importantly, when inferring such an ERGM $\mathcal{G}$ from a single observed network G, it is appropriate to use the *unbiased* graph cumulants, such that $\kappa_{rg}(\mathcal{G}) = \hat{\kappa}_{rg}(G)$ for all subgraphs through the chosen order (see supplementary materials S4 for the detailed protocol).

There are two main differences between our proposed family of ERGMs and those typically

16

used in the literature. First, our family prescribes the expected counts of *all* subgraphs (including disconnected) with at most the desired order (see supplementary materials S9 for a motivation of this choice based on a spectral representation of distributions over networks). In contrast, current ERGMs only consider some (usually connected) subgraphs of interest, perhaps due to the fact that not all of these subgraphs are deemed important to model the observed network, or because disconnected subgraphs are not usually thought of as motifs (*25, 78*). Second, the use of unbiased graph cumulants results in a distribution with expected subgraph counts different from those of the observed network. Nevertheless, the ERGM distribution induced by this choice generates samples that are appropriately clustered around the observed network (see fig. 4 and supplementary materials S5 for detailed intuition).
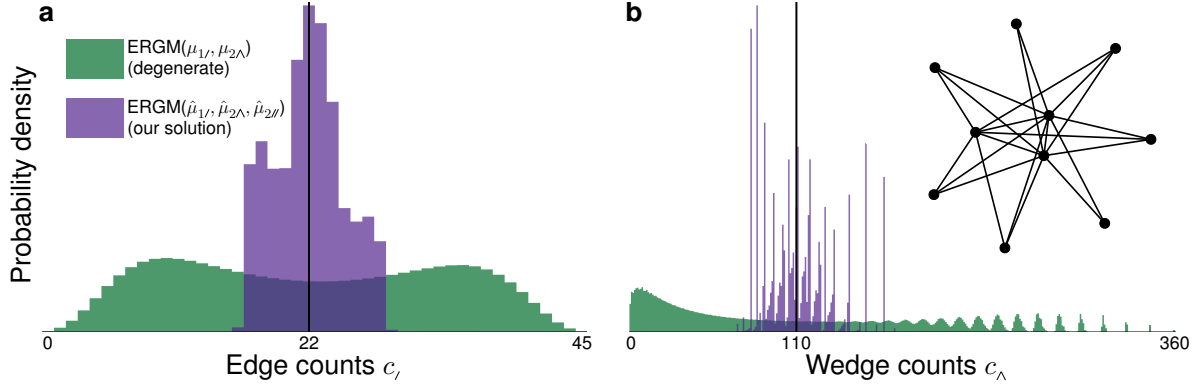
Fig. 4: Our proposed family of ERGMs does not suffer from the "degeneracy problem".
We compute the exact distributions over simple graphs with 10 nodes resulting from fitting two
ERGMs to the same observed network (shown on the right), and plot the resulting distributions
of: **a)** edge counts, and **b)** wedge (i.e., 2-star) counts. Green corresponds to a network
model frequently used in the literature: $\text{ERGM}(\mu_{1/}, \mu_{2\wedge})$, the maximum entropy distribution
with prescribed expected counts of edges and wedges. Purple corresponds to our analogous
second-order network model: $\text{ERGM}(\hat{\mu}_{1/}, \hat{\mu}_{2\wedge}, \hat{\mu}_{2/\!/})$, with prescribed expected *unbiased* counts
of substructures of first and second order, i.e., edges, wedges, *and* two edges that do not share
any node. We fit the model using the procedure described in supplementary materials S4, with
unbiasing parameter $\eta = \frac{1}{11}$. Black lines denote the counts in the observed network that these
distributions are intended to model. In sharp contrast to our proposed $\text{ERGM}(\hat{\mu}_{1/}, \hat{\mu}_{2\wedge}, \hat{\mu}_{2/\!/})$, the
currently used $\text{ERGM}(\mu_{1/}, \mu_{2\wedge})$ can result in a distribution whose typical samples are notably
different from the observed network. This is reflected in the fact that the green distributions
have maxima far from their means. This undesirable behavior, known as the degeneracy
problem, tends to become even more pronounced for larger networks (*75*). In supplementary
materials S5, we explain why this occurs and why our proposed family of ERGMs does not
suffer from this problem.

Discussion

Over a century ago, Thiele introduced cumulants (*79*), a concept now fundamental to the field

of statistics (*32, 33, 46, 45, 39, 40*), which has justifiably percolated throughout the scientific

community (*35, 36, 37, 38, 80*). In this work, we introduce *graph cumulants*, their generalization

to networks (fig. 2). This principled hierarchy of network statistics provides a framework to

systematically describe and compare networks (figs. 3 and S1), naturally including those with additional features, such as directed edges (fig. S3), node attributes (fig. S2), and edge weights (fig. S4). Moreover, through the lens of the maximum entropy principle, these statistics induce a natural hierarchical family of network models. These models are immune to the "degeneracy problem", providing a principled prescription for obtaining distributions that are clustered around the properties of the network they intend to model (figs. 4, S5, and S6).

To make appropriate predictions, one must acknowledge that the observed data are but one instantiation, inherently incomplete and stochastic, of some underlying process. In network science, one typically has a single network observation, and would like to make inferences about the distribution from which it came. This is analogous to characterizing the distribution of a classical random variable given a finite collection of samples. However, aside from the mean, the cumulants of a finite sample are not the same (in expectation) as those of the underlying distribution. The desired unbiased estimators are known as the k-statistics (*52, 56, 58*) (e.g., the $\frac{n}{n-1}$ correction factor for the sample variance). Characterizing the distributions of these unbiased estimators allows for a variety of principled statistical tests (e.g., the use of the χ^2 distribution for analyzing the sample variance). In supplementary materials S3, we provide a procedure for deriving the analogous unbiased estimators of graph cumulants given a single network observation, and in supplementary materials S6.2, we provide a procedure for deriving their variance. While the derivations are incredibly tedious, once the expressions are obtained, they could be used to systematically measure the statistical significance of the propensity (or aversiveness) for arbitrary substructures (see supplementary materials S6.1). This is an incredibly promising avenue, as it circumvents the need for constructing and sampling from a network null model, a major challenge for many current methods. As with the rest of our framework, such an analysis naturally incorporates additional features, such as directed edges, node attributes and edge weights.

Graph cumulants quantify the propensity for substructures throughout the *entire* network. However, in some applications, statistics that quantify the propensity of *individual* nodes (or edges) to participate in these substructures are more appropriate. For example, the local clustering coefficient (*13*) aims to describe the propensity of a node to participate in triangles (as does the edge clustering coefficient for edges). However, as before, there is no general framework for arbitrary substructures. Our graph cumulant formalism again offers a systematic prescription. Essentially, the local graph cumulants are obtained by giving the node (or edge) of interest a unique identity and computing the graph cumulants of the entire network with this augmented information (see supplementary materials S7 for details). These local graph cumulants could serve as useful primitives in machine learning tasks such as node classification (*81, 82, 83*) and link prediction (*84*).

Just as the scientific community has converged upon the variance as the canonical measure of the spread of a distribution, the field of network science could greatly benefit from a similarly standardized measure of the propensity for an arbitrary substructure. Inspired by over a century of work in theoretical statistics, the framework of graph cumulants introduced in this paper provides a uniquely principled solution.

References

1. D. J. Klein, M. Randić, Resistance distance. *J. Math. Chem.* **12**, 81-95 (1993).

2. C. W. Lynn, D. S. Bassett, The physics of brain network structure, function and control. *Nat. Rev. Phys.* **1**, 318-332 (2019).

3. P. Landi, H. O. Minoarivelo, Å. Brännström, C. Hui, U. Dieckmann, Complexity and stability of ecological networks: a review of the theory. *Popul. Ecol.* **60**, 319-345 (2018).

4. S. Wasserman, K. Faust, *Social Network Analysis: Methods and applications* (Cambridge Univ. Press, Cambridge, 1994).

5. M. G. Bell, Y. Iida, *Transportation network analysis* (Wiley, 1997).

6. W. Hall, T. Tiropanis, Web evolution and web science. *Comput. Netw.* **56**, 3859-3865 (2012).

7. G. Cimini, *et al.*, The statistical physics of real-world networks. *Nat. Rev. Phys.* **1**, 58-71 (2019).

8. A.-L. Barabási, *et al.*, *Network Science* (Cambridge university press, 2016).

9. M. Newman, *Networks* (Oxford university press, 2018).

10. S. Emmons, S. Kobourov, M. Gallant, K. Börner, Analysis of network clustering algorithms and cluster quality metrics at scale. *PloS One* **11**, e0159161 (2016).

11. S. Goswami, C. Murthy, A. K. Das, Sparsity measure of a network graph: Gini index. *Inf. Sci.* **462**, 16-39 (2018).

12. E. D. Demaine, *et al.*, Structural sparsity of complex networks: Bounded expansion in random models and real-world graphs. *J. Comput. Syst. Sci.* **105**, 199-241 (2019).

13. D. J. Watts, S. H. Strogatz, Collective dynamics of "small-world" networks. *Nature* **393**, 440-442 (1998).

14. A.-L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509-512 (1999).

15. E. R. Colman, G. J. Rodgers, Complex scale-free networks with tunable power-law exponent and clustering. *Physica A* **392**, 5501-5510 (2013).

16. D. Cai, T. Campbell, T. Broderick, "Edge-exchangeable graphs and sparsity", in *NIPS'16 Proceedings of the 29th International Conference on Neural Information Processing Systems* (2016).

17. A. Kartun-Giles, D. Krioukov, J. Gleeson, Y. Moreno, G. Bianconi, Sparse power-law network model for reliable statistical predictions based on sampled data. *Entropy* **20**, 257 (2018).

18. P. Holme, B. J. Kim, Growing scale-free networks with tunable clustering. *Phys. Rev. E* **65**, 026107 (2002).

19. C. I. Del Genio, T. Gross, K. E. Bassler, All scale-free networks are sparse. *Phys. Rev. Lett.* **107**, 178701 (2011).

20. C. Orsini, *et al.*, Quantifying randomness in real networks. *Nat. Commun.* **6**, 8627 (2015).

21. F. Caron, E. B. Fox, Sparse graphs using exchangeable random measures. *J. Royal Stat. Soc.* **79**, 1295-1366 (2017).

22. C. I. Sampaio Filho, A. A. Moreira, R. F. Andrade, H. J. Herrmann, J. S. Andrade Jr, Mandala networks: ultra-small-world and highly sparse graphs. *Sci. Rep.* **5**, 9082 (2015).

23. K. Zuev, M. Boguná, G. Bianconi, D. Krioukov, Emergence of soft communities from geometric preferential attachment. *Sci. Rep.* **5**, 9421 (2015).

24. L. Lovász, *Large Networks and Graph Limits* (American Mathematical Soc., 2012).

25. R. Milo, *et al.*, Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827 (2002).

26. U. Alon, Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450-461 (2007).

27. F. Xia, H. Wei, S. Yu, D. Zhang, B. Xu, A survey of measures for network motifs. *IEEE Access* **7**, 106576–106587 (2019).

28. C. Fretter, M. Müller-Hannemann, M.-T. Hütt, Subgraph fluctuations in random graphs. *Phy. Rev. E* **85**, 056119 (2012).

29. R. Ginoza, A. Mugler, Network motifs come in sets: Correlations in the randomization process. *Phys. Rev. E* **82**, 011921 (2010).

30. M. Ritchie, L. Berthouze, I. Z. Kiss, Generation and analysis of networks with a prescribed degree sequence and subgraph family: higher-order structure matters. *J. Complex Netw.* **5**, 1-31 (2017).

31. T. N. Thiele, *Theory of observations* (Charles & Edwin Layton, 1903).

32. G.-C. Rota, J. Shen, On the combinatorics of cumulants. *J. Combin. Theory A* **91**, 283-304 (2000).

33. P. McCullagh, *Tensor Methods in Statistics: Monographs on Statistics and Applied Probability* (Chapman and Hall, 1987).

34. B. Gunderson, M. Aliaga, *Interactive statistics* (Prentice Hall, 1999).

35. X. Wu, M. Xia, H. Zhang, Forecasting var using realized egarch model with skewness and kurtosis. *Finance Res. Lett.* (2019).

36. S. Vähämaa, Skewness and kurtosis adjusted black-scholes model: A note on hedging performance. *Finance Lett.* **1**, 6-12 (2003).

37. M. J. Blanca, J. Arnau, D. López-Montiel, R. Bono, R. Bendayan, Skewness and kurtosis in real data samples. *Meth. Eur. J. Res. Meth. Behav. Soc. Sci.* **9**, 78-84 (2013).

38. M. Kardar, *Statistical Physics of Particles* (Cambridge Univ. Press., 2007).

39. T. Speed, Cumulants and partition lattices i. *J. Aust. Math. Soc.* **25**, 378-388 (1983).

40. T. Speed, Cumulants and partition lattices ii: Generalised k-statistics. *J. Aust. Math. Soc.* **40**, 34-53 (1986).

41. R. Curticapean, H. Dell, D. Marx, "Homomorphisms are a good basis for counting small subgraphs", in *STOC'17 Proceedings of the 49th ACM Symposium on Theory of Computing* (2017).

42. F. V. Fomin, D. Lokshtanov, V. Raman, S. Saurabh, B. R. Rao, Faster algorithms for finding and counting subgraphs. *J. Comput. Syst. Sci.* **78**, 698-706 (2012).

43. O. Amini, F. V. Fomin, S. Saurabh, "Counting subgraphs via homomorphisms", in *International Colloquium on Automata, Languages, and Programming* (2009).

44. A. Pinar, C. Seshadhri, V. Vishal, "Escape: Efficiently counting all 5-vertex subgraphs", in *Proceedings of the 26th International Conference on World Wide Web* (2017).

45. A. Hald, The early history of the cumulants and the gram-charlier series. *Int. Stat. Rev.* **68**, 137-153 (2000).

46. B. Gnedenko, A. Kolmogorov, *Limit distributions for sums of independent random variables* (Addison-Wesley, 1954).

47. J. Stehl'e, *et al.*, High-resolution measurements of face-to-face contact patterns in a primary school. *PloS One* **6**, e23176 (2011).

48. G. Bravo-Hermsdorff, *et al.*, Gender and collaboration patterns in a temporal scientific authorship network. *Appl. Netw. Sci.* **4**, 112 (2019).

49. C. Campbell, S. Yang, R. Albert, K. Shea, A network model for plant–pollinator community assembly. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 197-202 (2011).

50. L. Isella, *et al.*, What's in a crowd? analysis of face-to-face behavioral networks. *J. Theor. Biol.* **271**, 166-180 (2011).

51. J. W. Tukey, Some sampling simplified. *J. Am. Stat. Assoc.* **45**, 501-519 (1950).

52. R. A. Fisher, Moments and product moments of sampling distributions. *P. Lond. Math. Soc.* **2**, 199-238 (1930).

53. P. R. Halmos, The theory of unbiased estimation. *Ann. Math. Stat.* **17**, 34-43 (1946).

54. M. Kendall, Some properties of k-statistics. *Ann. Eugen.* **10**, 106–111 (1940).

55. M. Kendall, Proof of fisher's rules for ascertaining the sampling semi-invariants of k-statistics. *Ann. Eugen.* **10**, 215-222 (1940).

56. J. F. Kenney, E. Keeping, *Mathematics of statistics Vol. II* (D. Van Nostrand Co., 1951).

57. G. James, On moments and cumulants of systems of statistics. *Sankhyā: The Indian Journal of Statistics* **20**, 1-30 (1958).

58. N. Fisher, Unbiased estimation for some non-parametric families of distributions. *Ann. Stat.* **10**, 603-615 (1982).

59. C. T. Butts, A perfect sampling method for exponential family random graph models. *J. Math. Sociol.* **42**, 17-36 (2018).

60. G. Robins, P. Pattison, Y. Kalish, D. Lusher, An introduction to exponential random graph (p*) models for social networks. *Soc. Netw.* **29**, 173-191 (2007).

61. P. Wang, K. Sharpe, G. L. Robins, P. E. Pattison, Exponential random graph p* models for affiliation networks. *Soc. Netw.* **31**, 12-25 (2009).

62. V. Veitch, D. M. Roy, Sampling and estimation for (sparse) exchangeable graphs. *Ann. Stat.* **47**, 3274-3299 (2019).

63. E. T. Jaynes, Information theory and statistical mechanics. *Phy. Rev.* **106**, 620-630 (1957).

64. E. T. Jaynes, Information theory and statistical mechanics. ii. *Phys. Rev.* **108**, 171-190 (1957).

65. S.-I. Amari, *Information geometry and its applications*, vol. 194 (Springer, 2016).

66. Z. M. Saul, V. Filkov, Exploring biological network structure using exponential random graph models. *J. Bioinform.* **23**, 2604-2611 (2007).

67. A. Chakraborty, H. Krichene, H. Inoue, Y. Fujiwara, Exponential random graph models for the japanese bipartite network of banks and firms. *J. Comput. Soc. Sci.* **2**, 3-13 (2019).

68. J. Koskinen, P. Wang, G. Robins, P. Pattison, Outliers and influential observations in exponential random graph models. *Psychometrika* **83**, 809-830 (2018).

69. M. E. J. N. Juyong Park, Statistical mechanics of networks. *Phy. Rev. E* **70**, 066117 (2004).

70. C. E. Buddenhagen, *et al.*, Epidemic network analysis for mitigation of invasive pathogens in seed systems: Potato in ecuador. *Phytopathology* **107**, 1209-1218 (2017).

71. D. Strauss, On a general class of models for interaction. *SIAM Review* **28**, 513-527 (1986).

72. M. E. J. N. Juyong Park, Solution of the two-star model of a network. *Phy. Rev. E* **70**, 066146 (2004).

73. S. Chatterjee, P. Diaconis, Estimating and understanding exponential random graph models. *Ann. Stat.* **41**, 2428-2461 (2013).

74. M. E. J. N. Juyong Park, Solution for the properties of a clustered network. *Phy. Rev. E* **72**, 026136 (2005).

75. S. Horvát, É. Czabarka, Z. Toroczkai, Reducing degeneracy in maximum entropy models of networks. *Phys. Rev. Lett.* **114**, 158701 (2015).

76. T. A. Snijders, P. E. Pattison, G. L. Robins, M. S. Handcock, New specifications for exponential random graph models. *Sociol. Method.* **36**, 99-153 (2006).

77. A. Caimo, N. Friel, Bayesian inference for exponential random graph models. *Soc. Netw.* **33**, 41-55 (2011).

78. A. R. Benson, D. F. Gleich, J. Leskovec, Higher-order organization of complex networks. *Science* **353**, 163-166 (2016).

79. T. Thiele, The general theory of observations. *Repitzel, Copenhagen* (1889).

80. R. Kubo, Generalized cumulant expansion method. *J. Phys. Soc. Jpn.* **17**, 1100-1120 (1962).

81. T. N. Kipf, M. Welling, "Semi-supervised classification with graph convolutional networks", in *ICRL'17 Proceedings of the 7th International Conference on Learning Representations* (2017).

82. W. L. Hamilton, R. Ying, J. Leskovec, Representation learning on graphs: Methods and applications. *IEEE Data Eng. Bull.* **40**, 52-74 (2017).

83. W. Hamilton, Z. Ying, J. Leskovec, "Inductive representation learning on large graphs", in *NeurIPS'17 Proceedings of the 31st International Conference on Neural Information Processing Systems* (2017).

84. D. Liben-Nowell, J. Kleinberg, The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Tec.* **58**, 1019-1031 (2007).

85. G. Csardi, T. Nepusz, *et al.*, The igraph software package for complex network research. *InterJournal* **1695**, 1-9 (2006).

86. I. Wolfram Research, *Mathematica* (Wolfram Research, Inc., Champaign, Illinois, 2019).

87. M. Danisch, O. Balalau, M. Sozio, "Listing k-cliques in sparse real-world graphs", in *WWW'18 Proceedings of the 27th World Wide Web Conference* (2018).

88. N. Chiba, T. Nishizeki, Arboricity and subgraph listing algorithms. *SIAM J. Comput.* **14**, 210-223 (1985).

89. M. Kuba, A. Panholzer, On the degree distribution of the nodes in increasing trees. *J. Combin. Theory A* **114**, 597-618 (2007).

90. M. Houbraken, *et al.*, The index-based subgraph matching algorithm with general symmetries (ismags): exploiting symmetry for faster subgraph enumeration. *PloS one* **9**, e97896 (2014).

91. S. Wernicke, Efficient detection of network motifs. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **3**, 347-359 (2006).

92. G. M. Slota, K. Madduri, "Fast approximate subgraph counting and enumeration", in *ICPP'13 Proceedings of the 42nd International Conference on Parallel Processing* (2013).

93. P. Ribeiro, F. Silva, L. Lopes, "Efficient parallel subgraph counting using g-tries", in *Proceedings of the 11th IEEE International Conference on Cluster Computing* (2010).

94. M. Aliakbarpour, *et al.*, Sublinear-time algorithms for counting star subgraphs via edge sampling. *Algorithmica* **80**, 668-697 (2018).

95. K. Miyajima, T. Sakuragawa, Continuous and robust clustering coefficients for weighted and directed networks. *arXiv:1412.0059* (2014).

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

97. A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA* **101**, 3747-3752 (2004).

98. I. E. Antoniou, E. T. Tsompa, Statistical analysis of weighted networks. *Discrete Dyn. Nat. Soc.* **2008**, 375452 (2008).

99. G. Robins, M. Alexander, Small worlds among interlocking directors: Network structure and distance in bipartite graphs. *Comput. Math. Organ. Theory* **10**, 69-94 (2004).

100. J. C. Brunson, Triadic analysis of affiliation networks **3**, 480-508 (2015).

101. M. C. G. Pedro G. Lind, H. J. Herrmann, Cycles and clustering in bipartite networks. *Phys. Rev. E* **72**, 056127 (2005).

102. N. Bhardwaj, K.-K. Yan, M. B. Gerstein, Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels. *Proc. Natl. Acad. Sci. USA* **107**, 6841-6846 (2010).

103. https://oeis.org/A000088, *The on-line encyclopedia of integer sequences (OEIS), entry A000088* (founded by Neil Sloane in 1964).

104. M. Nouri, S. Talatahari, A. S. Shamloo, Graph products and its applications in mathematical formulation of structures. *J. Appl. Math.* **2012**, 510180 (2012).

105. X. Gao, B. Xiao, D. Tao, X. Li, A survey of graph edit distance. *Pattern Anal. Appl.* **13**, 113-129 (2010).

106. T. Opsahl, P. Panzarasa, Clustering in weighted networks. *Soc. Netw.* **31**, 155-163 (2009).

107. P. Zhang, *et al.*, Clustering coefficient and community structure of bipartite networks. *Phys. A.* **387**, 6869-6875 (2008).

108. C. Borgs, *et al.*, "Graph limits and parameter testing", in *Proceedings of the 38th annual ACM symposium on Theory of Computing* (2006).

Supplementary Materials

S1   Python module for computing graph cumulants

We provide a python module that computes graph moments of networks, including networks with directed edges, edge weights, and binary node attributes. Conversions to and from graph cumulants (as well as their unbiased counterparts) are also implemented.

We use the python package igraph (*85*) to count instances of subgraphs in a network. In conjunction with the symbolic computation available in Mathematica (*86*), we automatically derived the expressions for the counts of disconnected subgraphs in terms of the connected counts (see supplementary materials S1.1), as well as the expressions for converting graph moments to graph cumulants (see supplementary materials S10).

Our code contains the expressions to obtain graph cumulants up to: sixth order for undirected unweighted networks, fifth order for undirected weighted networks (plus a single K_4 at sixth order), fifth order for directed unweighted networks (plus a single K_3 at sixth order), subgraphs of $K_{2,2}$ for bipartite networks, and subgraphs of K_3 for networks with binary node attributes. Expressions for the unbiased graph cumulants are implemented up to third order for undirected unweighted networks (see supplementary materials S3).

S1.1   Efficiently computing graph moments

The scalability of our framework is determined by the computational time required to count the relevant connected subgraphs. This is because the counts of the disconnected subgraphs can be inferred by the counts of the connected subgraphs, and therefore do not need to be explicitly enumerated. This is an important fact, as the counts of the disconnected subgraphs are generally orders of magnitude larger than those of the connected ones. Moreover, counting connected subgraphs is an active area of research, and much work has gone into their efficient

computation (*41, 42, 43, 44*).

To illustrate how the counts of disconnected subgraphs are derivable from the connected subgraph counts, consider the case of second-order moments for simple graphs. From first order, one has the counts of edges in the network, $c_/ = \binom{n}{2}\mu_{1/}$. Consider all unordered pairs of distinct edges; each pair corresponds to a single second-order count: either of a wedge, or of two edges that do not share any node, thus $\binom{c_/}{2} = c_\wedge + c_{/\!/}$. Hence, the count of two edges that do not share any node $c_{/\!/}$ is directly derivable from the count of edges $c_/$ and the count of wedges $c_\wedge$.

A similar argument applies to all orders. For instance, at third order, there are two disconnected subgraphs: three edges that do not share any node, and a wedge and an edge that do not share any node. By enumerating all triplets of distinct edges, as well as all pairs of a wedge and an edge not contained in that wedge, we obtain the following expressions:

$$\binom{c_/}{3} = c_{/\!/\!/} + c_\triangle + c_\curlywedge + c_\sqcap + c_{/\!\wedge},$$

$$c_\wedge(c_/ - 2) = c_{/\!\wedge} + 3c_\triangle + 3c_\curlywedge + 2c_\sqcap.$$

We now discuss the scalability of counting the instances of a connected subgraph g with n' nodes in a network G with m edges and n nodes. The complexity of a naïve enumeration of the $\frac{n!}{(n-n')!}$ potential node mappings scales as $\mathcal{O}(n^{n'})$. However, there exist notably more efficient algorithms for certain subgraphs (such as triangles, stars, and cliques), especially when G has particular properties, such as sparsity (*78, 87, 90*). For example, the worst-case computational time complexity for counting n'-cliques is known to be at most $\mathcal{O}(n'm^{n'/2})$ time (*88*). The counts of the r-stars can be quickly computed, as they are proportional to the rth factorial moments of the degree distribution (*89, 94*). Moreover, some of these algorithms can be substantially accelerated through parallel computation (*92, 93*) and approximate values can be obtained by stochastic methods (*91*). Asymptotics aside, from a pragmatic perspective, Pinar et al. (*44*)

showed that the exact counts of all connected subgraphs with up to 5 nodes can be obtained for networks with tens of millions of edges in minutes on a commodity machine (64GB memory).

S2 Applications to networks with additional features

In this section, we first demonstrate how our framework provides a natural notion of clustering in bipartite networks. We then illustrate the utility of incorporating additional network features by analyzing real networks with directed edges, node attributes, and weighted edges.

S2.1 Quantifying clustering

Quantifying clustering in networks with additional features is an active domain of research that has arguably not reached a consensus. For example, there have been multiple proposals for weighted (*106, 97, 98*), directed (*95, 96*) and bipartite networks (*99, 107, 100, 101*). For all cases, our framework provides a principled measure of clustering, viz., the relevant scaled graph cumulant. For example, for directed networks, the two third-order scaled triangle cumulants provide two measures of clustering: one with cyclic orientation $\tilde{\kappa}_{3\triangle}$ and one with transitive $\tilde{\kappa}_{3\triangle}$. For bipartite networks, extensions are somewhat less straightforward, as triangles are now excluded. Several proposed measures consider the appearance of 4-cycles, similarly compared to the number of incomplete cycles. In fig. S1, we compare the scaled graph cumulant of the 4-cycle subgraph $\tilde{\kappa}_{4\square}$ with the clustering coefficient proposed by (*99*), expressed in our framework as $C_{\square} = \mu_{4\square}/\mu_{3\sqcap}$. Again our measure is more directly sensitive to the propensity for clustering.
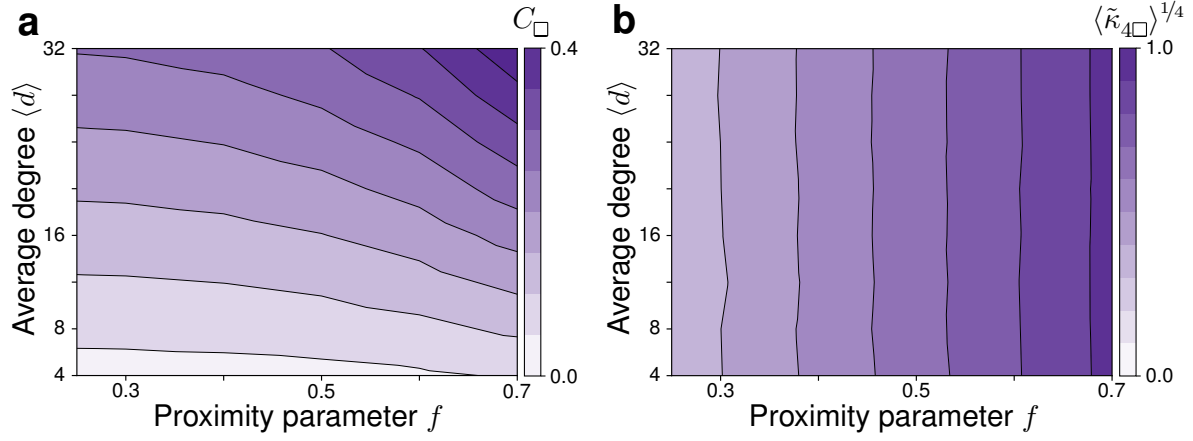
Fig. S1: A principled measure of clustering in bipartite networks. We simulated a bipartite geometric graph model with two parameters, f and $\langle d \rangle$, which determine the propensity for clustering (horizontal axis) and the edge density (vertical axis), respectively. The model first divides the nodes (here, 256) into two groups of equal size and randomly places them on the unit sphere. Each node may only connect to nodes from the other group, and only when they are within a certain radius, such that the area it contains is a fraction $1 - f$ of the entire unit sphere. Among these possible connections, a random subset is chosen, so as to match the desired average degree $\langle d \rangle$. For each set of parameters, we compute the average of both the global bipartite clustering coefficient $C_\square$ from (99) (left) and the scaled square cumulant for bipartite networks $\tilde{\kappa}_{4\square}$ (right), displaying the signed fourth root for $\tilde{\kappa}_{4\square}$. While both clustering measures increase with f, the bipartite clustering coefficient $C_\square$ also notably increases with average degree, whereas $\tilde{\kappa}_{4\square}$ is insensitive to such changes in edge density.

S2.2 Networks with node attributes
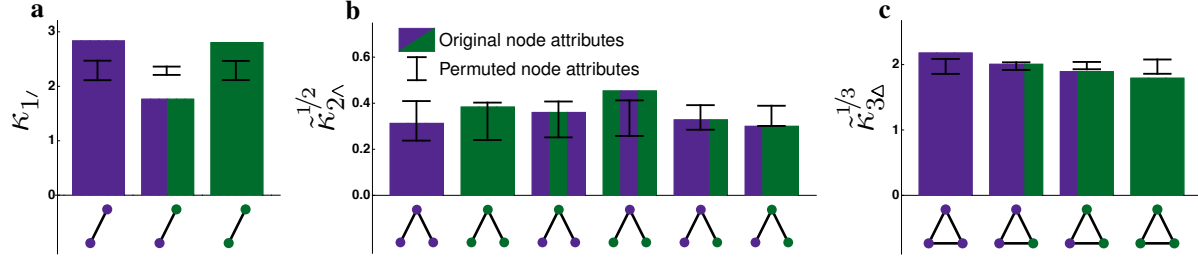
Fig. S2: Including node attributes reveals additional structure. We use the scaled graph cumulants for weighted networks with a binary node attribute to analyze a social network (222 nodes and 5364 edges) of interactions between primary school students from (*47*), with edge weights proportional to the number of interactions between pairs of students and node attributes corresponding to their (binary) sex. Purple nodes indicate female students, and green nodes indicate male students. Colored bars denote the value of (scaled) graph cumulant in the original network, and error bars denote the mean plus or minus one standard deviation for randomly shuffled node attributes (64 runs). **a)** The three first-order cumulants. **b)** The six second-order scaled wedge cumulants. **c)** The four third-order scaled triangle cumulants. The first-order graph cumulants (i.e., the density of edges between nodes of the indicated type) reveal a symmetric preference for homophily between the sexes, an effect well-documented in the social science literature (*4*). As all second-order scaled wedge cumulants are positive, we can infer a preference for hubs (i.e., nodes with degree notably larger than the average). Likewise, as all third-order scaled triangle cumulants are notably positive, we can infer a preference for triadic closure, a feature also present in many social networks (*13, 60*). While there does not appear to be much of a difference between the sexes at second order, the third-order scaled cumulants suggest that triadic closure is more prevalent when more participants are female.

S2.3 Networks with directed edges
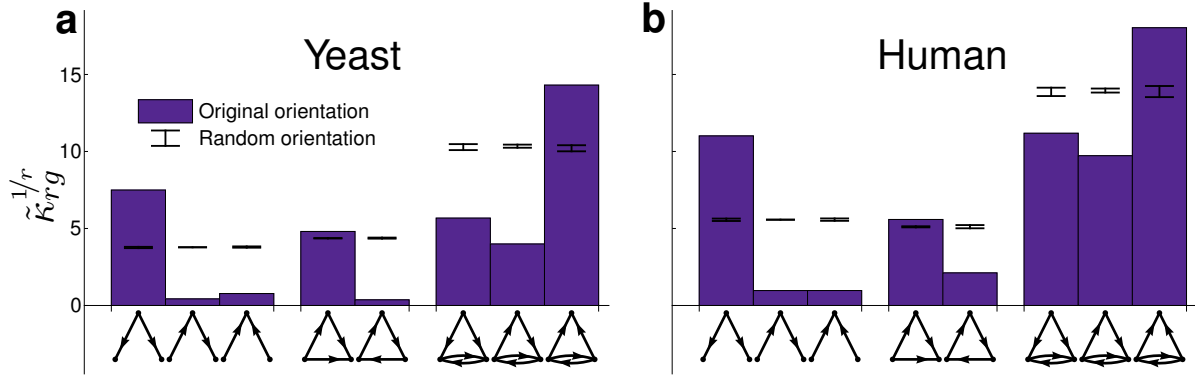
Fig. S3: The graph cumulant formalism naturally incorporates directed structure. We use the scaled cumulants for directed networks to analyze regulatory networks of protein interactions from (*102*) of: **a**) yeast (4441 nodes and 12873 edges), and **b**) humans (3197 nodes and 6896 edges). Error bars denote the mean and one standard deviation for the same network with randomized edge orientations. Despite the marked phenotypical differences between these two species, their regulatory networks display notable similarities. In particular, of the triangular substructures ($\tilde{\kappa}_{3\triangle}$), the feedforward (or transitive) substructure ($\tilde{\kappa}_{3\triangle}$) is significantly more prevalent than the cyclic ($\tilde{\kappa}_{3\triangle}$). Interestingly, while there is a higher prevalence of a central protein regulating many others ($\tilde{\kappa}_{2\wedge}$), proteins that regulate each other display a propensity to both regulate the same other protein ($\tilde{\kappa}_{4\triangle}$).

S2.4 Networks with weighted edges
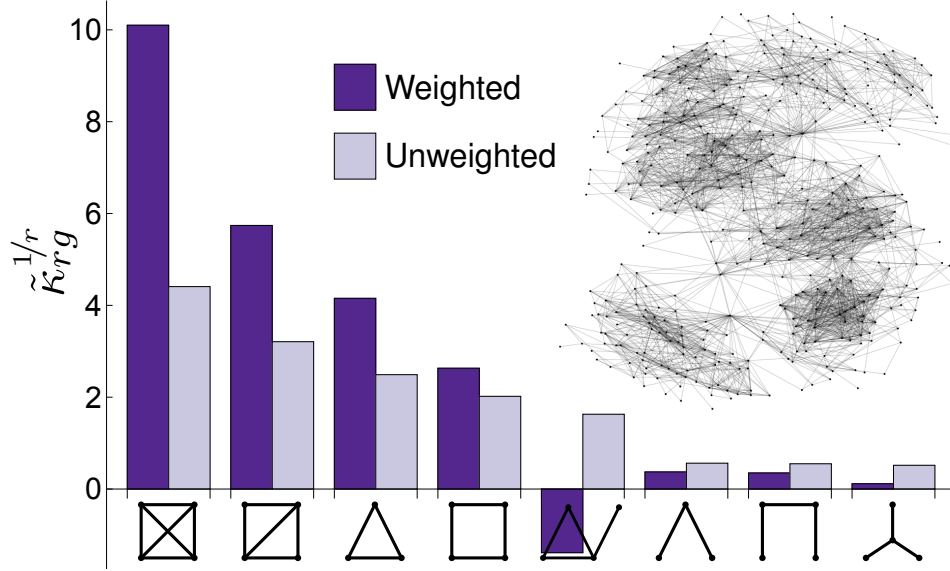
Fig. S4: Incorporating edge weights can increase the signal and possibly change the resulting interpretations. We use the scaled graph cumulants for weighted networks to analyze a social network (410 nodes and 2765 edges) of face-to-face interactions during an exhibition on infectious diseases from (*50*), with edge weights proportional to the total time a pair of people spent interacting. The scaled cumulants associated with clustering (i.e., $\tilde{\kappa}_{3\triangle}$ and $\tilde{\kappa}_{6\boxtimes}$) increase when using the true edge weights (dark purple) as opposed to the using the corresponding unweighted network (light purple). Conversely, many of the others become smaller, in particular, $\tilde{\kappa}_{2\wedge}$ and $\tilde{\kappa}_{3\curlywedge}$, both of which are associated with power-law properties of the degree distribution. The most notable deviation occurs for $\tilde{\kappa}_{4\triangle\!\!\!\vee}$, which is positive for the unweighted network, but negative for the weighted network. This negative cumulant could be interpreted as an anticorrelation between participating in triadic closure and interacting with many others, not unreasonable for such an exhibition: hosts are likely to talk with many different people, whereas groups of visitors tend to interact amongst themselves.

S3   Unbiased graph cumulants

In this section, we discuss the desired properties of the unbiased estimators $\hat{\kappa}_{rg}$ of graph cumulants, focusing our discussion on simple graphs. We describe the procedure for deriving

these unbiased graph cumulants $\hat{\kappa}_{rg}$, using the two second-order cumulants as an example, and provide the expressions up to third order.

In the spirit of k-statistics (*52*), we consider a large network G_N with N nodes (the "population"), randomly select a subset of n nodes from G_N, and observe the induced subgraph G (the "sample"). We require the expectations of the unbiased graph cumulants $\hat{\kappa}_{rg}$ to be invariant under this node subsampling (i.e., $\langle \hat{\kappa}_{rg}(G) \rangle = \hat{\kappa}_{rg}(G_N)$), and to have the appropriate limit (i.e., $\hat{\kappa}_{rg}(G_N) \to \kappa_{rg}(G_N)$ as $N \to \infty$). Moreover, suppose one generates a maximum entropy distribution with graph cumulants $\kappa_{rg}(\mathcal{G})$ equal to these $\hat{\kappa}_{rg}(G)$ up to some choice of order r' (i.e., an ERGM of order r'). Consider sampling single networks from this distribution and computing their $\hat{\kappa}_{rg}$. For $r \leq r'$, the expectations $\langle \hat{\kappa}_{rg} \rangle$ are equal to the cumulants κ_{rg} of the distribution itself. Thus, the $\hat{\kappa}_{rg}(G)$ provide unbiased estimators for the graph cumulants of the ERGM distribution $\mathcal{G}$ from which G was sampled.

As for real-valued random variables, the unbiased cumulants are polynomials in the moments, with coefficients that depend on the sample size n. We first consider how the relevant graph moments and their products change when removing a *single* random node from a network G. By requiring that the expectations of the expressions for the unbiased cumulants are unchanged, we obtain recursion relations for the coefficients of the polynomials as a function of n. We solve these recursion relations using Mathematica (*86*), and fix the undetermined constants such that the $n \to \infty$ limit agrees with the original combinatorial expressions for κ_{rg} (equation 2 and fig. 2).

As is the case with classical cumulants, the first-order unbiased graph cumulant $\hat{\kappa}_{1/}$ is simply $\hat{\kappa}_{1/} = \kappa_{1/} = \mu_{1/}$. Here, we illustrate our procedure applied to the second-order graph cumulants. The general form of a second-order polynomial in the moments is given by

$$\hat{\kappa}_{2g} = A(n)\mu_{2\wedge} + B(n)\mu_{1/}^2 + C(n)\mu_{1/}. \tag{3}$$

Note that it is not necessary to include a term for $\mu_{2/\!/}$, as it can be expressed as a function of the other terms (see supplementary materials S1.1).

We now describe how to determine the expectation of each of the terms ($\mu_{2\wedge}$, $\mu_{1/}^2$, and $\mu_{1/}$) when removing a single random node from G. The easiest term is $\mu_{1/}$. Let $c_{/}$ be the counts of edges in the original network on n nodes (as usual), d_i be the degree of node i, and $c'_{/}$ be the counts of edges in the network with node i removed. Clearly, $c'_{/} = c_{/} - d_i$, so $\langle c'_{/}\rangle = c_{/} - \langle d_i\rangle$. As $\langle d_i\rangle = \frac{2c_{/}}{n}$, we obtain $\langle c'_{/}\rangle = (1 - \frac{2}{n})c_{/}$. Dividing by the edge counts in the corresponding complete networks, we have $\langle \mu'_{1/}\rangle = \frac{\binom{n}{2}}{\binom{n-1}{2}}(1 - \frac{2}{n})\mu_{1/} = \mu_{1/}$. In fact, this is general: all graph moments are preserved in expectation under subsampling of the nodes,

$$\langle \mu'_{rg}\rangle = \mu_{rg}. \tag{4}$$

Products of moments, however, are generally not preserved in expectation. For instance, consider $\mu_{1/}^2$. The expected squared count of edges in the subsampled network is

$$\langle c'^2_{/}\rangle = \langle (c_{/} - d_i)^2\rangle = c_{/}^2 - 2c_{/}\langle d_i\rangle + \langle d_i^2\rangle.$$

Fortunately, $\langle d_i^2\rangle$ is easily expressed in terms of the wedge moment:

$$c_\wedge = \sum_i \binom{d_i}{2}$$
$$= \frac{n}{2}\left(\langle d_i^2\rangle - \langle d_i\rangle\right),$$

thus, $\langle d_i^2\rangle = \frac{2}{n}c_\wedge + \langle d_i\rangle$ and $\langle c'^2_{/}\rangle = \frac{2}{n}c_\wedge + (1 - \frac{4}{n})c_{/}^2 + \frac{2}{n}c_{/}$. Dividing by the appropriate counts in the corresponding complete networks yields

$$\langle \mu'^2_{1/}\rangle = \frac{4}{(n-1)(n-2)}\mu_{2\wedge} + \frac{n(n-4)}{(n-2)^2}\mu_{1/}^2 + \frac{4}{(n-1)(n-2)^2}\mu_{1/}.$$

39

Therefore, the resulting recursion relations for $A(n)$, $B(n)$, and $C(n)$ are:

$$A(n) = A(n-1) + \frac{4}{(n-1)(n-2)} B(n-1), \tag{5}$$

$$B(n) = \frac{n(n-4)}{(n-2)^2} B(n-1), \tag{6}$$

$$C(n) = C(n-1) + \frac{4}{(n-1)(n-2)^2} B(n-1). \tag{7}$$

The solution to equation 6 is

$$B(n) = b \frac{n(n-1)}{(n-2)(n-3)}, \tag{8}$$

and substituting this into equations 5 and 7 results in

$$A(n) = b \frac{4(n-4)}{(n-3)} + a, \tag{9}$$

$$C(n) = b \frac{(n-1)(n-4)}{(n-2)(n-3)} + c, \tag{10}$$

where a, b, and c are constants to be determined.

To obtain $\hat{\kappa}_{2\wedge}$, we require that the $n \to \infty$ limit agrees with $\kappa_{2\wedge}$, viz.,

$$A_\wedge \xrightarrow[n \to \infty]{} 1, \quad B_\wedge \xrightarrow[n \to \infty]{} -1, \quad C_\wedge \xrightarrow[n \to \infty]{} 0.$$

Thus, the relevant constants are $a_\wedge = 5$, $b_\wedge = -1$, and $c_\wedge = 1$, so $\hat{\kappa}_{2\wedge}$ is given by

$$\hat{\kappa}_{2\wedge} = \frac{n+1}{n-3} \mu_{2\wedge} - \frac{n(n-1)}{(n-2)(n-3)} \mu_{1/}^2 + \frac{2}{(n-2)(n-3)} \mu_{1/}. \tag{11}$$

Recalling that $c_\wedge + c_\# = \binom{c}{2}$, we repeat the same procedure to obtain $\hat{\kappa}_{2\#}$. This case is a bit less straightforward (although the result, $\hat{\kappa}_{2\#} = 0$, is comparatively simpler). As $n \to \infty$, the number of pairs of edges in the network grows as $\binom{c}{2} = \frac{n^4}{8} \mu_{1/}^2$, while the number of wedges is bounded by $c_\wedge < \frac{n^3}{2}$. Therefore, the number of pairs of edges that do not share any node is $c_\# = \frac{n^4}{8} \mu_{1/}^2 (1 - \mathcal{O}(\frac{1}{n}))$. Thus, $\mu_{2\#} = \mu_{1/}^2 + \mathcal{O}(\frac{1}{n})$, and $\kappa_{2\#} = \mathcal{O}(\frac{1}{n}) \to 0$. Hence, the (unique) coherent solution is $a_\# = b_\# = c_\# = 0$, so $\hat{\kappa}_{2\#} = 0$. Indeed, the same result holds for graphons

40

(the natural limit of a sequence of dense graphs of increasing size (*108*)), where it can be shown (*24*) that $\mu_{2/\!\!/} \xrightarrow[n\to\infty]{} \mu_{1/}^2$, and therefore $\kappa_{2/\!\!/} \xrightarrow[n\to\infty]{} 0$. More generally, for a single network observation, $\hat{\kappa}_{rg} = 0$ for all disconnected subgraphs g.

To summarize, as with a real-valued random variable, the unbiased estimator of the first-order cumulant is simply the first-order cumulant itself,

$$\hat{\kappa}_{1/} = \mu_{1/}, \tag{12}$$

the unbiased estimators of the second-order cumulants are

$$\hat{\kappa}_{2\wedge} = \frac{n+1}{n-3}\mu_{2\wedge} - \frac{n(n-1)}{(n-2)(n-3)}\mu_{1/}^2 + \frac{2}{(n-2)(n-3)}\mu_{1/}, \tag{13}$$

$$\hat{\kappa}_{2/\!\!/} = 0. \tag{14}$$

Applying the same procedure to the third-order graph cumulants yields

$$\hat{\kappa}_{3\triangle} = \alpha_{1/}\mu_{1/} + \alpha_{1/1/}\mu_{1/}^2 + \alpha_{2\wedge}\mu_{2\wedge} + \alpha_{1/1/1/}\mu_{1/}^3 + \alpha_{2\wedge 1/}\mu_{2\wedge}\mu_{1/}$$
$$+ (1 + \alpha_{3\triangle})\mu_{3\triangle} + \alpha_{3\lambda}\mu_{3\lambda} + \alpha_{3\sqcap}\mu_{3\sqcap}, \tag{15}$$

$$\hat{\kappa}_{3\lambda} = \alpha_{1/}\mu_{1/} + \alpha_{1/1/}\mu_{1/}^2 + \alpha_{2\wedge}\mu_{2\wedge} + \alpha_{1/1/1/}\mu_{1/}^3 + \alpha_{2\wedge 1/}\mu_{2\wedge}\mu_{1/}$$
$$+ \alpha_{3\triangle}\mu_{3\triangle} + (1 + \alpha_{3\lambda})\mu_{3\lambda} + \alpha_{3\sqcap}\mu_{3\sqcap}, \tag{16}$$

$$\hat{\kappa}_{3\sqcap} = \alpha_{1/}\mu_{1/} + \alpha_{1/1/}\mu_{1/}^2 + \alpha_{2\wedge}\mu_{2\wedge} + \alpha_{1/1/1/}\mu_{1/}^3 + \alpha_{2\wedge 1/}\mu_{2\wedge}\mu_{1/}$$
$$+ \alpha_{3\triangle}\mu_{3\triangle} + \alpha_{3\lambda}\mu_{3\lambda} + (1 + \alpha_{3\sqcap})\mu_{3\sqcap}, \tag{17}$$

$$\hat{\kappa}_{3/\!\!\wedge} = 0, \tag{18}$$

$$\hat{\kappa}_{3/\!\!/\!\!/} = 0, \tag{19}$$

41

where

$$\alpha_{1/} = \frac{16}{(n-2)(n-3)(n-4)(n-5)}, \tag{20}$$

$$\alpha_{1/1/} = \frac{-12n(n-1)}{(n-2)(n-3)(n-4)(n-5)}, \tag{21}$$

$$\alpha_{2\wedge} = \frac{12(n+3)}{(n-3)(n-4)(n-5)}, \tag{22}$$

$$\alpha_{1/1/1/} = \frac{2n^2(n-1)^2}{(n-2)(n-3)(n-4)(n-5)}, \tag{23}$$

$$\alpha_{2\wedge 1/} = \frac{-3(n+3)n(n-1)}{(n-3)(n-4)(n-5)}, \tag{24}$$

$$\alpha_{3\triangle} = \frac{6n+2}{(n-3)(n-4)(n-5)}, \tag{25}$$

$$\alpha_{3\curlywedge} = \frac{6n+2}{(n-4)(n-5)}, \tag{26}$$

$$\alpha_{3\sqcap} = \frac{12(n-1)}{(n-4)(n-5)}. \tag{27}$$

These expressions may be inverted by exploiting the expressions relating the disconnected counts to those of the connected (supplementary materials S1.1), e.g.,

$$\hat{\mu}_{1/} = \hat{\kappa}_{1/}, \tag{28}$$

$$\hat{\mu}_{2\wedge} = \frac{n-3}{n+1}\hat{\kappa}_{2\wedge} + \frac{n(n-1)}{(n+1)(n-2)}\hat{\kappa}_{1/}^2 - \frac{2}{(n+1)(n-2)}\hat{\kappa}_{1/}, \tag{29}$$

$$\hat{\mu}_{2//} = \frac{1}{\#_{//}}\left(\binom{\#_{/}\hat{\mu}_{1/}}{2} - \#_{\wedge}\hat{\mu}_{2\wedge}\right), \tag{30}$$

where $\#_g$ is the count of subgraph g in the complete network with n nodes (e.g., the denominators of equations 60–71 in supplementary materials S10).

S4   Fitting ERGMs using unbiased graph cumulants

We now describe how to infer a model from our proposed hierarchical family of ERGMs using a *single* observed network G, again specializing our discussion to simple graphs. In particular, we

consider ERGMs with prescribed expected graph moments of at most order r (or, equivalently, the associated subgraph counts, as the number of nodes n is fixed). These distributions have the following form:

$$p(G) = \frac{1}{Z} \, \text{ER}_{n,1/2}(G) \, \exp\left(\sum_g \beta_g c_g(G) \right), \tag{31}$$

$$Z = \sum_{G' \in \Omega} \left[\text{ER}_{n,1/2}(G') \, \exp\left(\sum_g \beta_g c_g(G') \right) \right], \tag{32}$$

where $\text{ER}_{n,p}(G)$ is the probability of the network G in an Erdős–Rényi random graph model (i.e., the presence of an edge between any pair of nodes occurs independently with the given probability p)[3]; β_g is the parameter (or Lagrange multiplier) associated with subgraph g; $c_g(G)$ is the count of subgraph g in the network G; Ω is the space of all simple graphs with n nodes; and Z is the normalization constant (or partition function).

While an ERGM is typically specified by the desired expectations of the statistics of interest (in this case, subgraph counts/moments), the parameters needed to compute the distribution are the β_g, and in general must be determined numerically. Moreover, as the number of unique graphs grows super-exponentially in the number of nodes (e.g., there are 12005168 simple graphs with 10 nodes (*103*)), the partition function Z often cannot be exactly computed. However, there exists a large body of literature on sampling and variational techniques for efficiently approximating Z as a function of β_g (*76*).

Given a single observed network, the protocol for inferring an ERGM from our hierarchical family is as follows:

1. Choose the order of the desired ERGM. Use the observed network to compute the unbiased estimators $\hat{\kappa}_{rg}$ of all graph cumulants up to and including this order (see

---

supplementary materials S3).

2. Substitute these unbiased cumulants into the combinatorial formula (equation 2) to obtain the desired unbiased graph moments $\hat{\mu}_{rg}$ for this ERGM.

3. Fit the parameters β_g such that the resulting ERGM distribution has expected graph moments equal to these $\hat{\mu}_{rg}$.

S4.1 Partially unbiased ERGMs

The expressions for the unbiased graph cumulants are derived by assuming that the nodes were sampled randomly from a much larger underlying network (the "population"). However, this assumption may not always be appropriate, such as when the observed network is small, or when it could feasibly represent a significant fraction of the system of interest. For these cases, we introduce an adjustable unbiasing parameter $\eta \in [0, 1]$, where 1 corresponds to the aforementioned "fully" unbiased case, and 0 corresponds to using the original moments μ_{rg} of the observed network ("no unbiasing"). Essentially, instead of assuming that the underlying network is infinite, η controls its size (N nodes) relative to that of the observed network (n nodes), and is defined as $\eta = 1 - \frac{n}{N}$.

The procedure is essentially the same as before, with a modified second step. In particular, one first computes the unbiased estimates of the cumulants using the moments μ of the observed network (step 1 of our procedure):

$$\mu \xrightarrow[\text{eq. 12–19 with } n = n]{} \hat{\kappa}. \tag{33}$$

Then, one uses these $\hat{\kappa}$ in the inverse expressions, now using $n = N$:

$$\hat{\kappa} \xrightarrow[\text{eq. 28–30 with } n = N]{} \hat{\mu}. \tag{34}$$

44

Finally, one fits the parameters β_g such that the resulting ERGM distribution has expected graph moments equal to these (partially unbiased) graph moments $\hat{\mu}_{rg}$ (step 3 of our procedure).

S5   A geometric understanding of the degeneracy problem

The degeneracy problem refers to the appearance of undesirable large-scale multimodality in the distribution induced by an ERGM; despite the fact that averaging over this distribution gives expected subgraph counts equal to those of the observed network (as desired), typical samples from it have counts vastly different from these average values.

Essentially, this arises due to the shape of the base distribution (i.e., $\mathrm{ER}_{n,1/2}$) as a function of the statistics whose expected values are constrained (75) (here the subgraph counts, or equivalently, the corresponding graph moments). Recall from equation 31 that these ERGM distributions have the following form:

$$p(G) \propto \mathrm{ER}_{n,1/2}(G) \; \exp\!\left(\sum_g \beta_g c_g(G)\right). \tag{35}$$

Projecting this distribution to the space of the relevant subgraph counts (i.e., summing the probability of all networks for which these counts are the same), and taking its logarithm yields:

$$\ln p(\vec{c}) = \ln\!\big(\mathrm{ER}_{n,1/2}(\vec{c})\big) + \vec{\beta} \cdot \vec{c}, \tag{36}$$

where $\vec{c}$ is the vector of relevant subgraphs counts, $\vec{\beta}$ is the vector of their associated parameters, and we have dropped the term associated with the partition function (as it does not depend on $\vec{c}$). Thus, to understand the behavior of $p(G)$, it is geometrically instructive to look at the shape of $\ln\!\big(\mathrm{ER}_{n,1/2}\big)$ as a function of $\vec{c}$.

To provide intuition about the degeneracy problem and our proposed solution, here we give attention to a commonly used (and easily visualizable) 2D model, denoted by $\mathrm{ERGM}(\mu_{1/}, \mu_{2\wedge})$,

which prescribes the expected counts of edges and wedges in the distribution to be equal to those of the observed network. For comparison, we consider our second-order ERGM, which additionally prescribes the expected counts of pairs of edges that do not share any node. We will discuss both the case when the expectations of these three subgraph counts are prescribed to be those of the observed network, denoted by $\text{ERGM}(\mu_{1/}, \mu_{2\wedge}, \mu_{2\sslash})$, and when they are prescribed to be equal to the unbiased values (see supplementary materials S4), denoted by $\text{ERGM}(\hat{\mu}_{1/}, \hat{\mu}_{2\wedge}, \hat{\mu}_{2\sslash})$.

Consider all tuples representing realizable subgraph counts of a single network in their respective 2D (for $\text{ERGM}(\mu_{1/}, \mu_{2\wedge})$) or 3D (for $\text{ERGM}(\mu_{1/}, \mu_{2\wedge}, \mu_{2\sslash})$ and $\text{ERGM}(\hat{\mu}_{1/}, \hat{\mu}_{2\wedge}, \hat{\mu}_{2\sslash})$) spaces (fig. S5). Any point within the convex hull formed by these points may serve as the prescribed expected values of some ERGM. However, some of these choices *require* degenerate distributions. For example, consider $\langle c_/ \rangle = \frac{1}{2}\#_/$, $\langle c_\wedge \rangle = \frac{1}{2}\#_\wedge$ (where $\#_g$ is the count of subgraph g in the complete graph with n nodes). Indeed, the *only* distribution with these expected values is an equal mixture of the empty and complete networks — in a sense, the most "degenerate" distribution possible!

Even if one restricts attention to tuples of subgraph counts that are realizable by a single network, $\text{ERGM}(\mu_{1/}, \mu_{2\wedge})$ still does not always concentrate around these values. In particular, this occurs when one chooses a network that lies along the concave boundary of the support of $\ln(\text{ER}_{n, 1/2}(c_/, c_\wedge))$ (i.e., the region in fig. S5b, where $c_\wedge$ is large for a given number of edges). This can be understood by considering equation 36: the $\vec{\beta}$ term (which serves to enforce the prescribed expected subgraph counts) is linear and essentially "pushes" on the distribution with the same direction and magnitude everywhere. Thus, increasing the expected counts of wedges is inevitably coupled with a motion of the probability density toward the "tips" of this crescent-shaped domain. Hence, the expected counts of wedges and the *spread* in the counts of edges cannot be independently controlled, and the distribution can become degenerate.

In contrast, the $\vec{\beta}$ in ERGM($\mu_{1/}, \mu_{2\wedge}, \mu_{2/\!/}$) has an additional degree of freedom. Thus, it is able to independently control the expected counts of edges and wedges, as well as the spread in the counts of edges. However, if one requires that the expected counts $(c_{/}, c_{\wedge}, c_{/\!/})$ are exactly equal to those of the observed network, the resulting distribution necessarily concentrates on networks with precisely this edge count. Essentially, this occurs because the triplet $(c_{/}, c_{\wedge}, c_{/\!/})$ of any individual network lies on the boundary of the convex hull formed by all such triplets. For a fixed number of edges, the relationship between the second-order moments is linear: $\#_{\wedge}\mu_{2\wedge} + \#_{/\!/}\mu_{2/\!/} = C$ (see fig. S5a). Additionally, the relationship between the counts of edges and this invariant sum C has a curvature that does not change sign: $C = \frac{1}{2}(\#_{/}^2\mu_{1/}^2 - \#_{/}\mu_{1/})$. Thus, any distribution with expected counts equal to those of an observed network must have support only in the linear direction given by the set of triplets with the same invariant sum C (and therefore the same number of edges). Thus, we have "solved" the degeneracy problem by essentially fixing the number of edges in the ERGM. However, such a solution is not satisfactory for many applications.

In order to obtain a non-degenerate distribution containing networks with different numbers of edges, the triplet of expected counts must be slightly in the interior of the convex hull, in the direction of the red arrow in figs. S5a and S6. The unbiased graph cumulants derived in supplementary materials S3 provide a natural and consistent prescription for obtaining such modified triplets of expected counts (and, more generally, modified tuples of expected counts for higher order ERGMs). While this may seem to be an unusual choice (as such tuples are not realizable by any individual network), it is indeed quite natural: even the $\mathrm{ER}_{n,p}$ distributions have tuples of expected counts that lie in this direction.
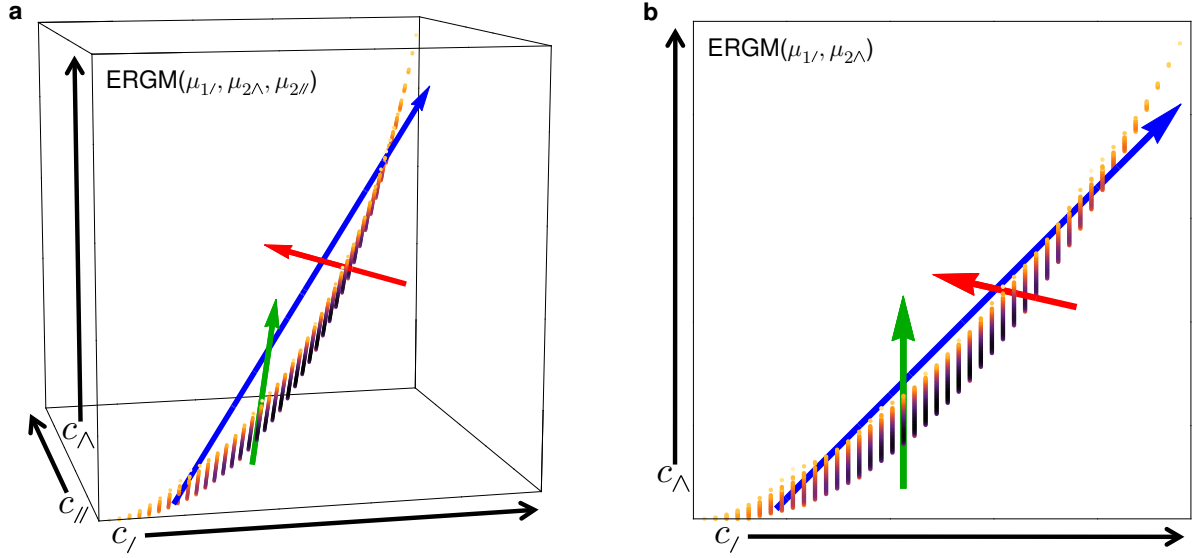
Fig. S5: Prescribing the expected counts of all subgraphs of first and second order allows for greater control over the resulting distribution. Here, we explicitly enumerate all graphs with 10 nodes. Each point corresponds to a tuple of subgraph counts, and the color corresponds to $\ln\big(\mathrm{ER}_{10,1/2}\big)$, with darker colors denoting higher probability. **a)** When represented as a function of $(c_{/}, c_{\wedge}, c_{/\!/})$, the density of $\mathrm{ER}_{10,1/2}$ lies on a 2D submanifold with extrinsic curvature. Three orthogonal directions can independently control the edge counts (blue arrow), the wedge counts (green arrow), and the spread in the edge counts (red arrow). **b)** In contrast, when representing the distribution as a function of $(c_{/}, c_{\wedge})$, these three quantities cannot be independently controlled. This can lead to the "degenerate" distributions with significant bimodality observed in certain $\mathrm{ERGM}(\mu_{1/}, \mu_{2\wedge})$.
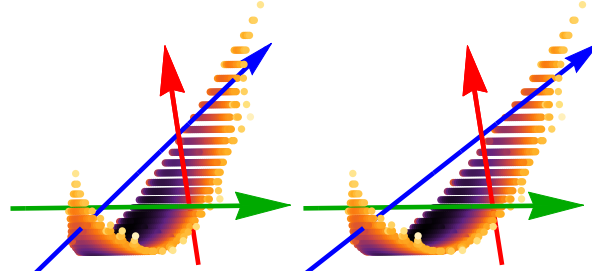
Fig. S6: A stereographic image of the base distribution for graphs with 10 nodes. Displayed is the logarithm of the base distribution $\ln(\mathrm{ER}_{10,1/2}(c_/, c_\wedge, c_{/\!/}))$ embedded in the space of all subgraph counts up to and including second order, i.e., edges, wedges, and two edges that do not share any node. Darker colors indicate higher probability. To fully appreciate the stereographic effect, print the image using either standard A4 paper or US letter size. Begin with the paper close to your eyes. Allow your left eye to focus on the left image, while your right eye focuses on the right image. Slowly move the paper away from your eyes, while maintaining focus on the middle of the "three" images, until it is approximately 20–40 cm (8–16 in) away. The "middle" image should be more apparent than those on either side, and its upper right corner should appear farther away than its bottom left. See here for an animation.

For a few extremal networks, our prescription for obtaining the modified expected counts may result in tuples that lie outside the convex hull, and thus do not lead to realizable ERGMs. This tends to occur for networks that are unlikely to be observed when subsampling nodes from a large network (such as regular or nearly-regular graphs). From a pragmatic perspective, this is unlikely to be an issue, as real networks tend not to have such properties. Moreover, if one does observe a network for which this is the case, the issue is often alleviated by using an intermediate choice of the unbiasing parameter η to obtain the modified tuples of expected counts (see supplementary materials S4.1).

S6 Statistical inference without constructing an explicit null model

In order to assess the statistical significance of a network's propensity for substructures, one needs to compare the observed cumulants with the distribution obtained from some appropriate null model. While our proposed hierarchical family of ERGMs is a principled option, unfortunately, obtaining the parameters $\vec{\beta}$ is often computationally prohibitive. Fortunately, our procedure to derive the unbiased graph cumulants $\hat{\kappa}_{rg}$ (supplementary materials S3) can also be used to derive their variance $\mathrm{Var}(\hat{\kappa}_{rg})$, allowing for statistical tests of a network's propensity for substructures *without* explicitly constructing a null model. We first explain how to perform such a statistical test, then we illustrate how to obtain the variance of the unbiased graph cumulants, using the derivation of $\mathrm{Var}(\hat{\kappa}_{1\prime})$ as an example.

S6.1 Statistical test using $\hat{\kappa}_{rg}$ and $\mathrm{Var}(\hat{\kappa}_{rg})$

To analyze a substructure g with r edges, first measure the moments of the observed network up to order $2r$, and use these to compute the unbiased cumulant $\hat{\kappa}_{rg}$, as well as its variance $\mathrm{Var}(\hat{\kappa}_{rg})$. If $\hat{\kappa}_{rg} \neq 0$, this potentially indicates a propensity (or aversiveness) for the substructure g, depending on the sign of $\hat{\kappa}_{rg}$. To determine if such an assessment is statistically significant, one should compute the (squared) Z-score associated with the null hypothesis that $\hat{\kappa}_{rg} \sim 0$: $Z_{rg}^2 = \frac{\hat{\kappa}_{rg}^2}{\mathrm{Var}(\hat{\kappa}_{rg})}$. If $Z_{rg}^2 \gg 1$, one can be reasonably confident that the observed network has a propensity (or aversiveness) for the substructure g. This procedure can also be used to measure the similarity between two networks, by applying a two-sample t-test to the pair of unbiased cumulants associated to each particular substructure.

The standard conversion from a Z-score to a p-value tacitly assumes normality, which does not hold in general. However, our procedure to obtain unbiased graph cumulants and the variance of their distributions can also be used to obtain higher-order cumulants of their

distributions (although the derivations become incredibly tedious). Future work on these unbiased cumulants (and the cumulants of their distributions) could lead to more principled statistical tests of the propensity (or aversiveness) for arbitrary substructures.

S6.2 Deriving $\text{Var}(\hat{\kappa}_{rg})$

To obtain the variance of the unbiased graph cumulants, one can apply a procedure similar to that described in supplementary materials S3. Here, we illustrate the derivation of the variance of the first-order unbiased cumulant $\hat{\kappa}_{1/} = \mu_{1/}$.

In general, the expressions for $\text{Var}(\hat{\kappa}_{rg})$ require moments up to order $2r$, a property analogous to that for real-valued random variables. Thus, for first order, we consider expressions of the form

$$\text{Var}(\mu_{1/}) = A(n)\mu_{2\wedge} + B(n)\mu_{1/}^2 + C(n)\mu_{1/}. \tag{37}$$

and again determine how the functions $A(n)$, $B(n)$, and $C(n)$ should change when removing a single random node. This provides a recursion relation, and determines these functions up to a few constants, which are now chosen so to give zero variance as $n \to \infty$.

Assume that we started with an initial network with $N \gg n$ nodes, and have subsampled it down to n nodes, thereby accruing a variance $\text{Var}(\mu_{1/})$ in the first moment. We consider the increase in variance due to removing another random node. The change in the first moment when removing node i is:

$$
\begin{aligned}
\mu_{1/} \to \mu'_{1/} &= \frac{1}{\binom{n-1}{2}} \left(\binom{n}{2}\mu_{1/} - d_i \right) \\
&= \mu_{1/} + \frac{2}{n-2}\mu_{1/} - \frac{2}{(n-1)(n-2)}d_i.
\end{aligned} \tag{38}
$$

Thus, $\mathrm{Var}(\mu'_{1/}) = \mathrm{Var}(\mu_{1/}) + \Delta\mathrm{Var}$, where

$$
\begin{aligned}
\Delta\mathrm{Var} &= \left\langle \left(\frac{2}{n-2}\mu_{1/} - \frac{2}{(n-1)(n-2)}d_i \right)^2 \right\rangle_{n \leftarrow N} \\
&= \frac{4}{(n-2)^2}\left\langle \mu_{1/}^2 \right\rangle_{n \leftarrow N} - \frac{8}{(n-1)(n-2)^2}\left\langle \mu_{1/}\langle d_i \rangle \right\rangle_{n \leftarrow N} \\
&\quad + \frac{4}{(n-1)^2(n-2)^2}\left\langle \langle d_i^2 \rangle \right\rangle_{n \leftarrow N},
\end{aligned}
\tag{39}
$$

where the angle brackets $\langle \cdot \rangle_{n \leftarrow N}$ denote the expectation over networks with n nodes obtained by randomly subsampling from the original N nodes, and angle brackets without a subscript denote expectation with respect to a single such network (i.e., randomly choosing a node i from these n nodes).

Starting with the last term of equation 39, recall from supplementary materials S3 that $\langle d_i \rangle = \frac{2c_/}{n}$ and $\langle d_i^2 \rangle = \frac{2c_\wedge}{n} + \langle d_i \rangle$. Thus, we have

$$
\begin{aligned}
\left\langle \langle d_i^2 \rangle \right\rangle_{n \leftarrow N} &= \left\langle \frac{2c_\wedge}{n} + \frac{2c_/}{n} \right\rangle_{n \leftarrow N} \\
&= (n-1)(n-2)\langle \mu_{2\wedge} \rangle_{n \leftarrow N} + (n-1)\langle \mu_{1/} \rangle_{n \leftarrow N} \\
&= (n-1)(n-2)\hat{\mu}_{2\wedge} + (n-1)\hat{\mu}_{1/}.
\end{aligned}
\tag{40}
$$

where $\hat{\mu}_{rg}$ denote the graph moments of the large network with N nodes (equivalent to the fully unbiased moments as $N \to \infty$, supplementary materials S4.1), and we have used the fact that graph moments are preserved in expectation under subsampling of nodes.

For the middle term of equation 39, as the nodes are removed randomly,

$$
\begin{aligned}
\left\langle \mu_{1/}\langle d_i \rangle \right\rangle_{n \leftarrow N} &= \left\langle \mu_{1/}\frac{2\binom{n}{2}\mu_{1/}}{n} \right\rangle_{n \leftarrow N} \\
&= (n-1)\left\langle \mu_{1/}^2 \right\rangle_{n \leftarrow N},
\end{aligned}
\tag{41}
$$

thus combining with the first term of equation 39, giving $-\frac{4}{(n-2)^2}\langle \mu_{1/}^2 \rangle_{n \leftarrow N}$.

Evaluating $\langle \mu_{1/}^2 \rangle_{n \leftarrow N}$ requires the inclusion of the current amount of variance:

$$
\begin{aligned}
\left\langle \mu_{1/}^2 \right\rangle_{n \leftarrow N} &= \left\langle \mu_{1/} \right\rangle_{n \leftarrow N}^2 + \text{Var}(\mu_{1/}) \\
&= \hat{\mu}_{1/}^2 + \text{Var}(\mu_{1/}).
\end{aligned}
\tag{42}
$$

Substituting these into the expression for $\text{Var}(\mu_{1/}')$ yields a recursion relation:

$$
\begin{aligned}
\text{Var}(\mu_{1/}') &= \text{Var}(\mu_{1/}) - \frac{4}{(n-2)^2} \left\langle \mu_{1/}^2 \right\rangle_{n \leftarrow N} + \frac{4}{(n-1)^2(n-2)^2} \left\langle \langle d_i^2 \rangle \right\rangle_{n \leftarrow N} \\
&= \text{Var}(\mu_{1/}) - \frac{4}{(n-2)^2} \left(\hat{\mu}_{1/}^2 + \text{Var}(\mu_{1/}) \right) \\
&\quad + \frac{4}{(n-1)^2(n-2)^2} \left((n-1)(n-2)\hat{\mu}_{2\wedge} + (n-1)\hat{\mu}_{1/} \right) \\
&= \frac{n(n-4)}{(n-2)^2} \text{Var}(\mu_{1/}) - \frac{4}{(n-2)^2} \hat{\mu}_{1/}^2 \\
&\quad + \frac{4}{(n-1)(n-2)} \hat{\mu}_{2\wedge} + \frac{4}{(n-1)(n-2)^2} \hat{\mu}_{1/}.
\end{aligned}
\tag{43}
$$

The solution to this recursion relation (along with the condition that $\text{Var}(\mu_{1/}) \to 0$ as $n \to \infty$) is given by

$$
\begin{aligned}
\text{Var}(\hat{\kappa}_{1/}) &= \frac{1}{\binom{n}{2}} \left(\hat{\mu}_{1/} + (3 - 2n)\hat{\mu}_{1/}^2 + 2(n-2)\hat{\mu}_{2\wedge} \right) \\
&= \frac{1}{\binom{n}{2}} \left(\mu_{1/}(1 - \mu_{1/}) + 2(n-2)\hat{\kappa}_{2\wedge} \right).
\end{aligned}
\tag{44}
$$

S7   Local graph cumulants

Graph cumulants are statistics of the *entire* network, quantifying its overall propensity for a given substructure. However, in some applications, such as node classification (*81, 82, 83*) and link prediction (*84*), one often desires statistics of the propensity of an *individual* node or edge to participate in a given substructure. The graph cumulant framework naturally incorporates both of these "local" cases. In this section, we describe how to derive these local graph moments and cumulants for both nodes and edges, providing the expressions necessary to compute both local triangle cumulants.

53

S7.1 Node local graph cumulants

The node local graph moments and cumulants are defined by giving a unique identity to the node of interest (here, symbolically distinguished by an empty circle), and applying the equations for general node attributes (see supplementary materials S10.3). For example, for simple graphs, there are now two first-order moments. One is defined as the count of edges between the distinguished node and any other node (i.e., the degree of the distinguished node), again normalized by the corresponding count in the associated complete graph:

$$\mu_{1\rho} = \frac{c_\rho}{n-1}.$$

The other first-order moment is defined as the count of edges that do not use the distinguished node, normalized by the corresponding count in the associated complete graph:

$$\mu_{1\underline{\circ}} = \frac{c_{\underline{\circ}}}{\binom{n-1}{2}}.$$

Likewise:

$$\mu_{2\wedge} = \frac{c_{\wedge}}{\binom{n-1}{2}},$$

$$\mu_{2\underline{\wedge}} = \frac{c_{\underline{\wedge}}}{2\binom{n-1}{2}},$$

$$\mu_{3\underline{\triangle}} = \frac{c_{\underline{\triangle}}}{\binom{n-1}{2}}.$$

The definition of node local graph cumulants follows the same procedure as before, now taking care to incorporate the presence of this distinguished node, e.g.,

$$\kappa_{3\underline{\triangle}} = \mu_{3\underline{\triangle}} - \mu_{2\wedge}\mu_{1\underline{\circ}} - 2\mu_{2\underline{\wedge}}\mu_{1\rho} + 2\mu_{1\rho}^2\mu_{1\underline{\circ}}. \tag{45}$$

The same care must be taken when scaling the node local graph cumulants, e.g.,

$$\tilde{\kappa}_{3\underline{\triangle}} = \frac{\kappa_{3\underline{\triangle}}}{\mu_{1\rho}^2\mu_{1\underline{\circ}}}. \tag{46}$$

S7.2 Edge local graph cumulants

A similar procedure can be used to obtain edge local graph cumulants, where instead of distinguishing a node, one now distinguishes an edge (here, represented by a four-pointed star at the midpoint of that edge). Again, there are two first-order edge local graph cumulants, although the one associated with the distinguished edge itself is trivial:

$$\mu_{1\chi} \equiv 1.$$

The other, associated with the remaining edges, is given by

$$\mu_{1\lambda} = \frac{c_{\lambda}}{\binom{n}{2} - 1},$$

where the floating star indicates that the distinguished edge is not included in the illustrated subgraph. In particular, as the nodes associated with the distinguished edge are equivalent to any other, they are neither required in nor excluded from the illustrated subgraph. Thus, $c_{\lambda} = c_{\ell} - 1$, i.e., the count of edges in the network minus the one distinguished edge.

Likewise,

$$\mu_{2\wedge} = \frac{c_{\wedge}}{2(n-2)},$$

$$\mu_{2\triangle} = \frac{c_{\triangle}}{3\binom{n}{3} - 2(n-2)},$$

$$\mu_{3\triangle} = \frac{c_{\triangle}}{n-2}.$$

Again, our procedure requires no modification; the edge local graph cumulants are given by a straightforward application of the combinatorial definition, e.g.,

$$\kappa_{3\triangle} = \mu_{3\triangle} - 2\mu_{2\wedge}\mu_{1\lambda} - \mu_{2\triangle}\mu_{1\chi} + 2\mu_{1\chi}\mu_{1\lambda}^2. \tag{47}$$

Scaling the edge local graph cumulants also incorporates the distinguished edge, e.g.,

$$\tilde{\kappa}_{3\triangle} = \frac{\kappa_{3\triangle}}{\mu_{1\chi}\mu_{1\lambda}^2}. \tag{48}$$

S8 Graph cumulants are additive

Essentially, the defining property of cumulants is their unique additive nature when applied to sums of independent random variables (*31, 32*) (e.g., $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ when X and Y are independent). This property is integral to foundational results in probability and statistics, such as the central limit theorem and its generalizations (*46, 45*). In this section, we first define a natural notion of "summing" (denoted by $\oplus$) graph-valued random variables with the same number of nodes. We then show that the graph cumulants of these distributions sum when they are independent.

There are a variety of operations that compose two graphs, such as the disjoint union and a variety of graph products (*104*). Here, we consider the sum of two graphs $G_a \oplus G_b$ to be at the level of their adjacency matrices, defined by simply adding the entries component-wise. In general, as the same graph can be represented by many adjacency matrices, we must assign equal probability to each. In particular, for a graph G with n nodes represented by an adjacency matrix $A = \{a_{ij}\}$, then one distributes the probability associated to this graph uniformly over all matrices $A' = \{a_{\sigma(i)\sigma(j)}\}$ for all permutations σ of $\{1, \ldots, n\}$. Thus, when summing two graphs, one considers all the ways that their sets of representative adjacency matrices could sum. The result is a graph-valued random variable over weighted graphs (see figs. S7 and S8). This notion extends to graph-valued random variables by the distributive property,

$$\left(\sum_{G \in \Omega} p_a(G)\, G \right) \oplus \left(\sum_{G' \in \Omega} p_b(G')\, G' \right) = \sum_{G \in \Omega} \sum_{G' \in \Omega} p_a(G)\, p_b(G')\, G \oplus G'. \qquad (49)$$

Moreover, as $\oplus$ is clearly commutative and associative, it is also well-defined for multiple graph-valued random variables.
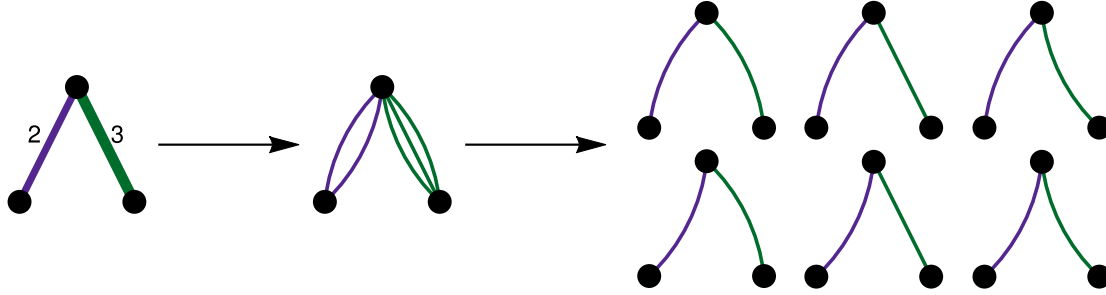
Fig. S7: Subgraphs are counted with multiplicity equal to the product of their edge weights. To motivate this prescription, consider a network with integer edge weights, and represent each integer-weighted edge as that number of unit-weighted edges. For the weighted wedge graph shown here, by counting the number of pairs of unit-weighted edges that share one node, we find that there are 2×3 unweighted wedges in this weighted graph. Indeed, for *any* weighted network (including those with real-valued edge weights), each subgraph should be counted with weight equal to the product of its edge weights.

Even when summing unweighted graph-valued random variables, the result is a graph-valued random variable over weighted graphs. Thus, to obtain the graph cumulants of the resulting distribution, we must generalize the notion of subgraph density to weighted networks (itself a useful extension). Several ways have been proposed to generalize counts of subgraphs to weighted networks (*95, 106, 97, 98*). Within our framework, the consistent prescription is to treat a weighted edge as a collection of multiple edges that sum to its weight. Hence, when counting subgraphs, one should consider each instance with multiplicity equal to the product of its edge weights (*24*). The normalization for graph moments is the same as before, i.e., the counts of the subgraphs in the unweighted complete network (thus, the graph moments of weighted networks may be greater than one). Likewise, the conversion from graph moments to graph cumulants remains identical (see expressions in supplementary materials S10.4).

With the definitions for summing graph-valued random variables and for computing

moments and cumulants of weighted networks, we can now state the main result of this section (see fig. S8): *For two independent graph-valued random variables over n nodes, $\mathcal{G}_a$ and $\mathcal{G}_b$, the graph cumulants of their sum is the sum of their cumulants:*

$$\kappa_{rg}(\mathcal{G}_a \oplus \mathcal{G}_b) = \kappa_{rg}(\mathcal{G}_a) + \kappa_{rg}(\mathcal{G}_b). \tag{50}$$

By induction, this holds for the sum of any number of independent graph-valued random variables.
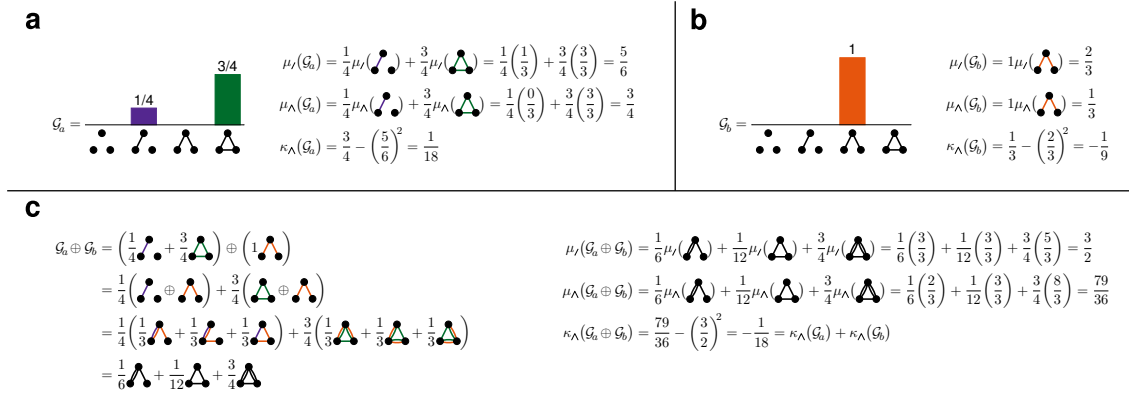
Fig. S8: Graph cumulants have the same additive property as classical cumulants.
a) A graph-valued random variable $\mathcal{G}$ corresponds to a probability distribution over graphs with n nodes. Here, $n = 3$ and $\mathcal{G}_a$ yields a single edge with probability $1/4$ and a triangle with probability $3/4$. **b)** A graph-valued random variable may also be concentrated on a single graph, as is the case for $\mathcal{G}_b$, which always yields a wedge. **c)** Our graph sum $\oplus$ is defined for graph-valued random variables with the same number of nodes. Specifically, the adjacency matrices of each network are summed component-wise, with probability uniformly distributed over all permutations of the nodes. If the graph-valued random variables being summed are independent, the resulting (weighted) graph-valued random variable has cumulants equal to the sum of those of the original distributions.

To demonstrate how to verify this property, we first consider the specific cases of $\kappa_{1/}$, $\kappa_{2\wedge}$, and $\kappa_{3\triangle}$, and then give the combinatorial argument for the general case. Clearly, for

the first moment:

$$\mu_{1/}(\mathcal{G}_a \oplus \mathcal{G}_b) = \mu_{1/}(\mathcal{G}_a) + \mu_{1/}(\mathcal{G}_b), \tag{51}$$

as edge weights simply sum and the normalization remains the same.

For $\mu_{2\wedge}(\mathcal{G}_a \oplus \mathcal{G}_b)$, one must consider the 2^2 ways in which a wedge could be formed: both edges from $\mathcal{G}_a$, giving $\mu_{2\wedge}(\mathcal{G}_a)$; both edges from $\mathcal{G}_b$, giving $\mu_{2\wedge}(\mathcal{G}_b)$; a "left" edge from $\mathcal{G}_a$ and a "right" edge from $\mathcal{G}_b$, giving $\mu_{1/}(\mathcal{G}_a)\mu_{1/}(\mathcal{G}_b)$; and a "left" edge from $\mathcal{G}_b$ and a "right" edge from $\mathcal{G}_a$, giving $\mu_{1/}(\mathcal{G}_b)\mu_{1/}(\mathcal{G}_a)$. Thus,

$$\mu_{2\wedge}(\mathcal{G}_a \oplus \mathcal{G}_b) = \mu_{2\wedge}(\mathcal{G}_a) + \mu_{2\wedge}(\mathcal{G}_b) + 2\mu_{1/}(\mathcal{G}_a)\mu_{1/}(\mathcal{G}_b) \tag{52}$$

Substituting 51 and 52 into the expression for $\kappa_{2\wedge}$ (equation 73), we have

$$\begin{aligned}
\kappa_{2\wedge}(\mathcal{G}_a \oplus \mathcal{G}_b) &= \mu_{2\wedge}(\mathcal{G}_a \oplus \mathcal{G}_b) - \mu_{1/}(\mathcal{G}_a \oplus \mathcal{G}_b)^2 \\
&= \left(\mu_{2\wedge}(\mathcal{G}_a) + \mu_{2\wedge}(\mathcal{G}_b) + 2\mu_{1/}(\mathcal{G}_a)\mu_{1/}(\mathcal{G}_b)\right) - \left(\mu_{1/}(\mathcal{G}_a) + \mu_{1/}(\mathcal{G}_b)\right)^2 \\
&= \mu_{2\wedge}(\mathcal{G}_a) + \mu_{2\wedge}(\mathcal{G}_b) - \mu_{1/}^2(\mathcal{G}_a) - \mu_{1/}^2(\mathcal{G}_b) \\
&= \kappa_{2\wedge}(\mathcal{G}_a) + \kappa_{2\wedge}(\mathcal{G}_b),
\end{aligned}$$

as desired.

Likewise, for $\mu_{3\triangle}(\mathcal{G}_a \oplus \mathcal{G}_b)$, one must again consider the 2^3 ways in which a triangle could be formed: all edges from $\mathcal{G}_a$, giving $\mu_{3\triangle}(\mathcal{G}_a)$; all edges from $\mathcal{G}_b$, giving $\mu_{3\triangle}(\mathcal{G}_b)$; a wedge from $\mathcal{G}_a$ and an edge from $\mathcal{G}_b$ (occurring for three configurations), giving $3\mu_{2\wedge}(\mathcal{G}_a)\mu_{1/}(\mathcal{G}_b)$; and a wedge from $\mathcal{G}_b$ and an edge from $\mathcal{G}_a$ (again occurring for three configurations), giving $3\mu_{2\wedge}(\mathcal{G}_b)\mu_{1/}(\mathcal{G}_a)$. Thus,

$$\mu_{3\triangle}(\mathcal{G}_a \oplus \mathcal{G}_b) = \mu_{3\triangle}(\mathcal{G}_a) + \mu_{3\triangle}(\mathcal{G}_b) + 3\mu_{2\wedge}(\mathcal{G}_a)\mu_{1/}(\mathcal{G}_b) + 3\mu_{2\wedge}(\mathcal{G}_b)\mu_{1/}(\mathcal{G}_a). \tag{53}$$

Substituting 51, 52 and 53 into the expression for $\kappa_{3\triangle}$ (equation 75), we again find that

$$\kappa_{3\triangle}(\mathcal{G}_a \oplus \mathcal{G}_b) = \kappa_{3\triangle}(\mathcal{G}_a) + \kappa_{3\triangle}(\mathcal{G}_b).$$

as desired.

More generally,

$$\mu_g(\mathcal{G}_a \oplus \mathcal{G}_b) = \sum_{x \subseteq g} \mu_x(\mathcal{G}_a)\mu_{x^c}(\mathcal{G}_b) \tag{54}$$

$$= \sum_{x \subseteq g}\left(\sum_{\pi \in P_x}\prod_{p \in \pi}\kappa_p(\mathcal{G}_a)\right)\left(\sum_{\pi \in P_{x^c}}\prod_{p \in \pi}\kappa_p(\mathcal{G}_b)\right) \tag{55}$$

$$= \sum_{\pi \in P_g}\sum_{y \subseteq \pi}\left(\prod_{p \in y}\kappa_p(\mathcal{G}_a)\prod_{p \in y^c}\kappa_p(\mathcal{G}_b)\right) \tag{56}$$

$$= \sum_{\pi \in P_g}\prod_{p \in \pi}\left(\kappa_p(\mathcal{G}_a) + \kappa_p(\mathcal{G}_b)\right) \tag{57}$$

$$= \sum_{\pi \in P_g}\prod_{p \in \pi}\kappa_p\big(\mathcal{G}_a \oplus \mathcal{G}_b\big). \tag{58}$$

where, for notational convenience, g now represents the edge set of the subgraph (replacing the pair rg). The superscript c denotes the complement with respect to the set from which this subset was taken. Line 54 enumerates all possible 2^r ways in which subgraph g could be formed. Line 55 expands the moments in terms of cumulants (equation 2). Line 56 exchanges the order of the summands. Line 57 applies the distributive law. Line 58 demonstrates consistency with the additive property of graph cumulants (equation 50).

S9 Spectral motivation

In this section, we describe a spectral motivation for graph moments and their generalizations.

S9.1 Simple graphs

Consider the problem of parameterizing a distribution over simple graphs with n nodes. We will represent such networks by ordered binary vectors of length $\binom{n}{2}$, where each entry represents an unordered pair of nodes, with 1 indicating that these two nodes are connected by an edge

and 0 that they are not. Let $\mathcal{X}$ be the space of all such vectors. In general, the same graph can be represented by multiple vectors, and distributions over these graphs must give the same probability to all vectors that represent the same graph.

When parametrizing a distribution, it is often desirable that the distribution be "smooth", in the sense that similar graphs are assigned similar probabilities. As our notion of similarity, we consider a "graph edit distance" (*105*), defined as the minimum number of edge changes (i.e., additions or deletions) needed to transform one graph into the other. For example, the wedge is distance 2 from the empty graph, and distance 1 from both the single edge and the triangle.

One common method for parameterizing smooth functions is via a Fourier representation, i.e., in terms of the eigenfunctions of some Laplacian operator. For example, a low-pass filter is equivalent to giving preference to the low-frequency (i.e., long-wavelength) terms, where the location of the cutoff determines the smoothness of the output. To obtain similarly smooth parameterizations over the space of networks, we consider a Laplacian operator based on this graph edit distance.

To this end, we define a (weighted, directed) "edit graph" H_n, with nodes representing unique (i.e., non-isomorphic) networks with n nodes. A directed edge from one node in H_n to another appears whenever the network it represents can be transformed into the other by adding or removing an edge at a single location. The weight of an edge in H_n is given by the number of locations that could be altered to effect this transformation (see Fig. S9 for the case of simple graphs with 4 nodes). The Laplacian of H_n is defined as $L_H = D_{\text{out}} - A_H^\top$, where D_{out} is the diagonal matrix of out-degrees, and $A_H = \{a_{ij}\}$ is the (asymmetric) adjacency matrix with entries a_{ij} equal to the weight of the transition from i to j.

The lowest eigenvalue of L_H is 0, and is associated with a left eigenvector that is uniform over the *unique* networks and a right eigenvector that is uniform over all representations of the networks (i.e., uniform over all binary vectors of length $\binom{n}{2}$), thus corresponding to the $\text{ER}_{n,1/2}$

distribution. The remainder of the spectrum contains additional structure. Its support is the set of positive integers up to and including $\binom{n}{2}$, and each integer has a predictable degeneracy: the multiplicity of an eigenvalue $\lambda = r$ is equal to the number of distinct subgraphs with exactly r edges (with at most n nodes). This is not just a combinatorial coincidence; the span of left eigenvectors (with eigenvalue at most r) is precisely the span of subgraph counts (of order at most r) in each network.

The structure of this spectrum gives rise to a hierarchical parameterization of distributions over networks that is equivalent to our proposed family of hierarchical ERGMs, namely

$$p(G) \propto v_{d,0}(G) \ \exp\left(\sum_{i=1}^{r} \sum_{j=1}^{r_i} \beta_{i,j} v_{l,i,j}(G) \right), \tag{59}$$

where $v_{d,0}$ is the right eigenvector of L_H with eigenvalue 0 (i.e., the $\mathrm{ER}_{n,1/2}$ base distribution); $v_{l,i}$ is the set of left eigenvectors of L_H with eigenvalue i, and $v_{l,i,j}$ is one such vector; r_i is the number of eigenvectors with eigenvalue i; and $\beta_{i,j}$ are the parameters to be determined. In particular, as eigenvectors with the same eigenvalue are intrinsically intertwined, this degeneracy offers a principled motivation for the use of *all* subgraphs up to some chosen order.
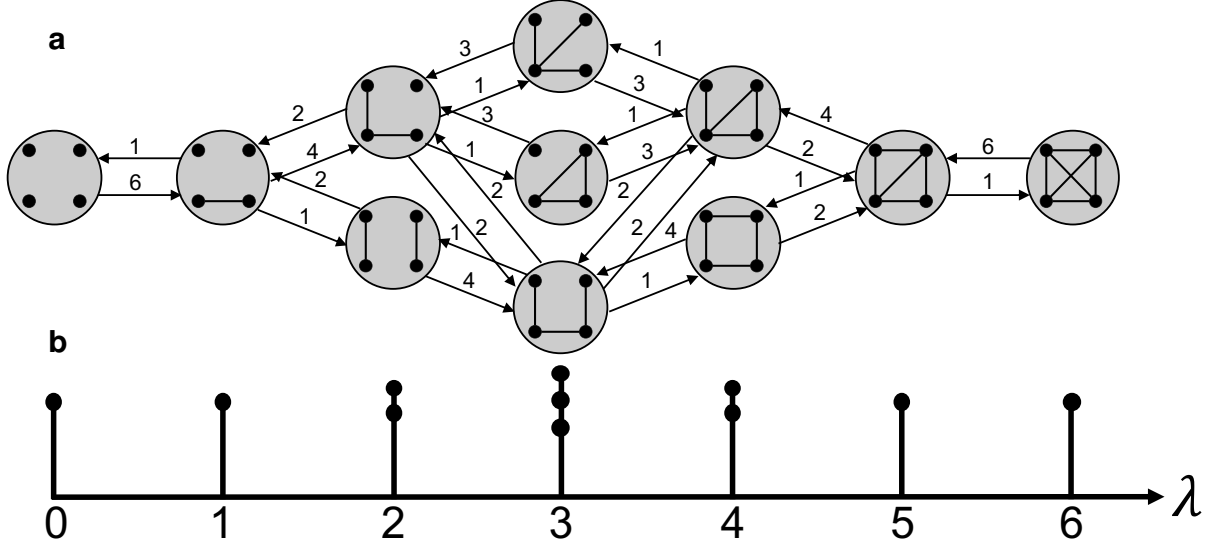
Fig. S9: Spectral motivation for graph moments. **a)** Directed weighted "edit graph" H_4 for simple networks with 4 nodes. The large nodes represent each of the 11 unique (non-isormophic) networks. A directed edge from node i to node j indicates that network i can be transformed into network j by adding or removing an edge at a single location, where the weight of this directed edge is equal to the number of locations that could affect this transformation. Note that every node has an out-degree of 6, corresponding to the $\binom{4}{2}$ possible locations. **b)** Schematic of the spectrum of the Laplacian of H_4. The eigenvalues are integers ranging from 0 to $\binom{4}{2}$, and are degenerate with multiplicity equal to the number of distinct subgraphs (with at most 4 nodes) with that number of edges. This correspondence has a combinatorial interpretation: the span of left eigenvectors with eigenvalue $\lambda \leq r$ is precisely the span of subgraph counts up to and including order r.

S9.2 Generalizations

We now generalize the concept of moments and cumulants for distributions over a set $\mathcal{X} \equiv \mathcal{A}^\ell$, i.e., the vectors of length $\ell \in \mathbb{N}$ over the alphabet $\mathcal{A} = \{a_1, a_2, \ldots\}$, invariant with respect to a group $\mathbb{G}$ acting on this set $\mathcal{X}$. The action of $\mathbb{G}$ induces an equivalence relation on $\mathcal{X}$: $x \sim x' \Leftrightarrow \exists g \in \mathbb{G} \mid x = g \circ x'$, partitioning it into orbits. The distribution over $\mathcal{X}$ is then characterized by assigning a probability to each of these orbits. For example, for the case of simple graphs, the group $\mathbb{G}$ is S_n, acting by permuting the n nodes of a graph. The set $\mathcal{X}$

63

consists of ordered binary vectors of length $\ell = \binom{n}{2}$, where each entry represents an unordered pair of nodes, with 1 indicating that these two nodes are connected by an edge and 0 that they are not. Nonisomorphic graphs are in different orbits, and all the elements in a given orbit correspond to the same graph, with the probability associated to that orbit distributed uniformly over all of its elements.

With this framework, we can construct the weighted directed "edit graph" described in the previous section for an arbitrary set $\mathcal{X}$ and a group $\mathbb{G}$ acting upon it. We can then use the spectrum of the Laplacian of this edit graph to obtain the number of moments at each order. Again, the nodes of the edit graph are the orbits of $\mathcal{X}$ under the action of $\mathbb{G}$, and a directed edge from orbit i to orbit j indicates that an element in orbit i can be transformed into an element in orbit j by changing one of its ℓ entries. The weight of this directed edge is given by the number of elements in orbit j that differ by a single entry from any single fixed element in orbit i.

This abstraction applies to a variety of situations, and naturally encompasses the generalizations previously presented in this paper. For example, for unweighted directed networks with no self-loops, $\mathcal{X}$ is the set of all ordered binary vectors of length $\ell = 2\binom{n}{2}$, where each entry represents an *ordered* pair of nodes, with 1 indicating that there is an edge from the first node to the second and 0 that there is not. The group $\mathbb{G}$ is again S_n. As another example, consider the case of undirected unweighted bipartite networks, i.e., every node has one of two possible "flavors" ("charm" and "strange"), and edges can occur only between nodes of different flavors. The set $\mathcal{X}$ consists of all ordered binary vectors of length $\ell = n_{\mathrm{ch}} n_{\mathrm{str}}$, where each entry represents a different unordered pair of nodes of different flavors, and 1 indicates that these two nodes are connected by an edge and 0 that they are not. The group $\mathbb{G}$ allows for permutations of nodes of the same flavor, namely $S_{n_{\mathrm{ch}}} \times S_{n_{\mathrm{str}}}$.

We now illustrate the versatility of this formalism by describing an additional generalization, namely, k-uniform hypergraphs, i.e., where each (hyper)edge represents a connection between

k distinct nodes. As in the standard graph case (i.e., $k = 2$), the group $\mathbb{G}$ is the symmetric group S_n acting by permuting the n nodes. The set $\mathcal{X}$ consists of all ordered binary vectors of length $\ell = \binom{n}{k}$, where each entry represents an unordered set of k nodes, and a 1 indicates the presence of a hyperedge between them and 0 its absence. The orbits induced by this group action again partition the elements of $\mathcal{X}$ into equivalence classes, one for each of the unique hypergraphs. The eigenvalues of the Laplacian of the corresponding edit graph follow a similar pattern: associated to the eigenvalue of 0 is a left eigenvector that is uniform over unique hypergraphs, and a right eigenvector that is uniform over all elements of $\mathcal{X}$. Likewise, for the remaining left eigenvectors, there is one eigenvector with associated eigenvalue of 1 that is linear in the number of hyperedges. At second order (associated eigenvalue of 2), there are now k eigenvalues (for $n \geq 2k$), corresponding to the k ways that two hyperedges can relate (sharing any number from 0 to $k - 1$ nodes).

S10 Formulas for graph moments
and graph cumulants

Here, we provide the expressions for graph moments and graph cumulants used to obtain the results presented in this paper. These expressions are also included explicitly in our associated code, and we have automated their derivation to arbitrary order. We first give the normalizations for obtaining the graph moments as well as the expressions for efficiently computing the disconnected subgraph counts (see supplementary materials S1.1). We then give the expressions for their conversion to graph cumulants (by inverting equation 2).

S10.1 Undirected, unweighted networks

For simple graphs, we now enumerate the expressions for all third-order graph moments and cumulants, as well as those that are necessary for computing the (sixth-order) cumulant

associated with the complete graph with four nodes. The remaining expressions up to and including sixth order are explicitly included on our code.

S10.1.1 Graph moments

$$\mu_{1/} = \frac{c_/}{\binom{n}{2}} \tag{60}$$

$$\mu_{2\wedge} = \frac{c_\wedge}{3\binom{n}{3}} \tag{61}$$

$$\mu_{2//} = \frac{c_{//}}{3\binom{n}{4}} = \frac{\binom{c_/}{2} - c_\wedge}{3\binom{n}{4}} \tag{62}$$

$$\mu_{3\triangle} = \frac{c_\triangle}{\binom{n}{3}} \tag{63}$$

$$\mu_{3\curlywedge} = \frac{c_\curlywedge}{4\binom{n}{4}} \tag{64}$$

$$\mu_{3\sqcap} = \frac{c_\sqcap}{12\binom{n}{4}} \tag{65}$$

$$\mu_{3/\!\!\wedge} = \frac{c_{/\!\!\wedge}}{30\binom{n}{5}} = \frac{c_\wedge(c_/ - 2) - 3c_\triangle - 3c_\curlywedge - 2c_\sqcap}{30\binom{n}{5}} \tag{66}$$

$$\mu_{3/\!/\!/} = \frac{c_{/\!/\!/}}{15\binom{n}{6}} = \frac{\binom{c_/}{3} - c_\triangle - c_\curlywedge - c_\sqcap - c_{/\!\!\wedge}}{15\binom{n}{6}} \tag{67}$$

$$\mu_{4\triangle\!\!\!/} = \frac{c_{\triangle\!\!\!/}}{12\binom{n}{4}} \tag{68}$$

$$\mu_{4\square} = \frac{c_\square}{3\binom{n}{4}} \tag{69}$$

$$\mu_{5\boxslash} = \frac{c_\boxslash}{6\binom{n}{4}} \tag{70}$$

$$\mu_{6\boxtimes} = \frac{c_\boxtimes}{\binom{n}{4}} \tag{71}$$

S10.1.2 Graph cumulants

$$\kappa_{1/} = \mu_{1/} \tag{72}$$

$$\kappa_{2\wedge} = \mu_{2\wedge} - \mu_{1/}^2 \tag{73}$$

$$\kappa_{2//} = \mu_{2//} - \mu_{1/}^2 \tag{74}$$

$$\kappa_{3\triangle} = \mu_{3\triangle} - 3\mu_{2\wedge}\mu_{1/} + 2\mu_{1/}^3 \tag{75}$$

$$\kappa_{3\curlywedge} = \mu_{3\curlywedge} - 3\mu_{2\wedge}\mu_{1/} + 2\mu_{1/}^3 \tag{76}$$

$$\kappa_{3\sqcap} = \mu_{3\sqcap} - 2\mu_{2\wedge}\mu_{1/} - \mu_{2/\!/}\mu_{1/} + 2\mu_{1/}^3 \tag{77}$$

$$\kappa_{3/\!\wedge} = \mu_{3/\!\wedge} - \mu_{2\wedge}\mu_{1/} - 2\mu_{2/\!/}\mu_{1/} + 2\mu_{1/}^3 \tag{78}$$

$$\kappa_{3/\!/\!/} = \mu_{3/\!/\!/} - 3\mu_{2/\!/}\mu_{1/} + 2\mu_{1/}^3 \tag{79}$$

$$\kappa_{4\triangle\!\vee} = \mu_{4\triangle\!\vee} - \mu_{3\triangle}\mu_{1/} - \mu_{3\curlywedge}\mu_{1/} - 2\mu_{3\sqcap}\mu_{1/}$$
$$- 2\mu_{2\wedge}^2 - \mu_{2\wedge}\mu_{2/\!/} + 10\mu_{2\wedge}\mu_{1/}^2 + 2\mu_{2/\!/}\mu_{1/}^2 - 6\mu_{1/}^4, \tag{80}$$

$$\kappa_{4\square} = \mu_{4\square} - 4\mu_{3\sqcap}\mu_{1/} - 2\mu_{2\wedge}^2 - \mu_{2/\!/}^2 + 8\mu_{2\wedge}\mu_{1/}^2 + 4\mu_{2/\!/}\mu_{1/}^2 - 6\mu_{1/}^4 \tag{81}$$

$$\kappa_{5\boxtimes} = \mu_{5\boxtimes} - 4\mu_{4\triangle\!\vee}\mu_{1/} - \mu_{4\square}\mu_{1/} - 2\mu_{3\triangle}\mu_{2\wedge} - 2\mu_{3\curlywedge}\mu_{2\wedge} - 4\mu_{3\sqcap}\mu_{2\wedge} - 2\mu_{3\sqcap}\mu_{2/\!/}$$
$$+ 4\mu_{3\triangle}\mu_{1/}^2 + 4\mu_{3\curlywedge}\mu_{1/}^2 + 8\mu_{3\sqcap}\mu_{1/}^2 + 4\mu_{3\sqcap}\mu_{1/}^2$$
$$+ 20\mu_{2\wedge}^2\mu_{1/} + 8\mu_{2\wedge}\mu_{2/\!/}\mu_{1/} + 2\mu_{2/\!/}^2\mu_{1/}$$
$$- 48\mu_{2\wedge}\mu_{1/}^3 - 12\mu_{2/\!/}\mu_{1/}^3 + 24\mu_{1/}^5 \tag{82}$$

$$\kappa_{6\boxtimes} = \mu_{6\boxtimes} - 6\mu_{5\boxtimes}\mu_{1/} - 12\mu_{4\triangle\!\vee}\mu_{2\wedge} - 3\mu_{4\square}\mu_{2/\!/} + 24\mu_{4\triangle\!\vee}\mu_{1/}^2 + 6\mu_{4\square}\mu_{1/}^2$$
$$- 4\mu_{3\triangle}\mu_{3\curlywedge} - 6\mu_{3\sqcap}^2$$
$$+ 24\mu_{3\triangle}\mu_{2\wedge}\mu_{1/} + 24\mu_{3\curlywedge}\mu_{2\wedge}\mu_{1/} + 48\mu_{3\sqcap}\mu_{2\wedge}\mu_{1/} + 24\mu_{3\sqcap}\mu_{2/\!/}\mu_{1/}$$
$$- 24\mu_{3\triangle}\mu_{1/}^3 - 24\mu_{3\curlywedge}\mu_{1/}^3 - 72\mu_{3\sqcap}\mu_{1/}^3$$
$$+ 12\mu_{2\wedge}^3 + 15\mu_{2\wedge}^2\mu_{2/\!/} + 3\mu_{2/\!/}^3$$
$$- 153\mu_{2\wedge}^2\mu_{1/}^2 - 90\mu_{2\wedge}\mu_{2/\!/}\mu_{1/}^2 - 27\mu_{2/\!/}^2\mu_{1/}^2$$
$$+ 288\mu_{2\wedge}\mu_{1/}^4 + 72\mu_{2/\!/}\mu_{1/}^4 - 120\mu_{1/}^6 \tag{83}$$

S10.2 Directed networks

We now enumerate the expressions necessary for computing the graph moments and cumulants of all directed subgraphs with three nodes, including the sixth-order graph cumulant associated

with the complete directed triad. The remaining expressions up to and including fifth order are explicitly included on our code.

S10.2.1 Graph moments

$$\mu_{1\nearrow} = \frac{c_{1\nearrow}}{2\binom{n}{2}} \tag{84}$$

$$\mu_{2\wedge} = \frac{c_{2\wedge}}{3\binom{n}{3}} \tag{85}$$

$$\mu_{2\wedge} = \frac{c_{2\wedge}}{3\binom{n}{3}} \tag{86}$$

$$\mu_{2\wedge} = \frac{c_{2\wedge}}{6\binom{n}{3}} \tag{87}$$

$$\mu_{2\emptyset} = \frac{c_{2\emptyset}}{\binom{n}{2}} \tag{88}$$

$$\mu_{3\wedge} = \frac{c_{3\wedge}}{6\binom{n}{3}} \tag{89}$$

$$\mu_{3\wedge} = \frac{c_{3\wedge}}{6\binom{n}{3}} \tag{90}$$

$$\mu_{3\triangle} = \frac{c_{3\triangle}}{6\binom{n}{3}} \tag{91}$$

$$\mu_{3\triangle} = \frac{c_{3\triangle}}{2\binom{n}{3}} \tag{92}$$

$$\mu_{4\triangle} = \frac{c_{4\triangle}}{3\binom{n}{3}} \tag{93}$$

$$\mu_{4\triangle} = \frac{c_{4\triangle}}{3\binom{n}{3}} \tag{94}$$

$$\mu_{4\triangle} = \frac{c_{4\triangle}}{6\binom{n}{3}} \tag{95}$$

$$\mu_{4\wedge} = \frac{c_{4\wedge}}{3\binom{n}{3}} \tag{96}$$

$$\mu_{5\triangle} = \frac{c_{5\triangle}}{6\binom{n}{3}} \tag{97}$$

$$\mu_{6\triangle} = \frac{c_{6\triangle}}{\binom{n}{3}} \tag{98}$$

S10.2.2 Graph cumulants

$$\kappa_{1\prime} = \mu_{1\prime} \tag{99}$$

$$\kappa_{2\wedge} = \mu_{2\wedge} - \mu_{1\prime}^2 \tag{100}$$

$$\kappa_{2\wedge} = \mu_{2\wedge} - \mu_{1\prime}^2 \tag{101}$$

$$\kappa_{2\wedge} = \mu_{2\wedge} - \mu_{1\prime}^2 \tag{102}$$

$$\kappa_{2\theta} = \mu_{2\theta} - \mu_{1\prime}^2 \tag{103}$$

$$\kappa_{3\wedge} = \mu_{3\wedge} - \mu_{2\wedge}\mu_{1\prime} - \mu_{2\wedge}\mu_{1\prime} - \mu_{2\theta}\mu_{1\prime} + 2\mu_{1\prime}^3 \tag{104}$$

$$\kappa_{3\wedge} = \mu_{3\wedge} - \mu_{2\wedge}\mu_{1\prime} - \mu_{2\wedge}\mu_{1\prime} - \mu_{2\theta}\mu_{1\prime} + 2\mu_{1\prime}^3 \tag{105}$$

$$\kappa_{3\triangle} = \mu_{3\triangle} - \mu_{2\wedge}\mu_{1\prime} - \mu_{2\wedge}\mu_{1\prime} - \mu_{2\wedge}\mu_{1\prime} + 2\mu_{1\prime}^3 \tag{106}$$

$$\kappa_{3\triangle} = \mu_{3\triangle} - 3\mu_{2\wedge}\mu_{1\prime} + 2\mu_{1\prime}^3 \tag{107}$$

$$\kappa_{4\triangle} = \mu_{4\triangle} - 2\mu_{3\wedge}\mu_{1\prime} - 2\mu_{3\triangle}\mu_{1\prime} - \mu_{2\wedge}^2 - \mu_{2\wedge}^2 - \mu_{2\wedge}\mu_{2\theta}$$
$$+ 2\mu_{2\wedge}\mu_{1\prime}^2 + 4\mu_{2\wedge}\mu_{1\prime}^2 + 4\mu_{2\wedge}\mu_{1\prime}^2 + 2\mu_{2\theta}\mu_{1\prime}^2 - 6\mu_{1\prime}^4 \tag{108}$$

$$\kappa_{4\triangle} = \mu_{4\triangle} - 2\mu_{3\wedge}\mu_{1\prime} - 2\mu_{3\triangle}\mu_{1\prime} - \mu_{2\wedge}^2 - \mu_{2\wedge}^2 - \mu_{2\wedge}\mu_{2\theta}$$
$$+ 4\mu_{2\wedge}\mu_{1\prime}^2 + 2\mu_{2\wedge}\mu_{1\prime}^2 + 4\mu_{2\wedge}\mu_{1\prime}^2 + 2\mu_{2\theta}\mu_{1\prime}^2 - 6\mu_{1\prime}^4 \tag{109}$$

$$\kappa_{4\triangle} = \mu_{4\triangle} - \mu_{3\wedge}\mu_{1\prime} - \mu_{3\wedge}\mu_{1\prime} - \mu_{3\triangle}\mu_{1\prime} - \mu_{3\triangle}\mu_{1\prime} - \mu_{2\wedge}\mu_{2\wedge} - \mu_{2\wedge}\mu_{2\wedge} - \mu_{2\wedge}\mu_{2\theta}$$
$$+ 2\mu_{2\wedge}\mu_{1\prime}^2 + 2\mu_{2\wedge}\mu_{1\prime}^2 + 6\mu_{2\wedge}\mu_{1\prime}^2 + 2\mu_{2\theta}\mu_{1\prime}^2 - 6\mu_{1\prime}^4 \tag{110}$$

$$\kappa_{4\wedge} = \mu_{4\wedge} - 2\mu_{3\wedge}\mu_{1\prime} - 2\mu_{3\wedge}\mu_{1\prime} - \mu_{2\wedge}\mu_{2\wedge} - \mu_{2\wedge}^2 - \mu_{2\theta}^2$$
$$+ 2\mu_{2\wedge}\mu_{1\prime}^2 + 2\mu_{2\wedge}\mu_{1\prime}^2 + 4\mu_{2\wedge}\mu_{1\prime}^2 + 4\mu_{2\theta}\mu_{1\prime}^2 - 6\mu_{1\prime}^4 \tag{111}$$

$$\kappa_{5\triangle} = \mu_{5\triangle} - \mu_{4\triangle}\mu_{1\prime} - \mu_{4\triangle}\mu_{1\prime} - 2\mu_{4\triangle}\mu_{1\prime} - \mu_{4\wedge}\mu_{1\prime} -$$
$$- \mu_{3\wedge}\mu_{2\wedge} - \mu_{3\wedge}\mu_{2\wedge} - \mu_{3\wedge}\mu_{2\theta} - \mu_{3\wedge}\mu_{2\wedge} - \mu_{3\wedge}\mu_{2\wedge}$$
$$- \mu_{3\wedge}\mu_{2\theta} - \mu_{3\triangle}\mu_{2\wedge} - \mu_{3\triangle}\mu_{2\wedge} - \mu_{3\triangle}\mu_{2\wedge} - \mu_{3\triangle}\mu_{2\wedge}$$
$$+ 6\mu_{3\wedge}\mu_{1\prime}^2 + 6\mu_{3\wedge}\mu_{1\prime}^2 + 9\mu_{3\triangle}\mu_{1\prime}^2 + 2\mu_{3\triangle}\mu_{1\prime}^2$$

$$+ 2\mu_{2\wedge}^2 + 2\mu_{2\wedge}^2 + 6\mu_{2\wedge}^2 + 2\mu_{2\emptyset}^2 + 2\mu_{2\wedge}\mu_{2\wedge}$$

$$+ 4\mu_{2\wedge}\mu_{2\wedge} + 4\mu_{2\wedge}\mu_{2\wedge} + 2\mu_{2\wedge}\mu_{2\emptyset} + 2\mu_{2\wedge}\mu_{2\emptyset} + 4\mu_{2\wedge}\mu_{2\emptyset}$$

$$- 12\mu_{2\wedge}\mu_{1\prime}^3 - 12\mu_{2\wedge}\mu_{1\prime}^3 - 24\mu_{2\wedge}\mu_{1\prime}^3 - 12\mu_{2\emptyset}\mu_{1\prime}^3 + 24\mu_{1\prime}^5, \tag{112}$$

$$\kappa_{6\triangle} = \mu_{6\triangle} - 6\mu_{5\triangle}\mu_{1\prime} - 3\mu_{4\triangle}\mu_{2\wedge} - 3\mu_{4\triangle}\mu_{2\wedge} - 6\mu_{4\triangle}\mu_{2\wedge} - 3\mu_{4\triangle}\mu_{2\emptyset}$$

$$+ 6\mu_{4\triangle}\mu_{1\prime}^2 + 6\mu_{4\triangle}\mu_{1\prime}^2 + 12\mu_{4\triangle}\mu_{1\prime}^2 + 6\mu_{4\triangle}\mu_{1\prime}^2 - 3\mu_{3\triangle}^2 - 3\mu_{3\triangle}^2 - 3\mu_{3\triangle}^2 - \mu_{3\triangle}^2$$

$$+ 12\mu_{3\triangle}\mu_{2\wedge}\mu_{1\prime} + 12\mu_{3\triangle}\mu_{2\wedge}\mu_{1\prime} + 12\mu_{3\triangle}\mu_{2\emptyset}\mu_{1\prime} + 12\mu_{3\triangle}\mu_{2\wedge}\mu_{1\prime} + 12\mu_{3\triangle}\mu_{2\wedge}\mu_{1\prime}$$

$$+ 12\mu_{3\triangle}\mu_{2\emptyset}\mu_{1\prime} + 12\mu_{3\triangle}\mu_{2\wedge}\mu_{1\prime} + 12\mu_{3\triangle}\mu_{2\wedge}\mu_{1\prime} + 12\mu_{3\triangle}\mu_{2\wedge}\mu_{1\prime} + 12\mu_{3\triangle}\mu_{2\wedge}\mu_{1\prime}$$

$$- 36\mu_{3\triangle}\mu_{1\prime}^3 - 36\mu_{3\triangle}\mu_{1\prime}^3 - 36\mu_{3\triangle}\mu_{1\prime}^3 - 12\mu_{3\triangle}\mu_{1\prime}^3$$

$$+ 2\mu_{2\wedge}^3 + 6\mu_{2\wedge}\mu_{2\wedge}\mu_{2\emptyset} + 6\mu_{2\wedge}\mu_{2\wedge}^2 + 2\mu_{2\wedge}^3 + 6\mu_{2\wedge}\mu_{2\wedge}^2 + 6\mu_{2\wedge}^2\mu_{2\emptyset} + 2\mu_{2\emptyset}^3$$

$$- 18\mu_{2\wedge}^2\mu_{1\prime}^2 - 18\mu_{2\wedge}\mu_{2\wedge}\mu_{1\prime}^2 - 36\mu_{2\wedge}\mu_{2\wedge}\mu_{1\prime}^2 - 18\mu_{2\wedge}\mu_{2\emptyset}\mu_{1\prime}^2 - 18\mu_{2\wedge}^2\mu_{1\prime}^2$$

$$- 36\mu_{2\wedge}\mu_{2\wedge}\mu_{1\prime}^2 - 18\mu_{2\wedge}\mu_{2\emptyset}\mu_{1\prime}^2 - 54\mu_{2\wedge}^2\mu_{1\prime}^2 - 36\mu_{2\wedge}\mu_{2\emptyset}\mu_{1\prime}^2 - 18\mu_{2\emptyset}^2\mu_{1\prime}^2$$

$$+ 60\mu_{2\wedge}\mu_{1\prime}^4 + 60\mu_{2\wedge}\mu_{1\prime}^4 + 120\mu_{2\wedge}\mu_{1\prime}^4 + 60\mu_{2\emptyset}\mu_{1\prime}^4 - 120\mu_{1\prime}^6, \tag{113}$$

S10.3   Networks with node attributes

Often, networks have additional attributes associated with the nodes. By incorporating this information into the graph cumulant formalism, one can reveal structure that is correlated with these attributes. Our example in Figure S2 considers cumulants associated with subgraphs containing up to three nodes for a network with a binary node attribute. We now enumerate the expressions for computing the graph moments and cumulants required for this analysis (as well as for the 3-star subgraphs). Note that the mapping from attributes to colors is arbitrary; the colors may be reversed in any expression (e.g., the expression for $\kappa_{2\wedge}$ can be obtained from that for $\kappa_{2\wedge}$ by exchanging all instances of purple and green with each other).

S10.3.1 Graph moments

$$\mu_{1\,\edge} = \frac{c_{1\,\edge}}{\binom{n_\bullet}{2}} \tag{114}$$

$$\mu_{1\,\edge} = \frac{c_{1\,\edge}}{n_\bullet n_\bullet} \tag{115}$$

$$\mu_{2\,\wedge} = \frac{c_{2\,\wedge}}{3\binom{n_\bullet}{3}} \tag{116}$$

$$\mu_{2\,\wedge} = \frac{c_{2\,\wedge}}{2\binom{n_\bullet}{2}n_\bullet} \tag{117}$$

$$\mu_{2\,\wedge} = \frac{c_{2\,\wedge}}{n_\bullet \binom{n_\bullet}{2}} \tag{118}$$

$$\mu_{3\,\triangle} = \frac{c_{3\,\triangle}}{\binom{n_\bullet}{3}} \tag{119}$$

$$\mu_{3\,\triangle} = \frac{c_{3\,\triangle}}{\binom{n_\bullet}{2}n_\bullet} \tag{120}$$

$$\mu_{3\,\star} = \frac{c_{3\,\star}}{4\binom{n_\bullet}{4}} \tag{121}$$

$$\mu_{3\,\star} = \frac{c_{3\,\star}}{3\binom{n_\bullet}{3}n_\bullet} \tag{122}$$

$$\mu_{3\,\star} = \frac{c_{3\,\star}}{2\binom{n_\bullet}{2}\binom{n_\bullet}{2}} \tag{123}$$

$$\mu_{3\,\star} = \frac{c_{3\,\star}}{n_\bullet \binom{n_\bullet}{3}} \tag{124}$$

S10.3.2 Graph cumulants

$$\kappa_{1\,\edge} = \mu_{1\,\edge} \tag{125}$$

$$\kappa_{1\,\edge} = \mu_{1\,\edge} \tag{126}$$

$$\kappa_{2\,\wedge} = \mu_{2\,\wedge} - \mu_{1\,\edge}^2 \tag{127}$$

$$\kappa_{2\,\wedge} = \mu_{2\,\wedge} - \mu_{1\,\edge}\mu_{1\,\edge} \tag{128}$$

$$\kappa_{2\,\wedge} = \mu_{2\,\wedge} - \mu_{1\,\edge}^2 \tag{129}$$

$$\kappa_{3\,\triangle} = \mu_{3\,\triangle} - 3\mu_{2\,\wedge}\mu_{1\,\edge} + 2\mu_{1\,\edge}^3 \tag{130}$$

71

$$\kappa_{3\triangle} = \mu_{3\triangle} - 2\mu_{2\wedge}\mu_{1/} - \mu_{2\wedge}\mu_{1/} + 2\mu_{1/}\mu_{1/}^2 \tag{131}$$

$$\kappa_{3\curlywedge} = \mu_{3\curlywedge} - 3\mu_{2\wedge}\mu_{1/} + 2\mu_{1/}^3 \tag{132}$$

$$\kappa_{3\curlywedge} = \mu_{3\curlywedge} - 2\mu_{2\wedge}\mu_{1/} - \mu_{2\wedge}\mu_{1/} + 2\mu_{1/}^2\mu_{1/} \tag{133}$$

$$\kappa_{3\curlywedge} = \mu_{3\curlywedge} - 2\mu_{2\wedge}\mu_{1/} - \mu_{2\wedge}\mu_{1/} + 2\mu_{1/}\mu_{1/}^2 \tag{134}$$

$$\kappa_{3\curlywedge} = \mu_{3\curlywedge} - 3\mu_{2\wedge}\mu_{1/} + 2\mu_{1/}^3 \tag{135}$$

S10.4 Weighted networks

For weighted networks, each instance of a subgraph is counted with weight equal to the product of its edge weights (see fig. S7 and supplementary materials S8)), but the normalization to moments and conversion to cumulants remain the same as in the unweighted case. However, the expressions for computing the disconnected counts from the connected counts requires a slight modification. For example, the counts of two weighted edges that do not share any node now includes a second-order term related to the square of the edge weights.

S10.4.1 Graph moments

$$\mu_{1/} = \frac{1}{\binom{n}{2}} \sum_{0 \leq i < j \leq n} w_{ij} \tag{136}$$

$$\mu_{2\varnothing} = \frac{1}{\binom{n}{2}} \sum_{0 \leq i < j \leq n} w_{ij}^2 \tag{137}$$

$$\mu_{2\wedge} = \frac{1}{3\binom{n}{3}} \sum_{0 \leq i < j < k \leq n} \left(w_{ij} w_{jk} + w_{jk} w_{ki} + w_{ki} w_{ij} \right) \tag{138}$$

$$\mu_{2//} = \frac{1}{3\binom{n}{4}} \left(\frac{1}{2}\left(\mu_{1/}\binom{n}{2} \right)^2 - \frac{1}{2}\left(\mu_{2\varnothing}\binom{n}{2} \right) - \left(\mu_{2\wedge}3\binom{n}{3} \right) \right) \tag{139}$$

$$\mu_{3\varnothing} = \frac{1}{\binom{n}{2}} \sum_{0 \leq i < j \leq n} w_{ij}^3 \tag{140}$$

$$\mu_{3\wedge} = \frac{1}{6\binom{n}{3}} \sum_{0 \leq i < j < k \leq n} \left(w_{ij}^2 w_{jk} + w_{jk}^2 w_{ki} + w_{ki}^2 w_{ij} + w_{ij} w_{jk}^2 + w_{jk} w_{ki}^2 + w_{ki} w_{ij}^2 \right) \tag{141}$$

72

S10.4.2 Graph cumulants

$$\kappa_{1/} = \mu_{1/} \tag{142}$$

$$\kappa_{2\emptyset} = \mu_{2\emptyset} - \mu_{1/}^2 \tag{143}$$

$$\kappa_{2\wedge} = \mu_{2\wedge} - \mu_{1/}^2 \tag{144}$$

$$\kappa_{2/\!/} = \mu_{2/\!/} - \mu_{1/}^2 \tag{145}$$

$$\kappa_{3\emptyset} = \mu_{3\emptyset} - 3\mu_{2\emptyset}\mu_{1/} + 2\mu_{1/}^3 \tag{146}$$

$$\kappa_{3\wedge} = \mu_{3\emptyset} - 2\mu_{2\wedge}\mu_{1/} - \mu_{2\emptyset}\mu_{1/} + 2\mu_{1/}^3 \tag{147}$$