Quantifying Network Similarity using Graph Cumulants

Gecia Bravo-Hermsdorff[*]   GECIA.BRAVO@GMAIL.COM
Department of Statistical Sciences
University College London
London, WC1E 7H8, United Kingdom

Lee M. GundersonGunderson[*]   L.GUNDERSON@UCL.AC.UK
Gatsby Computational Neuroscience Unit
University College London
London, W1T 4JG, United Kingdom

Pierre-André Maugis   PAMAUGIS@GOOGLE.COM
Google Research
Zürich, 8002, Switzerland

Carey E. Priebe   CEP@JHU.EDU
Department of Applied Mathematics and Statistics
Whiting School of Engineering
Johns Hopkins University
Baltimore, MD 21218, USA

Editor: Tina Eliassi-Rad

Abstract

How might one test the hypothesis that networks were sampled from the same distribution? Here, we compare two statistical tests that use subgraph counts to address this question. The first uses the empirical subgraph densities themselves as estimates of those of the underlying distribution. The second test uses a new approach that converts these subgraph densities into estimates of the *graph cumulants* of the distribution (without any increase in computational complexity). We demonstrate — via theory, simulation, and application to real data — the superior statistical power of using graph cumulants. In summary, when analyzing data using subgraph/motif densities, we suggest using the corresponding graph cumulants instead.

Keywords: subgraph counts, two-sample test, graph cumulants, graph moments, network comparison, motif analysis, graphons

1 Statement of the Problem

In this paper, we use statistics based on subgraph counts to address the following problem:

> Given two samples, $\boldsymbol{G}_A$ and $\boldsymbol{G}_B$, each containing s graphs
> sampled i.i.d. from unknown distributions $\mathcal{G}_A$ and $\mathcal{G}_B$, respectively,
> the goal is to infer whether $\mathcal{G}_A$ and $\mathcal{G}_B$ are different distributions.

---

[*]. *Shared first authorship.*

2 What we are Doing (Motivation)

The theory for statistical analysis of i.i.d. data (e.g., the height and weight of different breeds of dogs) is well-established (Casella and Berger, 2021; Stuart and Kendall, 1963). However, when the data of interest are *interactions* (e.g., who plays with whom in the dog park), a similar consensus has not yet been reached (Orbanz, 2017).

To analyze such networks of interactions, one common approach is to use statistics based on counts of small substructures (i.e., subgraphs or "motifs") (Aliakbarpour et al., 2018; Maugis et al., 2020; Pattillo et al., 2013; Xia et al., 2019). For example: the densities of "k-star" subgraphs provide increasingly detailed information about a network's degree distribution (Rauh, 2017), and the density of cliques of various sizes gives information about the scale of its clustering (Ouyang et al., 2019).

To make meaningful comparisons, one must first answer the question: "Compared to what?" The frequently-used clustering coefficient opts to divide the number of "complete" triangles by the number of "incomplete" triangles (Newman, 2001). The often-cited configuration model compares the observed subgraph counts with those of random networks with the same degree distribution (Alon, 2007; Barabási and Albert, 1999).

The recently-proposed graph cumulants (Gunderson and Bravo-Hermsdorff, 2020) offer a natural set of statistics based on a combinatorial view of cumulants (e.g., mean, variance, skew, kurtosis, etc.). Graph cumulants compare the density of a substructure in a network to the density that would be expected due to the prevalence of smaller substructures, thereby quantifying an intuitive notion of "excess propensity" for that substructure.

It is well-known that subgraph densities are useful for characterizing individual networks (Bickel et al., 2011; Bhattacharyya and Bickel, 2015) and for comparing different networks (Maugis et al., 2020; Bhattacharya et al., 2022; Zhang and Xia, 2022; Jin et al., 2021). A few works have also used statistics that combine subgraph densities to compare networks. In particular, Gao and Lafferty (2017) uses two statistics (one involving edge and 2-star densities, and the other involving edge and triangle densities) to devise a test for distinguishing an Erdős-Rényi model from an alternative stochastic block model (SBM).

Here, we describe a simple statistical test for comparing networks based on graph cumulants. For comparison, we consider the analogous test based on graph moments (i.e., subgraph densities) (Maugis et al., 2020). Our results strongly suggest that graph cumulants should be the default choice of statistics for problems involving subgraph counts.

3 What we are Not Doing (Related Work)

There are many ways to compare networks (Tantardini et al., 2019). Here, we describe some other common approaches, highlighting their differences with the setting considered in this paper.

3.1 Spectral Methods

Broadly, these methods use the eigendecomposition of matrices associated with the network, such as the adjacency matrix or the graph Laplacian (Chen et al., 2020; Bravo-Hermsdorff and Gunderson, 2019; Mukherjee et al., 2016). Here, we focus instead on statistics based on the frequency of small subgraphs. These two approaches are complementary (Lovász,

2012); flavorfully, spectral methods are more sensitive to the graph's global structure (its "shape"), whereas methods based on subgraph frequencies are more sensitive to the graph's local structure (its "texture").

3.2 Matching Nodes

When the networks being compared have the same set of unique names for all their nodes (Ghoshdastidar et al., 2020; Tang et al., 2017), it is highly advantageous for statistical tests to incorporate this one-to-one mapping (e.g., when comparing fMRI data of different subjects, it helps to assume that their hippocampi are sufficiently analogous). However, such information is not always available (e.g., when comparing the social interactions within different schools). This work addresses the latter problem; the nodes are considered to be indistinguishable (i.e., exchangeable), and only the statistics of their pairwise interactions (i.e., their edges) are known to the tests.

3.3 Obtaining Significance by Sampling

Many tests for comparing networks require sampling from the inferred distributions to estimate the significance of their differences. Common examples include the use of configuration models (Masuda et al., 2018), exponential random graph models (An, 2016), and geometric random graph models (Asta and Shalizi, 2014). Such methods are typically computationally intensive, rendering them difficult (or impossible) to implement in practice (Ginoza and Mugler, 2010). The two tests considered in this paper sidestep this issue entirely, analytically computing the significance of the observed differences between the networks.

4 How we Name Things (Notation)

A capital G denotes a single graph with n nodes. All graphs are assumed to be undirected, unweighted, simple graphs.[1] A calligraphic $\mathcal{G}$ denotes a distribution over such graphs. All graph distributions are assumed to be generated by a single graphon (Lovász, 2012) (see Section 7).[2] A bold $\boldsymbol{G}$ denotes a sample of s graphs from such a distribution. All graphs in a sample are assumed to have the same number of nodes.

A plebeian g denotes a subgraph. An emboldened $\boldsymbol{g}$ denotes a set of subgraphs. The set of subgraphs with at most r edges is denoted by $\boldsymbol{g}_r$, and the restriction to connected subgraphs is denoted by $\boldsymbol{g}_r^{(c)}$ (e.g., $\boldsymbol{g}_3 = \{/, \wedge, /\!/, \triangle, \curlywedge, \sqcap, /\!\!\wedge, /\!/\!/\}$ and $\boldsymbol{g}_3^{(c)} = \{/, \wedge, \triangle, \curlywedge, \sqcap\}$). Parenthetical superscripts are also used to distinguish the covariance matrices of graph moments $\boldsymbol{\Sigma}^{(\mu)}$ from those of graph cumulants $\boldsymbol{\Sigma}^{(\kappa)}$.

The statistics we consider here, namely, graph moments and graph cumulants, have associated with them a particular subgraph g (see Section 5). $\mu_g(\mathcal{G})$ denotes the graph moment (associated with subgraph g) of a distribution $\mathcal{G}$ (see Section 5.1), and $\kappa_g(\mathcal{G})$ the corresponding graph cumulant (see Section 5.2). A bold $\boldsymbol{\mu}$ (or $\boldsymbol{\kappa}$) denotes a vector of graph moments (or cumulants), one for each subgraph in $\boldsymbol{g}$.

---

1. This is for simplicity of presentation; the tests we describe here naturally extend to networks with additional information, such as directed edges, weighted edges, and node attributes (Gunderson and Bravo-Hermsdorff, 2020).

2. Again, this assumption may be relaxed.

The sample estimators of these quantities are given a hat (e.g., $\widehat{\boldsymbol{\mu}}(\boldsymbol{G})$ and $\widehat{\boldsymbol{\kappa}}(\boldsymbol{G})$, see Section 6). Expectation is denoted by angled brackets (e.g., for the unbiased estimators of cumulants $\langle\widehat{\boldsymbol{\kappa}}(\boldsymbol{G})\rangle_{\boldsymbol{G}\sim\mathcal{G}} = \boldsymbol{\kappa}(\mathcal{G})$, see Section 7.1.2).

5 Two Statistics based on Subgraphs (What we Measure)

We first define graph moments (Section 5.1), the statistics to which we compare our new method. Then we describe how to convert these to graph cumulants (Section 5.2), the statistics used by our new method.

5.1 Graph Moments (The Typically-used Statistics)

When discussing counts of a subgraph g in some larger graph G, it is important to distinguish between *induced* counts and *homomorphism* counts (Chen et al., 2008); here we are using the latter. Another important distinction is that we are using *injective* homomorphism counts. To obtain these counts, first consider all mappings from the nodes of g to the nodes of G.

Injective refers to a condition on these node mappings: consider only those mappings that do not send different nodes in g to the same node in G.

Homomorphism refers to how we decide which of those (injective) mappings are "counts": those for which G has edges at all the locations where there are edges in g (i.e., it is still a count even if there are additional edges in G).

Out of all injective mappings from the nodes of g to the nodes of G, the fraction that are "counts" is known as the injective homomorphism *density* (Lovász, 2012). These densities are the graph moments of a graph G, denoted by $\mu_g(G)$.

5.2 Graph Cumulants (The New Statistics)

Graph cumulants were recently introduced by Gunderson and Bravo-Hermsdorff (2020). We first review the defining features of cumulants, highlighting a less-well-known combinatorial definition (Section 5.2.1). We then describe the analogue for graphs (Section 5.2.2).

5.2.1 Combinatorial Cumulants (Background)

First, consider a scalar-valued random variable $X \in \mathbb{R}$ sampled from some distribution $\mathcal{X}$. The rth-order moment of $\mathcal{X}$ is the expectation of the rth power of X: $\mu_r(\mathcal{X}) = \langle X^r\rangle_{X\sim\mathcal{X}}$. These moments may be combined into certain polynomial expressions, known as the cumulants κ_r (e.g., mean, variance, skew, kurtosis, etc.).

Cumulants have a uniquely defining property related to sums of independent random variables (Rota and Shen, 2000): the cumulants of the resulting sum are equal to the sum of the cumulants of those independent random variables (e.g., $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$, when X and Y are independent). This is essentially the reason behind the central limit theorem and the ubiquity of the Gaussian distribution (Gnedenko and Kolmogorov, 1954; Hald, 2000).

While the classical cumulants are often defined via the cumulant generating function, they also have an equivalent combinatorial definition in terms of a Möbius transform (see Section 12) (Kardar, 2007; McCullagh, 1987; Lehner, 2004). Namely, the rth moment can

be expressed as a sum of cumulants of order r and lower, corresponding to all partitions of r unique elements (see Figure 1, top three rows). In particular, for our scalar-valued example:

$$\mu_r(\mathcal{X}) = \sum_{\pi \in P_r} \prod_{b \in \pi} \kappa_{|b|}(\mathcal{X}), \tag{1}$$

where μ_r is the rth moment, κ_r is the rth cumulant, P_r is the set of all partitions of r unique elements (i.e., a set of non-overlapping subsets, or "blocks", whose union contains all r elements), π is one such partition, b is a block in partition π, and $|b|$ is the number of elements in block b.

Equation 1 may be rearranged to obtain expressions for the cumulants in terms of moments. For example, the third (classical) cumulant (the "skew") is: $\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$. This combinatorial definition makes the generalization to random variables with additional structure, such as graphs, more transparent.

5.2.2 GRAPH CUMULANTS (DEFINITION)

Before describing graph cumulants, it is worth mentioning a subtle point. Notice that Equation 1 relates the moments and cumulants of the *distribution* $\mathcal{X}$ (and *not* of any finite sample $\mathbf{X}$). While this distinction is somewhat pedantic for graph moments (see Section 7.1.1), the combinatorial definition of cumulants (classical or graphical) should only be applied to distributions.

The moments and cumulants of real-valued random variables are indexed by their order $r \in \mathbb{N}$. For graph-valued random variables, moments (Lovász and Szegedy, 2010; Bickel et al., 2011) and cumulants (Gunderson and Bravo-Hermsdorff, 2020) are now indexed by subgraphs $g \in \boldsymbol{g}$, with order given by the number of edges in g.

To apply the combinatorial definition (Equation 1) to networks, the partitioning P_r of the edges of the subgraphs must respect their connectivity (see Figure 1, bottom row), i.e.:

$$\mu_g(\mathcal{G}) = \sum_{\pi \in P_{E(g)}} \prod_{b \in \pi} \kappa_{g_b}(\mathcal{G}), \tag{2}$$

where $E(g)$ is the set of edges forming subgraph g, $P_{E(g)}$ is the set of all partitions of these edges, and g_b is the subgraph formed by the edges in b.

Again, these may be rearranged to obtain the graph cumulants $\kappa_g(\mathcal{G})$ in terms of the graph moments $\mu_g(\mathcal{G})$. For example, the (3rd-order) graph cumulant associated with the path graph with 3 edges is:

$$\kappa_{\sqcap}(\mathcal{G}) = \mu_{\sqcap}(\mathcal{G}) - 2\mu_{\wedge}(\mathcal{G})\mu_{/}(\mathcal{G}) - \mu_{/\!\!/}(\mathcal{G})\mu_{/}(\mathcal{G}) + 2\mu_{/}^3(\mathcal{G}). \tag{3}$$
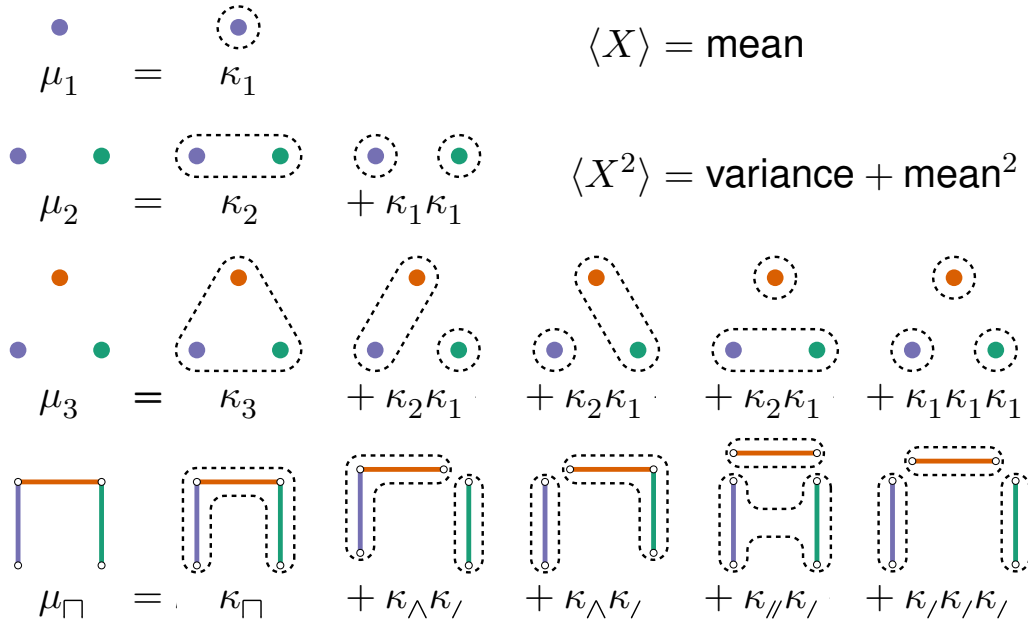
Figure 1: **To expand a graph moment μ_g in terms of graph cumulants,
enumerate all partitions of the edges forming subgraph g.**
The top three rows illustrate the combinatorial expansion of the first three
classical moments in terms of cumulants (Equation 1). Analogously, the bottom
row shows how to expand the graph moment $\mu_\sqcap$ in terms of graph cumulants
(Equation 2). The last term $(\kappa_/^3)$ corresponds to partitioning this subgraph
into three subsets, each with a single edge. The first term $(\kappa_\sqcap)$ corresponds
to "partitioning" this subgraph into a single subset with 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 $\kappa_\wedge\kappa_/$ terms), the subset with two edges has those edges sharing a node;
and in one case (the $\kappa_{/\!/}\kappa_/$ term), the subset with two edges has those edges not
sharing any node.

6 The Statistical Tests (How we Compare Graphs)

In this section, we describe the ways in which we ensured that the tests use the statistics
μ_g and κ_g in the same way, so as to compare graph cumulants on an equal footing with
the corresponding subgraph densities/moments. In particular, both tests have the same
structure (Section 6.1), rely on the same assumptions (Section 6.2), and have the same
computational complexity (Section 6.3).

6.1 Same Structure

To compare graph moments and graph cumulants on the same footing, we use them as analogous inputs to the same simple statistical test:

1. choose r, the maximum order of the subgraph statistics being considered;

2. for each sample, estimate its distribution using these subgraph statistics;

3. quantify the difference between distributions
   using a notion of distance for the space of these subgraph statistics.

In particular, we consider the graph moments/cumulants associated with $\boldsymbol{g}_r^{(\mathrm{c})}$, the set of connected subgraphs with at most r edges. To measure the difference between two samples of graphs $\mathbf{G}_A$ and $\mathbf{G}_B$, we use the squared Mahalanobis distance (Mahalanobis, 1936) between their inferred moments/cumulants:

$$\widehat{d_\mu^2}(\boldsymbol{G}_A, \boldsymbol{G}_B) = \left(\widehat{\boldsymbol{\mu}}_A - \widehat{\boldsymbol{\mu}}_B\right)^\top \left(\widehat{\boldsymbol{\Sigma}}_A^{(\mu)} + \widehat{\boldsymbol{\Sigma}}_B^{(\mu)}\right)^{-1} \left(\widehat{\boldsymbol{\mu}}_A - \widehat{\boldsymbol{\mu}}_B\right), \tag{4}$$

$$\widehat{d_\kappa^2}(\boldsymbol{G}_A, \boldsymbol{G}_B) = \left(\widehat{\boldsymbol{\kappa}}_A - \widehat{\boldsymbol{\kappa}}_B\right)^\top \left(\widehat{\boldsymbol{\Sigma}}_A^{(\kappa)} + \widehat{\boldsymbol{\Sigma}}_B^{(\kappa)}\right)^{-1} \left(\widehat{\boldsymbol{\kappa}}_A - \widehat{\boldsymbol{\kappa}}_B\right). \tag{5}$$

where $\widehat{\boldsymbol{\mu}}_A = \widehat{\boldsymbol{\mu}}(\boldsymbol{G}_A)$ is the vector of estimated moments of sample $\boldsymbol{G}_A$, and $\widehat{\boldsymbol{\Sigma}}_B^{(\kappa)}$ is the covariance estimate of the vector of estimated cumulants of sample $\boldsymbol{G}_B$, etc.

In Section 7, we describe how to compute these estimators.

6.2 Same Assumptions

We assume that the s observed graphs in a sample $\boldsymbol{G}$ were all obtained by subsampling n nodes i.i.d. from a single (much larger) "graph" $\mathcal{G}$. In other words, we assume that each graph distribution is generated by an underlying graphon (Borgs et al., 2008; Lovász, 2012). This assumption allows for several notable simplifications (in particular, Equation 8) without changing the primary message.

In addition, the graph distributions induced by a graphon are dense, and have empirical subgraph densities that converge to a Gaussian distribution as $n \to \infty$ (Bickel et al., 2011). As the empirical estimators of graph cumulants are linear combinations of the empirical subgraph densities (see Section 10), they also limit to a Gaussian distribution. Moreover, as the number of graphs per sample becomes large ($s \to \infty$), the averages will also approach a Gaussian distribution by the central limit theorem (see also Appendix A).

In light of these Gaussian limits, the well-established Mahalanobis distance (Anderson, 1958; Reiser, 2001) is appropriate for both tests considered in this paper. In Sections 8, 9, and 10, we show that (in contrast to moments) graph cumulants work well in practice, even when using this simple classical distance.

6.3 Same Complexity

The computational complexity of both tests is determined by the complexity of counting the relevant subgraphs. Computing the covariance matrices in Equations 4 and 5 requires counting connected subgraphs with at most $2r$ edges (see Section 7.2).

For this paper, we use $r = 3$. To count all but the complete graph on 4 nodes (K_4), we follow an approach similar to that in Maugis et al. (2016). Beginning with the observed

adjacency matrix, we require $n \times n$ matrix multiplication and entry-wise operations (thereby taking $\mathcal{O}(n^\omega)$ time, where ω is the matrix multiplication exponent). For K_4, we require multiplying an $n \times m$ and an $m \times n$ matrix, where m is the number of edges.

Our approach is not the best one can do. Efficiently counting subgraphs is a highly active area of research (Ribeiro et al., 2021), with a variety of powerful algorithms (Ye et al., 2022; Pinar et al., 2017), especially for certain (realistic) classes of graphs (Bera et al., 2019; Chiba and Nishizeki, 1985; Bressan, 2021), and for particular subgraphs (e.g., triangles (Gui et al., 2019)). Asymptotics aside, substantial acceleration can be obtained through parallel computation (Dodeja et al., 2022; Biswas et al., 2020), and more recently through the use of machine learning techniques (Zhao et al., 2021; Liu et al., 2020).

7 How to Estimate the Statistics (What we Compute)

In Section 7.1, we describe unbiased estimators for the graph moments and cumulants, highlighting an important relationship between subgraph densities in the process (Equation 8). In Section 7.2, we build on this idea to describe the typical fluctuations of such sample statistics in terms of their covariance matrices.

7.1 Obtaining Unbiased Estimators (Getting the Mean)

7.1.1 FOR GRAPH MOMENTS (SIMPLE SUBSTITUTION)

For a graphon model, graph moments are preserved under node subsampling:

$$\left\langle \mu_g(G) \right\rangle_{G \sim \mathcal{G}} = \mu_g(\mathcal{G}). \tag{6}$$

Thus, the empirical graph moments $\widehat{\boldsymbol{\mu}}(G)$ are themselves unbiased estimators of the moments of the distribution $\boldsymbol{\mu}(\mathcal{G})$ from which they were sampled. In particular, for a sample of s graphs:

$$\widehat{\mu}_g(\boldsymbol{G}) = \frac{1}{s} \sum_{i=1}^{s} \mu_g(G_i). \tag{7}$$

7.1.2 FOR GRAPH CUMULANTS (SLIGHTLY SUBTLE)

To obtain the analogous estimators for graph cumulants, we must be slightly more careful, as products of graph moments are not preserved in expectation under node subsampling. Fortunately, for graphs sampled from a graphon, products of graph moments are equal to the moment of their disjoint union[3] (Lovász, 2012; Maugis et al., 2020):

$$\mu_g(\mathcal{G})\mu_{g'}(\mathcal{G}) = \mu_{g \cup g'}(\mathcal{G}). \tag{8}$$

For example, using this relation in the expression for the cumulant associated with the path graph with 3 edges (Equation 3) results in its unbiased estimator:

$$\widehat{\kappa}_{\sqcap}(G) = \mu_{\sqcap}(G) - 2\mu_{\wedge\!\!\!\wedge}(G) + \mu_{/\!/\!/}(G). \tag{9}$$

---

3. Intuitively, nodes in disjoint components have nothing to do with each other, so the presence of edges in any component are independent from the presence of edges in the other.

For a sample of s graphs, the estimated cumulants of the underlying distribution are the average of the unbiased estimators of each individual graph:

$$\widehat{\kappa}_g(\boldsymbol{G}) = \frac{1}{s} \sum_{i=1}^{s} \widehat{\kappa}_g(G_i). \tag{10}$$

7.2 Analytically Computing Significance (Getting the Covariance)

Consider sampling graphs from a distribution $\mathcal{G}$ with known graph moments. The covariance between their observed graph moments is

$$\mathrm{Cov}\big(\mu_g(G), \mu_{g'}(G)\big) = \big\langle \mu_g(G)\mu_{g'}(G)\big\rangle - \big\langle \mu_g(G)\big\rangle\big\langle \mu_{g'}(G)\big\rangle. \tag{11}$$

The last term is trivial; as graph moments are preserved under node subsampling (Equation 6), the expectations of the empirical moments are exactly the moments of the underlying distribution. The first term, however, is the expectation of a *product* of graph moments.

Equation 8 applies to a single graphon $\mathcal{G}$. The analogous expression for the moments of a single finite graph G requires considering all the ways that the nodes of the subgraphs could overlap (Maugis et al., 2020). Thus, estimating the covariance of the empirical graph moments/cumulants up to order r requires the estimation of the distribution's graph moments/cumulants up to order $2r$ (as stated in Section 6.3). For example, in terms of injective homomorphism counts, we have the following expression:

$$c_{\wedge}(G)c_{/}(G) = 4c_{\wedge}(G) + 2c_{\triangle}(G) + 2c_{\lambda}(G) + 4c_{\sqcap}(G) + c_{/\!\!\wedge}(G). \tag{12}$$

The $4c_{\wedge}$ term corresponds to the four ways that the two nodes of the edge can be placed to overlap with the wedge ($\wedge$), the $2c_{\triangle}$ term corresponds to the two ways those nodes can be placed to turn a wedge into a triangle, and so on. Converting from counts to moments and taking the expectation allows one to express Equation 11 in terms of moments of the distribution.

As the s graphs in each sample are i.i.d. and have the same number of nodes, the covariance of the sample graph moments is

$$\widehat{\boldsymbol{\Sigma}}^{(\mu)}(\boldsymbol{G}) = \frac{1}{s} \sum_{i=1}^{s} \widehat{\boldsymbol{\Sigma}}^{(\mu)}(G_i). \tag{13}$$

As the unbiased graph cumulants are linear combinations of the sample graph moments, obtaining their covariance follows *mutatis mutandis*:

$$\widehat{\boldsymbol{\Sigma}}^{(\kappa)}(\boldsymbol{G}) = \frac{1}{s} \sum_{i=1}^{s} \widehat{\boldsymbol{\Sigma}}^{(\kappa)}(G_i). \tag{14}$$

8 A Controlled Competition (Between the Tests)

We first describe the models we use to generate the synthetic data. We then describe how we summarize the quality of the statistical tests, with illustrations for some representative simulations. Asymptotic results as $s \to \infty$ are presented in Appendix A.

8.1 Synthetic Data (A Pair of two-by-two SBMs)

We consider two graph distributions: one with heterogeneous degree distribution (but no community structure), and the other with community structure (but homogeneous degree distribution). Both distributions are Stochastic Block Models (SBMs) (Holland et al., 1983) with two equal-sized communities and expected edge density ρ.

The heterogeneous SBM is parameterized by ε_h, with connectivity matrix:

$$\mathbf{B}_h = \rho \times \begin{bmatrix} 1 + \varepsilon_h & 1 \\ 1 & 1 - \varepsilon_h \end{bmatrix}, \tag{15}$$

where $\varepsilon_h = 0$ gives uniform connection probability between all pairs of nodes, and $\varepsilon_h = 1$ gives zero connection probability within one of the two communities.

The assortative SBM is parameterized by ε_a, with connectivity matrix:

$$\mathbf{B}_a = \rho \times \begin{bmatrix} 1 + \varepsilon_a & 1 - \varepsilon_a \\ 1 - \varepsilon_a & 1 + \varepsilon_a \end{bmatrix}, \tag{16}$$

where $\varepsilon_a = 0$ again gives uniform connection probability between all pairs of nodes, and $\varepsilon_a = 1$ gives zero connection probability between the two communities.

After fixing two such SBMs, each instantiation of a two-sample test involves "flipping two coins" to decide which distributions each of the two samples will come from. That is, the instantiations are split evenly between "different distributions" and "same distribution", with the latter split evenly between the two SBMs.

8.2 ROC and AUC Curves (How we Compare Tests)

After computing the squared Mahalanobis distance between a pair of samples (Equations 4 and 5), we use a threshold to classify them as coming from the "same distribution" or "different distributions." Each choice of threshold induces: a rate of false positives (incorrectly concluding that the distributions are different), and a rate of true positives (correctly concluding that the distributions are different). All possible threshold choices are summarized in a Receiver Operating Characteristic (ROC) curve (see Figure 2).
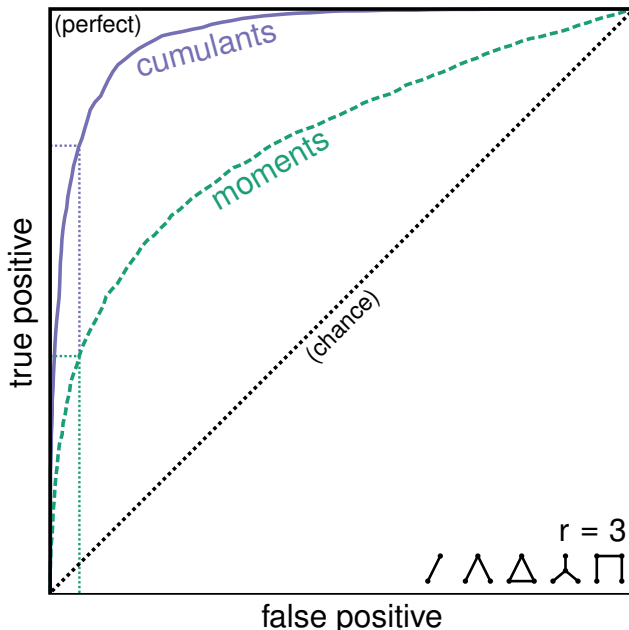
Figure 2: **The statistical power for all possible error rates.**
Known as an ROC curve, this plot visualizes the possible rates of false positives (Type I errors) and false negatives (Type II errors) of a binary classification method. Often, one specifies a maximum rate of false positives, commonly known as α. The vertical dotted lines in the lower left illustrate the canonical choice of $\alpha = 0.05$, and the point where this vertical line meets the ROC curves gives the corresponding rate of true positives (i.e., the statistical power, or $1 - $ (Type II error rate)). An ROC curve displays the results for all possible values of α: random guesses result in a line along the diagonal, while perfect answers result in a line that hugs the upper left. For this plot, both statistical tests (**moments** and **cumulants**) use the counts of connected subgraphs with up to $r = 3$ edges to distinguish between samples from two graph distributions: a heterogeneous SBM (Equation 15 with $\varepsilon_h = \frac{1}{16}$), and an assortative SBM (Equation 16 with $\varepsilon_a = \frac{1}{16}$), both with density $\rho = \frac{1}{2}$. Each sample contains $s = 4$ graphs, and all graphs have $n = 256$ nodes.

To compare ROC curves, we summarize them with a single scalar value, viz., the Area Under the Curve (AUC). Figure 3 compares the AUC for graphs with $n = 128$ and $n = 256$ nodes over a range of sample sizes. The use of graph cumulants consistently results in greater statistical power, especially when the number of graphs per sample is small (including the case of a *single* graph per sample, for which the test using moments is not applicable).

One might wonder if these results are sensitive to the model parameters ε_h and ε_a. While the values used to create the figures were judiciously chosen (e.g., such that the AUC spans the range from chance to perfect), the qualitative conclusions remain robust

for (essentially) any parameter choice. In the next section, we show that the use of graph cumulants is similarly promising for applications to more realistic networks.
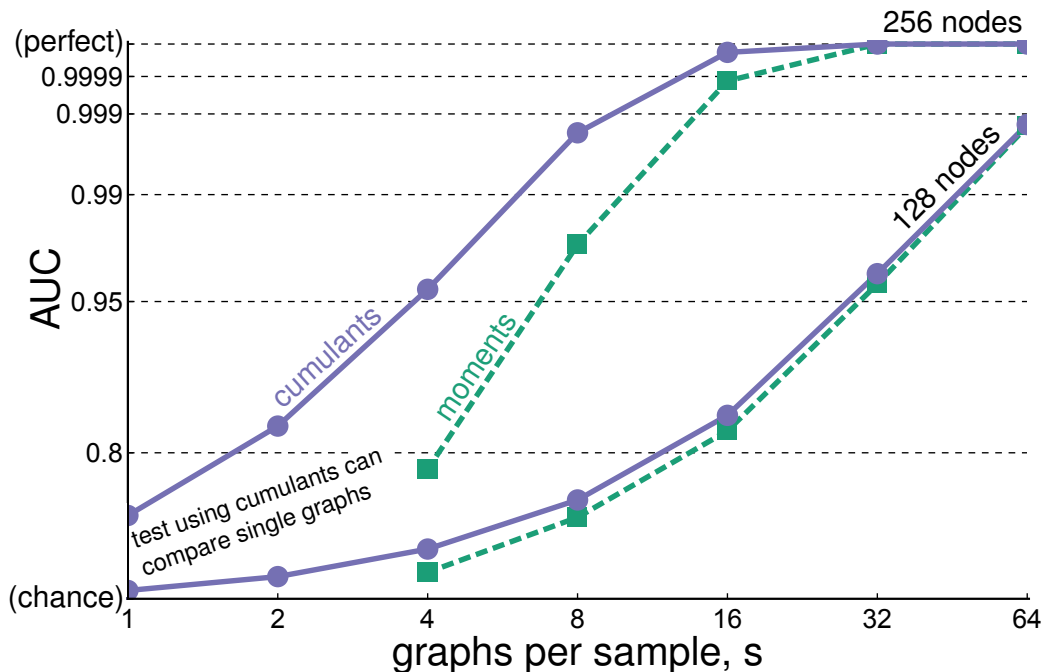

Figure 3: **Using graph cumulants outperforms using graph moments, particularly when the number of graphs per sample is small.**
We performed the comparison in Figure 2 for a range of graphs per sample s, for graphs with $n = 128$ nodes and graphs with $n = 256$ nodes. And we summarized the resulting ROCs in terms of the the Area Under the Curve (AUC). The (significantly distorted) vertical axis ranges from 0.5 (chance) to 1.0 (perfect). The test using moments (with at most $r = 3$ edges) fails to give an answer for $s < 4$ (see Appendix B). In contrast, the test using cumulants works even when there is only a *single* observed graph in each sample (see Section 11).

9 Studying the Tests in the Wild (Application to Real Networks)

Most real-world networks look rather different than the simulated graphs in the previous section. In this section, we compare the two tests using graphs sampled from more realistic graph distributions. In particular, we use gene interaction networks from four different species: Mouse, Rat, Human, and Arabidopsis (a small flowering plant related to cabbage and mustard), all curated by the FunCoup repository (Persson et al., 2021).

Although these networks have relatively similar edge densities, the differences are already enough for either test to easily distinguish samples from them using only the edge subgraph. As the differences between moments and cumulants only become apparent for subgraphs with more than one edge, we adjusted the networks being compared so they have the same

edge density. This was done by removing the lowest-weight edges from the network with larger edge density until it had the same density as the other.[4] The effect of this adjustment can be seen in Figure 4 (left); both tests using $r = 1$ edge can no longer distinguish between the samples.

Each sample contains s graphs, all obtained from one of these adjusted networks by subsampling its nodes. Specifically, each graph G in a sample contains n nodes in expectation (selected from the N nodes in the adjusted network independently with probability $\frac{n}{N}$), and nodes in G are connected by an edge if (and only if) their corresponding nodes in the adjusted network are connected by an edge. That is, G is an induced subgraph of the adjusted network.

Figures 4 and 5 show that using graph cumulants to discriminate between these biological networks outperforms the analogous test using graph moments.
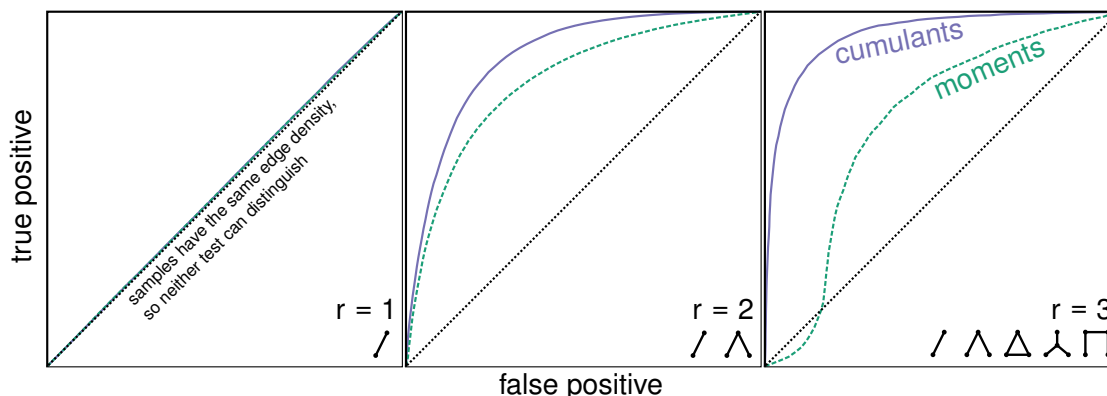

Figure 4: **Using more subgraphs increases accuracy for cumulants, but causes overfitting for moments.**
We use the two tests to compare the genetic interaction networks of Arabidopsis ($\sim 1.3 \times 10^4$ nodes, $\sim 7.9 \times 10^5$ edges) and Mouse ($\sim 1.4 \times 10^4$ nodes, $\sim 9.2 \times 10^5$ edges) (data from Persson et al. (2021)). As these networks were adjusted to have the same edge density (see Section 9), neither $r = 1$ test (left) can distinguish between them. The benefit of using cumulants can be seen for the tests using more subgraphs. When including subgraphs with $r = 2$ edges (middle), the test using cumulants consistently outperforms the test using moments. When including subgraphs with $r = 3$ edges (right), the test using cumulants continues to improve, while the test using moments performs worse. Each sample contained $s = 4$ graphs, and each graph had $n \sim 256$ nodes (see Section 9).

---

4. If one wishes to test if two distributions with different sparsities are otherwise the same, one could scale all subgraph densities by the edge density to the power of the number of edges in the subgraph.
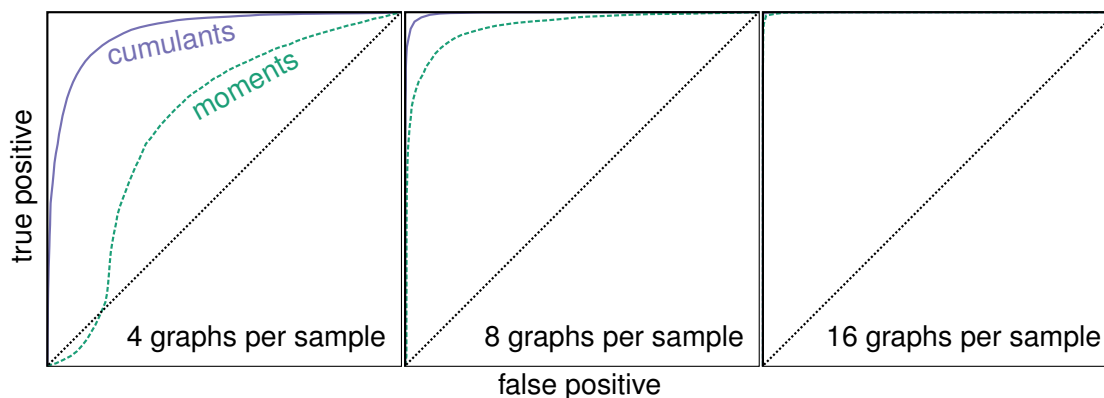
Figure 5: **Cumulants continue to consistently outperform moments in real data, especially when the number of graphs per sample is small.**
We compare the genetic interaction networks of Human ($\sim 1.6 \times 10^4$ nodes, $\sim 13.9 \times 10^5$ edges) and Rat ($\sim 1.1 \times 10^4$ nodes, $\sim 9.0 \times 10^5$ edges). The samples are generated with the same method as in Figure 4, and the tests use connected subgraphs with up to $r = 3$ edges.

10 Why Cumulants do Better

Why does using graph cumulants consistently provide better statistical power? After all, unbiased graph cumulants are simply linear combinations of graph moments! (Section 7.1.2) Loosely, the reason is that graph distributions "look more Gaussian" when represented in terms of cumulants (as compared to moments).

Both tests measure differences between distributions using the squared Mahalanobis distance (Equations 4 and 5). As this measure depends only on the mean and covariance, there is a tacit assumption that the sample statistics are well-characterized by a multivariate Gaussian (Rao, 1992).

Indeed, if the subgraph statistics *were* precisely Gaussian, and their covariance known *exactly*, then the squared Mahalanobis distance between pairs of samples from the same distribution would in fact follow a χ^2 distribution. However, as the sample statistics are *not* precisely Gaussian, and their covariance must be *estimated*, one should indeed expect some deviation from this limiting distribution.

As illustrated in Figure 6, when the graphs are sampled from the same distribution, the test using moments results in a distribution of squared Mahalanobis distances with overly-heavy tails. Whereas the distribution for the test using cumulants is well-characterized by the appropriate χ^2 distribution, thereby allowing one to control the false positive rate.
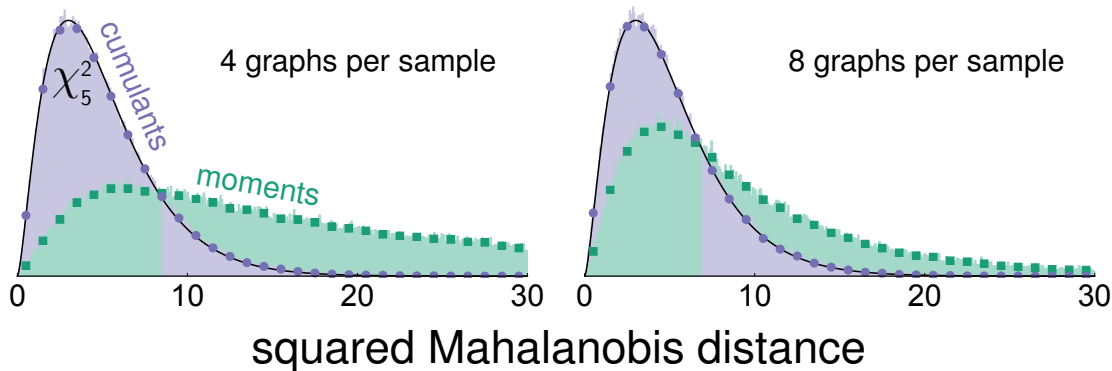
Figure 6: **The test statistic based on graph cumulants remains nearly χ^2,**
even when only few graphs are observed.
If a test statistic follows a known distribution under the null hypothesis, one
can control the false positive rate by choosing an appropriate threshold. Indeed,
for a large number of graphs per sample $s \to \infty$, both tests converge to a
χ^2 distribution. As the number of graphs per sample decreases, this becomes a
poor approximation for the test using moments. In contrast, the approximation
remains strikingly robust for the test using cumulants. Colored histograms are
the empirical distributions of the test statistics (squared Mahalanobis distance)
for simulations with the same parameters as in Figure 2, and black curves are
the limiting χ^2 distribution (with 5 degrees of freedom).

11 The Main Message (Discussion)

Perhaps the most salient advantage of using graph cumulants (instead of the subgraph
densities themselves) is the ability to compare distributions even when observing only a
single graph from each (e.g., Figure 3).

At first glance, it may seem strange that one could make inferences about a distribution
from a sample containing only a single "observation." Indeed, this certainly does not work
for scalar-valued random variables — a single observation provides no information about
the spread of its underlying distribution.

Essentially, this difference arises because, in graphs, the data reside in the edges, but
sampling is applied node-wise. Notationally, this manifests itself in our need to specify both
n (the number of nodes per graph) and s (the number of graphs per sample), whereas the
"quantity" of i.i.d. scalar data is specified by only the sample size.

Thus, there are two relevant limits: we may ask about the distributions of these statistics
as either n or s become sufficiently large. As $s \to \infty$ (with fixed n), both methods are
asymptotically normal by the central limit theorem. In Appendix A, we show that using
graph cumulants remains statistically advantageous even in this "classical" limit. As $n \to \infty$
(with fixed s), graph cumulants appear to more properly exploit the multiplicity of the data
within each graph, allowing for inference even when $s = 1$ (see Appendix B).

In a sense, graph cumulants provide a more "natural" set of coordinates than the subgraph densities themselves. For example, the edge density μ_I is strongly correlated with the wedge moment $\mu_\wedge$. In contrast, $\kappa_\wedge$ "takes into account" the lower-order effect of edge density, rendering it "more orthogonal" to μ_I. In fact, for an Erdős-Rényi distribution, the covariance between the edge density and all other graph cumulants $\widehat{\kappa}_g$ is precisely zero (see proof in Appendix C).

In practice, the covariance estimates for cumulants are more robust, leading to impressively accurate agreement with the classical χ^2 distribution. This allows one to convert these statistics into precise probabilistic statements (e.g.: "What is the likelihood that two graphs were generated by different graphons?"). This notion of "graphonic similarity" between any pair of graphs is a remarkably general tool.

12 Promising Sequels (Future Directions)

Towards the goal of estimating the relevance of various subgraphs/motifs in an observed network, recent works (Bhattacharya et al., 2022; Zhang and Xia, 2022) have provided higher-order characterizations of the sampling distributions of (appropriately rescaled) estimators of subgraph densities for various regimes and model assumptions. Given the success of graph cumulants compared to the bare densities/moments in determining network similarity, a natural step to help in this endeavor is to provide such characterizations for the estimators of graph cumulants (e.g., by generalizing the result in Appendix C to arbitrary graphons and pairs of subgraphs).

It might be argued that intelligent feature engineering has been rendered moot by powerful supercomputers fitting deep neural networks to big data. Nonetheless, some of the most notable breakthroughs in machine learning have been the result of architectures with cleverly engineered priors (e.g., convolutional neural networks for computer vision tasks (LeCun et al., 1998; Khan et al., 2018), and transformers for natural language processing tasks (Vaswani et al., 2017; Chan et al., 2022)). Like the edge and orientation detection cells in the first layers of visual processing (Hubel and Wiesel, 1959; Gabbiani and Cox, 2017), graph cumulants offer a principled first layer of contrastive features when looking at the relationships between edges in networks.

As a coda, we highlight a compelling analogy between the conversion between induced subgraph densities and injective homomorphism subgraph densities and the conversion between injective homomorphism subgraph densities (i.e., graph moments) and their corresponding graph cumulants. The former applies a Möbius transform to the poset induced by the inclusion of edges in a subgraph with a fixed number of nodes. The latter also applies a Möbius transform, though with respect to the partial order induced by the partitions of an arrangement of a fixed number of edges.

This suggests a general framework for the "cumulantification" of other types of combinatorial structures. Indeed, extensions to hypergraphs and directed networks were described in the Appendices of Gunderson and Bravo-Hermsdorff (2020), though the applications appear to be even more general — any combinatorial structure admitting a notion of "reduction" or "partitioning" naturally induces a corresponding poset over its substructures, offering an avenue for obtaining similarly useful statistics.

Acknowledgments and Disclosure of Funding

Appendix A. The Limit of Many Observed Graphs (Asymptotic Relative Efficiency of the Two Tests)

In Section 10, our simulations show that the test statistic based on graph cumulants closely approximates a χ^2 distribution under the null hypothesis even for a small number of observed graphs s, whereas the analogous test statistic based on graph moments does not. However, both statistics indeed do converge to a χ^2 distribution in the limit of many observed graphs $s \to \infty$ (Maugis et al., 2020).

In this section, we compare the statistical power of the two tests in this limit of large sample size, showing that using cumulants still outperforms using moments.

A.1 The Model

To this end, we consider the large sample size limit $s \to \infty$ as the distributions become increasingly similar $\mathcal{G}_1 \to \mathcal{G}_0$. We represent these distributions as stochastic block models (SBMs): a probability vector $\boldsymbol{\pi}$, giving the expected fractions of nodes contained in each community, and a connectivity matrix $\mathbf{B}$, giving the probability of an edge between two nodes based on their assigned communities.

Let $\mathcal{G}_1$ be a perturbation of the form $(\boldsymbol{\pi}_1, \mathbf{B}_1) = (\boldsymbol{\pi}_0, \mathbf{B}_0) + \sqrt{\gamma/s}\,(\delta\boldsymbol{\pi}, \delta\mathbf{B})$. The $1/\sqrt{s}$ scaling of this perturbation as $s \to \infty$ is such that, given some maximum allowable error rates of false positives α and false negatives β for a test, there exists a critical value γ_* above which these desiderata are achievable.

For two different tests, the ratio of their γ_* is known as the Pitman asymptotic relative efficiency (PARE) (Pitman, 1949). As the two tests we are comparing have the same $s \to \infty$ limiting distribution χ_5^2 (as we are considering $r = 3$), taking this ratio removes the dependence on α and β. Thus, the PARE is given by the asymptotic ratio of the squared Mahalanobis distances, and depends on two choices: the distribution $(\boldsymbol{\pi}_0, \mathbf{B}_0)$, and the perturbation $(\delta\boldsymbol{\pi}, \delta\mathbf{B})$.

For $(\boldsymbol{\pi}_0, \mathbf{B}_0)$, we choose a model that allow us to independently control assortativity and (degree) heterogeneity. Specifically, we blend the two previously mentioned SBMs (Equations 15 and 16), using the Kronecker product of their connectivity matrices $\mathbf{B}_h$ and $\mathbf{B}_a$, i.e.,

$$\mathbf{B}_0 = \rho \times \begin{bmatrix} (1+\varepsilon_h)(1+\varepsilon_a) & (1+\varepsilon_h)(1-\varepsilon_a) & (1+\varepsilon_a) & (1-\varepsilon_a) \\ (1+\varepsilon_h)(1-\varepsilon_a) & (1+\varepsilon_h)(1+\varepsilon_a) & (1-\varepsilon_a) & (1+\varepsilon_a) \\ (1+\varepsilon_a) & (1-\varepsilon_a) & (1-\varepsilon_h)(1+\varepsilon_a) & (1-\varepsilon_h)(1-\varepsilon_a) \\ (1-\varepsilon_a) & (1+\varepsilon_a) & (1-\varepsilon_h)(1-\varepsilon_a) & (1-\varepsilon_h)(1+\varepsilon_a) \end{bmatrix}, \tag{17}$$

and $\boldsymbol{\pi}_0$ is uniform over the $k = 4$ communities.

For $(\delta\boldsymbol{\pi}, \delta\mathbf{B})$, we choose random perturbations to be Gaussian with covariance proportional to typical fluctuations of realizations of this model $(\boldsymbol{\pi}_0, \mathbf{B}_0)$ for graphs with n nodes. Thus, $\delta\boldsymbol{\pi}$ has zero mean, with covariance given by the corresponding multinomial distribution:

$$\mathrm{Var}\big(\delta\boldsymbol{\pi}_i\big) = \frac{(\boldsymbol{\pi}_0)_i\big(1 - (\boldsymbol{\pi}_0)_i\big)}{n}, \quad \mathrm{Cov}\big(\delta\boldsymbol{\pi}_i, \delta\boldsymbol{\pi}_j\big) = -\frac{(\boldsymbol{\pi}_0)_i(\boldsymbol{\pi}_0)_j}{n}. \tag{18}$$

Likewise, $\delta\mathbf{B}$ has zero mean, with variance given by:

$$\mathrm{Var}\big(\delta B_{ii}\big) = \frac{(\mathbf{B}_0)_{ii}\big(1 - (\mathbf{B}_0)_{ii}\big)}{\binom{n}{2}(\boldsymbol{\pi}_0)_i^2}, \quad \mathrm{Var}\big(\delta B_{ij} = \delta B_{ji}\big) = \frac{(\mathbf{B}_0)_{ij}\big(1 - (\mathbf{B}_0)_{ij}\big)}{2\binom{n}{2}(\boldsymbol{\pi}_0)_i(\boldsymbol{\pi}_0)_j}. \tag{19}$$

A.2 The Computation

To evaluate the PARE, we construct a symmetric matrix $\mathbf{M}$ that yields the squared Mahalanobis distance of a random perturbation by evaluating its quadratic form with a vector $\boldsymbol{\eta}$ containing i.i.d. normal entries.

We first obtain the graph moments up to order $2r = 6$ of the distribution defined by $(\boldsymbol{\pi}_0, \mathbf{B}_0)$, and use these to compute the covariance of sample moments up to order $r = 3$ (for graphs with n nodes). We also obtain the derivatives $\mathbf{J}_{\partial\mu/\partial(\delta\boldsymbol{\pi}, \delta\mathbf{B})}$ of the moments up to order $r = 3$ with respect to the parameters defining $(\delta\boldsymbol{\pi}, \delta\mathbf{B})$. We then multiply both sides of the inverse of the covariance matrix with the matrix of partial derivatives:

$$\big(\mathbf{J}_{\partial\mu/\partial(\delta\boldsymbol{\pi}, \delta\mathbf{B})}\big)^\top \big(\boldsymbol{\Sigma}^{(\mu)}\big)^{-1} \big(\mathbf{J}_{\partial\mu/\partial(\delta\boldsymbol{\pi}, \delta\mathbf{B})}\big). \tag{20}$$

Taking the quadratic form of this matrix with a vector of the perturbations $(\delta\boldsymbol{\pi}, \delta\mathbf{B})$ gives the squared Mahalanobis distance (Equation 4).

Now we want a matrix $\mathbf{S}_{(\delta\boldsymbol{\pi}, \delta\mathbf{B})\leftarrow\boldsymbol{\eta}}$ that transforms a vector $\boldsymbol{\eta}$ of i.i.d. normal entries into a random vector of perturbations. For $\delta\mathbf{B}$, we only need to put the square root of the variances in Equation 19 along the corresponding diagonal of $\mathbf{S}$. For $\delta\boldsymbol{\pi}$, we need the individual realizations of the perturbations to sum to zero, so we put a projection matrix $(\mathbf{I} - \frac{1}{k}\mathbf{1})/\sqrt{nk}$ along the corresponding diagonal of $\mathbf{S}$.[5]

To obtain the desired matrix for moments $\mathbf{M}^{(\mu)}$, we multiply the matrix from Equation 20 on both sides by $\mathbf{S}$:

$$\mathbf{M}^{(\mu)} = \big(\mathbf{S}_{(\delta\boldsymbol{\pi}, \delta\mathbf{B})\leftarrow\boldsymbol{\eta}}\big)^\top \big(\mathbf{J}_{\partial\mu/\partial(\delta\boldsymbol{\pi}, \delta\mathbf{B})}\big)^\top \big(\boldsymbol{\Sigma}^{(\mu)}\big)^{-1} \big(\mathbf{J}_{\partial\mu/\partial(\delta\boldsymbol{\pi}, \delta\mathbf{B})}\big) \big(\mathbf{S}_{(\delta\boldsymbol{\pi}, \delta\mathbf{B})\leftarrow\boldsymbol{\eta}}\big). \tag{21}$$

The matrix for cumulants $\mathbf{M}^{(\kappa)}$ is defined analogously.

As the PARE is a ratio, it is natural to take the log before taking the expectation over the random perturbations. In particular, we can write the $\langle\log\mathrm{PARE}\rangle$ as follows:

$$\Big\langle \log\mathrm{PARE} \Big\rangle = \Big\langle \log\mathrm{PAE}^{(\kappa)} \Big\rangle - \Big\langle \log\mathrm{PAE}^{(\mu)} \Big\rangle$$
$$= \Big\langle \log\big(\boldsymbol{\eta}^\top \mathbf{M}^{(\kappa)} \boldsymbol{\eta}\big) \Big\rangle - \Big\langle \log\big(\boldsymbol{\eta}^\top \mathbf{M}^{(\mu)} \boldsymbol{\eta}\big) \Big\rangle,$$

---

5. For the case of uniform $\boldsymbol{\pi}_0 = \frac{1}{k}$, multiplying this projection matrix by k i.i.d. normal entries gives a vector of k values that sum to zero, with covariance given by Equation 18.

where expectation is taken over $\boldsymbol{\eta}$ having i.i.d. normal entries.

As the distribution for $\boldsymbol{\eta}$ is rotationally symmetric, the entries are i.i.d. normal for any orthogonal basis. In particular, we diagonalize the (positive semidefinite) matrices $\mathbf{M}^{(\mu)}$ and $\mathbf{M}^{(\kappa)}$. Thus, we only need their eigenvalues λ_i to compute the expected log PARE:

$$\left\langle \log \text{PARE} \right\rangle = \left\langle \log \sum_i \lambda_i^{(\kappa)} \eta_i^2 \right\rangle - \left\langle \log \sum_i \lambda_i^{(\mu)} \eta_i^2 \right\rangle. \tag{22}$$

At this point, we estimate $\langle \log \text{PARE} \rangle$ via Monte Carlo sampling, resulting in Figure 7.
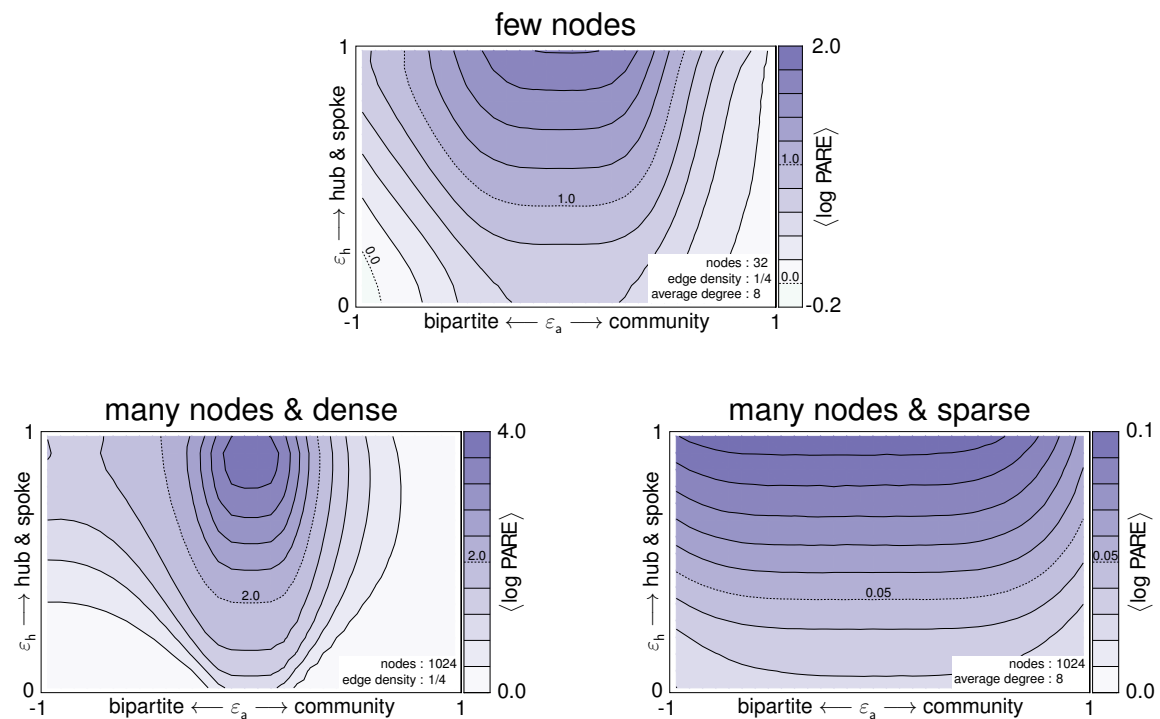
A.3 The Results

Figure 7: **The test using graph cumulants outperforms that using moments, even in the limit of many observed graphs.**
Here, we compare the asymptotic efficiencies of the two tests (with $r = 3$) in the limit $s \to \infty$. The contours correspond to the $\langle \log \text{PARE} \rangle$ for random perturbations to the SBM with $k = 4$ equal-sized communities (Equation 17). The test using cumulants does better than that using moments in nearly all regimes (aside from a small region for 32 nodes).

Appendix B. The Limit of Few Observed Graphs (Singular Covariance Estimates)

Below a certain number of observed graphs, the test using moments results in singular covariance matrices. If the sum of the two covariance matrices is also singular, one cannot take its inverse, so the Mahalanobis distance (Equation 4) is ill-defined. Here, we illustrate the simplest example: when there is only $s = 1$ graph per sample, the distributions inferred by the test using moments have zero variance in the edge density $\mu_/$.

Consider a single observed graph G_1 with n nodes and m edges. For the test using moments, the expected subgraph densities of the inferred distribution are taken to be those of this observed graph: $\mu_g(\mathcal{G}) = \mu_g(G_1)$. The estimation of the variance of $\mu_/$ assumes that the graph moments match to second order (and, as n is fixed, their expected counts as well). At first order, we have

$$\langle c_/(G) \rangle_{G \sim \mathcal{G}} = c_/(G_1)$$
$$= 2m. \tag{23}$$

And at second order, we have

$$\langle c_/^2(G) \rangle_{G \sim \mathcal{G}} = \langle 2c_/(G) + 4c_\wedge(G) + c_{/\!/}(G) \rangle_{G \sim \mathcal{G}}$$
$$= 2c_/(G_1) + 4c_\wedge(G_1) + c_{/\!/}(G_1)$$
$$= (2m)^2. \tag{24}$$

The *only* distributions satisfying both of these constraints are those containing graphs with precisely m edges,[6] and thus the variance in edge density is zero: $\mathrm{Var}(\mu_/) = 0$.

In a sense, this can be thought of as a $1{\times}1$ covariance matrix of rank 0. In general, as the number of graphs per sample s increases, the rank of the covariance matrix $\boldsymbol{\Sigma}^{(\mu)}$ does as well. For a given order r, one has covariance matrices of size $|\boldsymbol{g}|{\times}|\boldsymbol{g}|$ (where $|\boldsymbol{g}|$ is the number of connected subgraphs with r edges), and therefore requires sufficiently many graphs per sample s such that the rank of the sum of the two inferred covariance matrices is no less than $|\boldsymbol{g}|$. This is the reason behind the $s < 4$ cutoff for the test using moments in Figure 5.

We remark that, even with sufficient sample size, the covariance matrix may be singular (if a subset of the sample exhibits coincidental collinearity or any of its higher-dimensional analogues). To handle these infrequent cases in a consistent way, we use the *pseudo*inverse in Equations 4 and 5.

---

6. In Equations 23 and 24, we have $2m$ edges instead of m because we are considering (injective) homomorphism counts, which count both orientations of each edge.

Appendix C. Zero covariance between the edge density and the other graph cumulants for Erdős-Rényi distributions.

Theorem 1 *For any Erdős-Rényi distribution* $\mathrm{ER}(n,p)$, *the covariance between the edge density and all other unbiased estimators of graph cumulants* $\widehat{\kappa}_g$ *is zero.*

Proof We start by obtaining an expression for the covariance $\mathrm{Cov}^{(\mathrm{ER})}\big(\widehat{\mu}_/(G), \widehat{\mu}_g(G)\big)$ between the edge density[7] and the moment of any subgraph g, for graphs G sampled from an Erdős-Rényi distribution $\mathrm{ER}(n,p)$, where n is the number of nodes in G and p is the i.i.d. probability of an edge between any pair of nodes.

Let v denote the number of nodes in the subgraph g, and e its number of edges. Recall that to compute the covariance between two subgraphs (Equation 11), one must consider all the ways that the nodes of the subgraphs could overlap (Section 7.2). Here, there are four cases to consider:

1. Neither node in the edge maps to any node in g.
   There is a single way this can happen, and it results in
   a subgraph with $v + 2$ nodes and $e + 1$ edges.

2. A single node in the edge maps to a single node in g.
   There are $2v$ ways this can happen, each resulting in
   a subgraph with $v + 1$ nodes and $e + 1$ edges.

3. The two nodes of the edge map to two nodes in g that do not already have an edge.
   There are $2\big[\binom{v}{2}v - e\big] = v(v-1) - 2e$ ways this can happen, each resulting in
   a subgraph with v nodes and $e + 1$ edges.

4. The two nodes of the edge map to two nodes in g that already have an edge.
   There are $2e$ ways that this can happen, each resulting in
   a subgraph with v nodes and e edges.

For an $\mathrm{ER}(n,p)$ distribution, the moment of a subgraph g is p^e (i.e., p to the power of the number of edges in g). Thus, the moments of the resulting subgraphs in cases 1, 2, and 3 are p^{e+1}, while the moments of the resulting subgraphs in case 4 are p^e. Moreover, recall that the injective homomorphism counts $c_g(G)$ of a subgraph g (with v nodes) into a graph G (with n nodes) is $n^{\underline{v}}\mu_g(G)$, where the falling factorial is $n^{\underline{v}} = \prod_{j=0}^{v-1}(n-j)$.

We now substitute these results into the expression for the covariance:

$$
\begin{aligned}
\mathrm{Cov}^{(\mathrm{ER})}\big(\widehat{\mu}_/(G), \widehat{\mu}_g(G)\big) &= \big\langle\widehat{\mu}_/(G)\widehat{\mu}_g(G)\big\rangle_{G\sim\mathrm{ER}(n,p)} - \big\langle\widehat{\mu}_/(G)\big\rangle_{G\sim\mathrm{ER}(n,p)}\big\langle\widehat{\mu}_g(G)\big\rangle_{G\sim\mathrm{ER}(n,p)} \\
&= \big\langle\widehat{\mu}_/(G)\widehat{\mu}_g(G)\big\rangle_{G\sim\mathrm{ER}(n,p)} - \big(p\big)\big(p^e\big) \\
&= \frac{1}{n^{\underline{v}}n^{\underline{2}}}\bigg[\big(1\big)\big(n^{\underline{v+2}}\big)\big(p^{e+1}\big) + \big(2v\big)\big(n^{\underline{v+1}}\big)\big(p^{e+1}\big) \\
&\qquad + \big(v(v-1) - 2e\big)\big(n^{\underline{v}}\big)\big(p^{e+1}\big) + \big(2e\big)\big(n^{\underline{v}}\big)\big(p^e\big)\bigg] - p^{e+1}
\end{aligned}
$$

---

7. Recall that the edge density is both the first graph moment and the first graph cumulant, i.e.,
$\mu_/(\mathcal{G}) = \kappa_/(\mathcal{G}) = \big\langle\widehat{\mu}_/(\boldsymbol{G})\big\rangle_{\boldsymbol{G}\sim\mathcal{G}} = \big\langle\widehat{\kappa}_/(\boldsymbol{G})\big\rangle_{\boldsymbol{G}\sim\mathcal{G}}$.

Using "algebra autopilot" and grouping the terms, we get:

$$\frac{1}{n^2}\left[\left((n-v)(n-(v+1))+2v(n-v)+\left(v(v-1)-2e\right)-n^2\right)p^{e+1}+(2e)p^e\right]$$

$$=\frac{1}{n^2}\left[\left(n^2-2nv-n+v^2+v+2nv-2v^2+v^2-v-2e-n^2+n\right)p^{e+1}+(2e)p^e\right]$$

$$=\frac{1}{n^2}\left[(-2e)p^{e+1}+(2e)p^e\right]=\frac{1}{n^2}2ep^e\left(1-p\right)$$

$$\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\mu}_g\right)\;=\;\underbrace{p\left(1-p\right)/\binom{n}{2}}_{\text{"variance of }\widehat{p}\text{"}}\;\times\;\underbrace{ep^{e-1}}_{\text{"}d(p^e)/dp\text{"}} \tag{25}$$

Crucially, for an $\mathrm{ER}(n,p)$ distribution, the covariance between the edge density and the density of a subgraph g depends *only* on the number of edges in g.

The $\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\kappa}_/(G),\widehat{\kappa}_g(G)\right)\equiv\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\kappa}_g\right)$ is given by the sum of the covariances of the edge density with each subgraph g' that appears in the expression for the unbiased cumulant of g (weighted by its coefficient in the formula). For example:

$$\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\kappa}_\wedge\right)=\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\mu}_\wedge\right)-\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\mu}_{/\!/}\right);$$

$$\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\kappa}_\triangle\right)=\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\mu}_\triangle\right)-3\,\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\mu}_{/\!\wedge}\right)+2\,\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\mu}_{/\!/\!/}\right);$$

$$\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\kappa}_\sqcap\right)=\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\mu}_\sqcap\right)-2\,\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\mu}_{/\!\wedge}\right)+\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\mu}_{/\!/\!/}\right).$$

We now note two key facts about the expressions for the unbiased graph cumulants: *1)* the subgraphs all have the same number of edges, and *2)* the coefficients always sum to zero. Hence, $\mathrm{Cov}^{(\mathrm{ER})}\left(\widehat{\mu}_/,\widehat{\kappa}_g\right)=0$ for any other subgraph g, as desired. ∎

References

Maryam Aliakbarpour, Amartya Shankha Biswas, Themis Gouleakis, John Peebles, Ronitt Rubinfeld, and Anak Yodpinyanee. Sublinear-time algorithms for counting star subgraphs via edge sampling. *Algorithmica*, 80(2):668–697, 2018.

Uri Alon. Network motifs: Theory and experimental approaches. *Nature Review Genetics*, 8(6):450–461, 2007.

Weihua An. Fitting ERGMs on big networks. *Social Science Research*, 59:107–119, 2016.

Theodore Wilbur Anderson. *An introduction to multivariate statistical analysis*. Wiley, 1958.

Dena Asta and Cosma Rohilla Shalizi. Geometric network comparison. *arXiv:1411.1350*, 2014.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

Suman K Bera, Noujan Pashanasangi, and C Seshadhri. Linear time subgraph counting, graph degeneracy, and the chasm at size six. *arXiv:1911.05896*, 2019.

Bhaswar B Bhattacharya, Sayan Das, and Sumit Mukherjee. Motif estimation via subgraph sampling: The fourth-moment phenomenon. *The Annals of Statistics*, 50(2):987–1011, 2022.

Sharmodeep Bhattacharyya and Peter J Bickel. Subsampling bootstrap of count features of networks. *The Annals of Statistics*, 43(6):2384–2411, 2015.

Peter J Bickel, Aiyou Chen, and Elizaveta Levina. The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5):2280–2301, 2011.

Amartya Shankha Biswas, Talya Eden, Quanquan C Liu, Slobodan Mitrović, and Ronitt Rubinfeld. Massively parallel algorithms for small subgraph counting. *arXiv:2002.08299*, 2020.

Christian Borgs, Jennifer T Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851, 2008.

Gecia Bravo-Hermsdorff and Lee M Gunderson. A unifying framework for spectrum-preserving graph sparsification and coarsening. *Neural Information Processing Systems (NeurIPS)*, 33, 2019.

Marco Bressan. Faster algorithms for counting subgraphs in sparse graphs. *Algorithmica*, 83(8):2578–2605, 2021.

George Casella and Roger L Berger. *Statistical Inference*. Cengage Learning, 2021.

Stephanie CY Chan, Adam Santoro, Andrew K Lampinen, Jane X Wang, Aaditya Singh, Pierre H Richemond, Jay McClelland, and Felix Hill. Data distributional properties drive emergent few-shot learning in transformers. *arXiv:2205.05055*, 2022.

Li Chen, Nathaniel Josephs, Lizhen Lin, Jie Zhou, and Eric D Kolaczyk. A spectral-based framework for hypothesis testing in populations of networks. *arXiv:2011.12416*, 2020.

Yijia Chen, Marc Thurley, and Mark Weyer. Understanding the complexity of induced subgraph isomorphisms. In *Automata, Languages and Programming: 35th International Colloquium (ICALP)*, pages 587–596. Springer, 2008.

Norishige Chiba and Takao Nishizeki. Arboricity and subgraph listing algorithms. *SIAM Journal on Computing*, 14(1):210–223, 1985.

Vibhor Dodeja, Mohammad Almasri, Rakesh Nagi, Jinjun Xiong, and Wen-mei Hwu. PARSEC: Parallel subgraph enumeration in CUDA. In *International Parallel and Distributed Processing Symposium (IPDPS)*, pages 168–178. IEEE, 2022.

Fabrizio Gabbiani and Steven J Cox. *Mathematics for neuroscientists*. Academic Press, 2017.

Chao Gao and John Lafferty. Testing network structure using relations between small subgraph probabilities. *arXiv:1704.06742*, 2017.

Debarghya Ghoshdastidar, Maurilio Gutzeit, Alexandra Carpentier, and Ulrike Von Luxburg. Two-sample hypothesis testing for inhomogeneous random graphs. *The Annals of Statistics*, 48(4):2208–2229, 2020.

Reid Ginoza and Andrew Mugler. Network motifs come in sets: Correlations in the randomization process. *Physics Review E*, 82(1):011921, 2010.

Boris V Gnedenko and Andrey Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, 1954.

Chuangyi Gui, Long Zheng, Pengcheng Yao, Xiaofei Liao, and Hai Jin. Fast triangle counting on GPU. In *High Performance Extreme Computing Conference (HPEC)*, pages 1–7. IEEE, 2019.

Lee M Gunderson and Gecia Bravo-Hermsdorff. Introducing graph cumulants: What is the variance of your social network? *arXiv:2002.03959*, 2020.

Anders Hald. The early history of the cumulants and the Gram-Charlier series. *International Statistical Review*, 68(2):137–153, 2000.

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574, 1959.

Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. Optimal adaptivity of signed-polygon statistics for network testing. *The Annals of Statistics*, 49(6):3408–3433, 2021.

Mehran Kardar. *Statistical Physics of Particles*. Cambridge University Press, 2007.

Salman Khan, Hossein Rahmani, Syed Afaq Ali Shah, and Mohammed Bennamoun. A guide to convolutional neural networks for computer vision. *Synthesis lectures on computer vision*, 8(1):1–207, 2018.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Franz Lehner. Cumulants in noncommutative probability theory I. Noncommutative exchangeability systems. *Mathematische Zeitschrift*, 248(1):67–100, 2004.

Xin Liu, Haojie Pan, Mutian He, Yangqiu Song, Xin Jiang, and Lifeng Shang. Neural subgraph isomorphism counting. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1959–1969, 2020.

László Lovász. *Large Networks and Graph Limits*. American Mathematical Society, 2012.

László Lovász and Balázs Szegedy. The graph theoretic moment problem. *arXiv:1010.5159*, 2010.

Prasanta Chandra Mahalanobis. On the generalized distance in statistics. In *National Institute of Science of India*, pages 49–55, 1936.

Naoki Masuda, Sadamori Kojaku, and Yukie Sano. Configuration model for correlation matrices preserving the node strength. *Physical Review E*, 98(1):012312, 2018.

PA Maugis, Sofia C Olhede, and Patrick J Wolfe. Fast counting of medium-sized rooted subgraphs. *arXiv:1701.00177*, 2016.

Pierre-André Maugis, Sofia C Olhede, Carey E Priebe, and Patrick J Wolfe. Testing for equivalence of network distribution using subgraph counts. *Journal of Computational and Graphical Statistics*, 29(3):455–465, 2020.

Peter McCullagh. *Tensor Methods in Statistics: Monographs on Statistics and Applied Probability*. Chapman and Hall, 1987.

Soumendu Sundar Mukherjee, Purnamrita Sarkar, and Lizhen Lin. On clustering network-valued data. *arXiv:1606.02401*, 2016.

Mark EJ Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409, 2001.

Peter Orbanz. Subsampling large graphs and invariance in networks. *arXiv:1710.04217*, 2017.

Guang Ouyang, Dipak K Dey, and Panpan Zhang. Clique-based method for social network clustering. *Journal of Classification*, 37(1):1–21, 2019.

Jeffrey Pattillo, Nataly Youssef, and Sergiy Butenko. On clique relaxation models in network analysis. *European Journal of Operational Research*, 226(1):9–18, 2013.

Emma Persson, Miguel Castresana-Aguirre, Davide Buzzao, Dimitri Guala, and Erik L.L. Sonnhammer. FunCoup 5: Functional association networks in all domains of life, supporting directed links and tissue-specificity. *Journal of Molecular Biology*, 433(11):166835, 2021.

Ali Pinar, C Seshadhri, and Vaidyanathan Vishal. Escape: Efficiently counting all 5-vertex subgraphs. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1431–1440, 2017.

Edwin JG Pitman. Notes on non-parametric statistical inference. Technical report, North Carolina State University, Department of Statistics, 1949.

C Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in Statistics*, pages 235–247. Springer, 1992.

Johannes Rauh. The polytope of k-star densities. *The Electronic Journal of Combinatorics*, pages 1–4, 2017.

Benjamin Reiser. Confidence intervals for the Mahalanobis distance. *Communications in Statistics-Simulation and Computation*, 30(1):37–45, 2001.

Pedro Ribeiro, Pedro Paredes, Miguel EP Silva, David Aparicio, and Fernando Silva. A survey on subgraph counting: concepts, algorithms, and applications to network motifs and graphlets. *ACM Computing Surveys (CSUR)*, 54(2):1–36, 2021.

Gian-Carlo Rota and Jianhong Shen. On the combinatorics of cumulants. *Journal of Combinatorial Theory, Series A*, 91:283–304, 2000.

Alan Stuart and Maurice G Kendall. *The Advanced Theory of Statistics*. Charles Griffin and Co., Ltd., London, 1963.

Minh Tang, Avanti Athreya, Daniel L Sussman, Vince Lyzinski, Youngser Park, and Carey E Priebe. A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, 26(2):344–354, 2017.

Mattia Tantardini, Francesca Ieva, Lucia Tajoli, and Carlo Piccardi. Comparing methods for comparing networks. *Scientific Reports*, 9(1):1–19, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Neural Information Processing Systems (NIPS)*, 30, 2017.

Feng Xia, Haoran Wei, Shuo Yu, Da Zhang, and Bo Xu. A survey of measures for network motifs. *IEEE Access*, 7:106576–106587, 2019.

Xiaowei Ye, Rong-Hua Li, Qiangqiang Dai, Hongzhi Chen, and Guoren Wang. Lightning fast and space efficient k-clique counting. In *Proceedings of the ACM Web Conference 2022*, pages 1191–1202, 2022.

Yuan Zhang and Dong Xia. Edgeworth expansions for network moments. *The Annals of Statistics*, 50(2):726–753, 2022.

Kangfei Zhao, Jeffrey Xu Yu, Hao Zhang, Qiyan Li, and Yu Rong. A learned sketch for subgraph counting. In *International Conference on Management of Data*, pages 2142–2155, 2021.