



## Testing for Equivalence of Network Distribution Using Subgraph Counts

P.-A. G. Maugis, S. C. Olhede, C. E. Priebe & P. J. Wolfe

To cite this article: P.-A. G. Maugis, S. C. Olhede, C. E. Priebe & P. J. Wolfe (2020) Testing for Equivalence of Network Distribution Using Subgraph Counts, Journal of Computational and Graphical Statistics, 29:3, 455-465, DOI: [10.1080/10618600.2020.1736085](https://doi.org/10.1080/10618600.2020.1736085)

To link to this article: <https://doi.org/10.1080/10618600.2020.1736085>



© 2020 The Author(s). Published with License by Taylor & Francis Group, LLC



[View supplementary material](#)



Published online: 16 Apr 2020.



[Submit your article to this journal](#)



Article views: 278



[View related articles](#)



[View Crossmark data](#)

## Testing for Equivalence of Network Distribution Using Subgraph Counts

P.-A. G. Maugis<sup>a</sup>, S. C. Olhede<sup>b</sup>, C. E. Priebe<sup>c</sup>, and P. J. Wolfe<sup>d</sup>

<sup>a</sup>Department of Statistical Science, University College London, and Pivitar, London, UK; <sup>b</sup>School of Basic Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; <sup>c</sup>Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD; <sup>d</sup>Department of Statistics, Purdue University, West Lafayette, IN

### ABSTRACT

We consider that a network is an observation, and a collection of observed networks forms a sample. In this setting, we provide methods to test whether all observations in a network sample are drawn from a specified model. We achieve this by deriving the joint asymptotic properties of average subgraph counts as the number of observed networks increases but the number of nodes in each network remains finite. In doing so, we do not require that each observed network contains the same number of nodes, or is drawn from the same distribution. Our results yield joint confidence regions for subgraph counts, and therefore methods for testing whether the observations in a network sample are drawn from: a specified distribution, a specified model, or from the same model as another network sample. We present simulation experiments and an illustrative example on a sample of brain networks where we find that highly creative individuals' brains present significantly more short cycles than found in less creative people. Supplementary materials for this article are available online.

### ARTICLE HISTORY

Received November 2018  
Revised December 2019

### KEYWORDS

Connectomes; Statistical testing; Subgraph count statistics

### 1. Introduction

We show that subgraph counts are flexible and powerful statistics for testing distributional properties of networks, when more than one network is observed. Specifically, we use subgraph counts to test the hypotheses that all networks in a sample are generated either from a specified distribution, from distributions in a specified model, or from the same model as that of another sample.


Our results address the fundamental inference problem raised by the following experiment (Gray et al. 2012): Networks connecting brain regions of individuals of varying levels of creativity are observed. However, while these observations can be assumed to be independent, due to the variability of the brain structure and the instability of the observation technique they cannot be assumed to be identically distributed; for instance, they need not contain the same number of nodes and edges. This implies that if we identify each network realization with its adjacency matrix, these matrices will be of different sizes. This makes more difficult, for instance, estimating the distribution the adjacency matrices are drawn from compared to the case of a sample of independent and identically distributed realizations on a fixed number of nodes. How, while allowing for such variations, can we test for significant differences between individuals with different levels of creativity?

Formally, we consider that each network is an observation—say  $G_i$  for each subject—and a collection of observed networks form a sample—say  $\mathcal{G} = (G_1, \dots, G_N)$ . Our goal is to infer distributional properties of the  $G_i$ 's as  $N$  grows. This parallels more classical statistical settings, where an observation is a vector—such as  $X_i \in \mathbb{R}^k$ —and a sample is a matrix:  $\mathcal{X} =$

$(X_1, \dots, X_N) \in \mathbb{R}^{k \times N}$ . However, this parallel with the classical setting stops there. Indeed, the  $G_i$  need not be of the same size, or have nodes that are matched with a 1-1 correspondence between graphs; that is, the  $X_i$ 's would not have the same dimension, and the entries could be shuffled.

This setting strongly differs from the two settings that have already seen extensive research. First, there is the setting that focuses on the asymptotic properties of large random networks (see Bickel, Chen, and Levina 2012; Ho, Parikh, and Xing 2012; Hoff, Raftery, and Handcock 2012; Sussman et al. 2012; Tang, Sussman, and Priebe 2013; Olhede and Wolfe 2014; Bhat-tacharyya and Bickel 2015; Fosdick and Hoff 2015; Coulson, Gaunt, and Reinert 2016; Klopp, Tsybakov, and Verzelen 2016, to cite but a few). Available statistical tests for network comparison in that setting focus on the case where one or two large networks are observed (Asta and Shalizi 2014; Banerjee and Ma 2017; Gao and Lafferty 2017; Ghoshdastidar et al. 2017; Tang, Athreya, Sussman, Lyzinski, Priebe, et al. 2017), or where one finite network is compared to a fixed model alone (Birmele 2012; Ali et al. 2014, 2016). The second setting addresses graph samples, as we do here, but under additional assumptions: samples that are independent and identically distributed, where all graphs have the same size, and where nodes across networks can be matched one-to-one. Then, under these assumptions, it is possible to compare network summaries using classical statistical methods (Simpson, Moussa, and Laurienti 2012; Darianu et al. 2013; Stoffers et al. 2013; Ginestet et al. 2017), or by fitting a parametric or semiparametric model (Simpson, Moussa, and Laurienti 2012; Durante and Dunson 2018; Wang, Zhang, and Dunson 2019).

**CONTACT** P.-A. G. Maugis  [p.a.maugis@gmail.com](mailto:p.a.maugis@gmail.com)  Privitar, 15 Hatfields SE1 8DJ, London, UK.

 Supplementary materials for this article are available online. Please go to [www.tandfonline.com/r/JCGS](http://www.tandfonline.com/r/JCGS).

© 2020 The Author(s). Published with License by Taylor & Francis Group, LLC

This is an Open Access Article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We provide, in the graph sample setting described above, an analog of the multivariate  $t$ -test for network samples: methods to test whether a given network sample  $\mathcal{G}$  presents averages consistent with either a specified model, or with that of another sample. The averages we use are *subgraph counts*; for example, the number of  $\blacktriangle$  or  $\blacksquare$  in the sample. The choice of subgraph counts as statistics is motivated by their success in comparing large networks (Milo et al. 2002; Ali et al. 2014), but also by results in random graph theory and the study of large graphs. In both fields, subgraph counts have proved to be among the most powerful tools available to compare networks (Ali et al. 2014, 2016) and are known to have properties similar to moments of random variables when studying large networks (Diaconis and Janson 2008; Bickel, Chen, and Levina 2012; Lovász 2012; Chatterjee and Diaconis 2013). Finally, because of these results, many powerful algorithms exist to count subgraphs efficiently (e.g., Ortmann and Brandes 2016; Talukder and Zaki 2016).

Formally, we are representing network samples in a space defined by subgraph counts, and performing comparisons in that space. While other network comparison techniques also use embeddings (Asta and Shalizi 2014; Gao and Lafferty 2017; Tang, Athreya, Sussman, Lyzinski, Priebe, et al. 2017; Wang, Zhang, and Dunson 2019), using subgraph counts presents three key advantages: first, if the  $G_i$ 's are generated by a block-model (Hoff, Raftery, and Handcock 2012)—one of the most popular random network models to date—and for some families of subgraphs, the embedding is one-to-one. This result is known as the finite forcibility of a family of graphs (Lovász 2012, chap. 16). Second, very few assumptions on each  $G_i$  need to be made as  $N$  grows to obtain consistency and asymptotic normality of the image of  $\mathcal{G}$  in the embedding space. This enables us to work under a very flexible null model. Finally, because it relies on direct summaries of the  $G_i$ 's—the number of  $\blacktriangle$ ,  $\blacksquare$ , and so on—this embedding remains interpretable, part of the appeal of using subgraph counts for inference.

The main risk in using subgraph counts for testing purposes is that we cannot be sure that the subgraphs considered are sufficient to distinguish the null and the data generating mechanism. We provide several experiments supporting our claim that, while possible, this risk does not appear to be common. Furthermore, this risk is balanced by the interpretability of subgraph counts: if the null and the generating mechanism have the same number of subgraphs, perhaps they should be considered equivalent for the purpose of the study at hand. Another shortcoming is that subgraph counts can be highly correlated, especially in denser networks, making the estimation of the inverse of their covariance matrix unstable. One of our contributions is closed form formulas for these covariance matrices under our null, which mitigates but does not fully resolve this issue. Finally, because subgraph counting libraries are not standard, implementing the proposed methods is not as easy as for other methods, which tend to rely on more common, linear algebra related, data transformation pipelines. To this end we made available all the code to perform our analysis.<sup>1</sup>

In the remainder of this article, we first introduce subgraph counts and the kernel based random graph model. We then successively present the case where all the networks in the sample come from the same kernel model, and the case where each observed network may come from different kernels. In both cases, we prove asymptotic normality of our estimator, present a plug-in estimator of its variance, describe the limit of the estimator under the alternative hypothesis, and provide representative examples showing the practical utility of the approach. Methods to produce the figures, as well as supporting simulations, can be found in the supplementary materials. We conclude with an analysis of connectome data, and with a discussion.

## 2. Subgraph Counts in Kernel Based Random Graphs

We now define our statistics (subgraph counts) and our null model (the kernel based random graph model). Subgraph counts are natural statistics for comparing networks for two reasons. First, subgraph counts intuitively summarize a network through its fundamental building blocks. This has historically given them purchase in addressing hard fundamental and empirical problems (Rucinski 1988; Milo et al. 2002; Ali et al. 2014). Second, subgraph counts possess very tractable analytical properties. We will describe and leverage these properties below, in a manner paralleling what is done in related literature (Rucinski 1988; Bickel, Chen, and Levina 2012; Bhattacharyya and Bickel 2015).

A *subgraph count* is the number of copies of a given graph in another graph (see Figure 1). Throughout, we call the subgraph—denoted  $F$ —the graph which is counted and  $G$  the larger graph in which the counting takes place. All graphs will be simple (unweighted, no self loops, or multiple edges). Subgraphs are also termed motifs, pattern graphs, or shapes depending on the field (Alon, Yuster, and Zwick 1997; Milo et al. 2002; Hočevar and Demšar 2014; Benson, Gleich, and Leskovec 2016).

For clarity, we define subgraph counts formally as follows:

**Definition 1 (Graph equivalence “ $\equiv$ ”).** Fix two graphs  $F$  and  $F'$ . We say that  $F$  is equivalent to—or is a copy of— $F'$ , and write  $F \equiv F'$ , if there exists a bijective map  $\phi$  from the vertex set of  $F$  to the vertex set of  $F'$  such that  $ij$  is an edge in  $F$  iff  $\phi(i)\phi(j)$  is an edge in  $F'$ .

**Definition 2 (Subgraph count  $X_F(G)$ ).** Fix two graphs  $F$  and  $G$ . We denote by  $X_F(G)$  the number of subgraphs (not necessarily induced) of  $G$  equivalent to  $F$ ; that is,

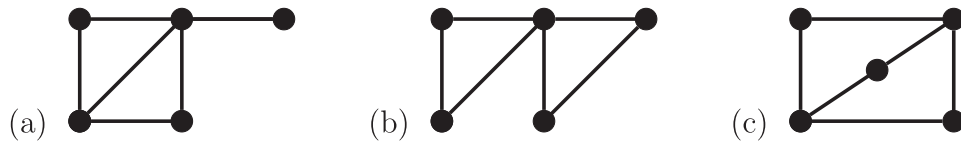
$$X_F(G) = \# \{F' \subset G : F' \equiv F\},$$

where  $F \subset G$  if the vertex and edge sets of  $F$  are subsets of those of  $G$ .

With this notation, calling  $G_a$ ,  $G_b$ , and  $G_c$  the graphs in Figure 1, we have  $X_{\blacksquare}(G_a) = 1$ ,  $X_{\blacksquare}(G_b) = 0$ ,  $X_{\blacksquare}(G_c) = 3$ . A more complete definition is provided in Definition 2.

The power of subgraph counts in the study of networks stems from their inherent linearity. Indeed, products of subgraph counts are but linear combinations of other subgraph counts. Intuitively, first observe that a product of two subgraph counts will involve counting pairs of copies. Thus, a product of

<sup>1</sup>Code to perform all methods, as well as reproduce all figures and tables reported in the article, can be found at: <https://github.com/PierreAndreMaugis/motifs-for-network-samples>.



**Figure 1.** Example of subgraph counts. There are 6 copies of the edge (—) in all three graphs. There are 2 copies of the triangle (▲) in (a) and (b), but 0 in (c). There are 1, 0, and 3 copies of the square (■) in (a), (b), and (c), respectively.

subgraph counts can be recovered by counting the number of copies of all subgraphs that can be induced by a pair of copies. More precisely, in the Appendix we show the following, which is implicitly used in Rucinski (1988):

**Lemma 1 (Linearity of subgraph counts (Rucinski 1988)).** For any two graphs  $F$  and  $F'$ , there are factors  $c_H$  and a set  $\mathcal{H}_{FF'}$  of subgraphs—the set of subgraphs that can be obtained using one copy each of  $F$  and  $F'$  as building blocks—such that for any graph  $G$

$$X_{F_1}(G)X_{F_2}(G) = \sum_{H \in \mathcal{H}_{FF'}} c_H X_H(G).$$

For instance, as these will be used later on, we have that in any graph  $G$ ,

$$\begin{aligned} X_{\triangle}(G)^2 &= 2X_{\triangle\triangle}(G) + 2X_{\square}(G) + 2X_{\text{pentagon}}(G) + X_{\text{hexagon}}(G), \\ X_{\square}(G)^2 &= 2X_{\square\square}(G) + 2X_{\text{pentagon}}(G) + 6X_{\text{hexagon}}(G) + 2X_{\text{heptagon}}(G) \\ &\quad + 2X_{\text{octagon}}(G) + 2X_{\text{nonagon}}(G) + 6X_{\text{decagon}}(G) + 6X_{\text{undecagon}}(G) \\ &\quad + X_{\text{dodecagon}}(G). \end{aligned}$$

This algebraic property of subgraph counts allows us to understand the proofs of Rucinski (1988) and Bhattacharyya and Bickel (2015). The property is also crucial to the subgraph counting algorithms of Hočevar and Demšar (2014), and can be found in many other examples. Crucially, as opposed to cases where it is the model that enforces linearity—such as with assumptions of Normality—it is the nature of the statistics (subgraph counts) and the system (graphs) that makes our problem linear.

The linearity of subgraph counts allows us to use as null the very flexible kernel based model (Lovász 2012). This framework subsumes most models used in the statistical literature on networks; for example, blockmodel (Hoff, Raftery, and Handcock 2012) and dot-product models (Sussman et al. 2012). It has the intuitive structure of affixing to each node  $i$  a latent feature (here  $x_i$ ) and of connecting nodes  $i$  and  $j$  (conditionally independently) with a probability determined by the node features (here  $f(x_i, x_j)$ ).

**Definition 3 (Kernel  $f$  and random graph  $G_n(f)$ ).** Fix a symmetric measurable map  $f : [0, 1]^2 \rightarrow [0, 1]$ , and call it a *kernel*. We call  $G_n(f)$  the random graph distribution over graphs with  $n$  nodes such that: to each node is randomly and independently assigned a feature  $x_i \in [0, 1]$ , with  $x_i \sim \text{Unif}([0, 1])$ ; and where edges form independently conditionally on  $\{x_i\}_{i \in [n]}$  with conditional probability

$$\mathbb{P}[ij \in G | x_i, x_j] = f(x_i, x_j).$$

To recover a blockmodel with  $K$  blocks, it suffices to consider a partition of  $[0, 1]$  in  $K$  sets (i.e.,  $(P_1, \dots, P_K) \in \mathcal{P}_K([0, 1])$ ) and set  $f$  as constant over each  $P_u \times P_v$ . The dot-product model is recovered with a kernel  $f$  of finite rank; that is,  $f(x, y) = \sum_{u \leq K} \lambda_u f_u(x) f_u(y)$ .

The model assumes that vertices' locations in the latent space are independent and identically distributed, so that the model does not assume any structure or symmetry among the nodes in the observed networks. However, it can accommodate any such structure, which would be recovered through estimation. Nonetheless, in some cases it may seem relevant to enforce such structure; for example, in our brain example, assume that nodes in different hemispheres are unlikely to connect. As always in statistics, making such assumptions risks misspecifying the model to gain faster convergence, which might be necessary in some cases. We did not find it necessary to make such assumptions in our application.

In the kernel framework, subgraph counts have direct interpretation as moments of  $f$  (Lovász 2012) (see Equation (1)). Specifically, if  $G \sim G_n(f)$ , then the moments of  $X_F(G)$  are moments of  $f$ . An infinite number of subgraph counts are sufficient statistics to distinguish between any two kernels (Bollobás and Riordan 2009; Lovász 2012), as the subgraph counts can be used to define the subgraph metric. However, there are no guarantees on which or how many subgraph counts are needed to distinguish between two kernels. For block models and finite rank models, we know only that a finite number is sufficient [a concept known as finite forcibility, see Lovász (2012, chap. 16.7 and Appendix 4) for more details]. The cost in using subgraph counts is that we compare the kernels, and therefore lose information pertaining to the latent locations (the  $x_i$ 's). For example, consider comparing brain samples where in one sample all connectomes have a fully connected left hemisphere, a fully disconnected right hemisphere and no connections across the two hemispheres, whereas in the other sample all connectomes have fully disconnected left hemisphere, a fully connected right hemisphere and no connections across the two hemispheres: in this setting there is a clear difference in the brains between the two samples, but the subgraph counts are the same.

This makes subgraph counts especially appropriate as statistics in an hypothesis test. Indeed, for any finite set of subgraphs  $\mathcal{F}$ , we could have two kernels  $f$  and  $f'$  such that  $f \neq f'$ , and yet  $\forall F \in \mathcal{F}, \mathbb{E}_{G \sim G_n(f)} X_F(G) = \mathbb{E}_{G \sim G_n(f')} X_F(G)$ . However, as we prove below, if  $\mathbb{E}_{G \sim G_n(f)} X_F(G) \neq \mathbb{E}_{G \sim G_n(f')} X_F(G)$  for any  $F$ , then  $f \neq f'$ . Therefore, a difference in subgraph counts is sufficient but not necessary to distinguish between kernels. Conversely, not observing a difference in subgraph counts can only imply that we do not observe a difference in the kernels, and therefore that we fail to reject the hypotheses that the kernels are equal. This has implications regarding the power of our test that we explore below.

Notably, all known results on subgraph counts under the kernel model consider the setting where one very large graph is observed. Here, we present the tools to address the problem where a sample of graphs is observed.

### 3. The Simple Case: Samples From One Kernel

We now present a central limit theorem as well as practical methods to build confidence regions for the subgraph counts observed in a network sample  $\mathcal{G} = (G_1, \dots, G_N)$ . In this section, we assume that there is a kernel  $f$  such that each  $G_i$  is drawn independently from  $G_{n_i}(f)$  (with  $n_i = |G_i|$ ), where for any graph  $F$  we write  $|F|$  for the number of nodes in  $F$ .

Fix  $F \in \mathcal{F}$  and  $G \in \mathcal{G}$ . In this setting,  $X_F(G)$  is a random variable, and the first parameter to consider is its mean. To compute this mean, let  $F_1, \dots, F_{X_F(K_G)}$  be all the copies of  $F$  in  $K_G$  (the complete graph over the nodes of  $G$ ), so that using the linearity of the expectation, we have that

$$\mathbb{E}X_F(G) = \mathbb{E} \sum_{j \in [X_F(K_G)]} \mathbb{1}_{\{F_j \subset G\}} = \sum_{j \in [X_F(K_G)]} \mathbb{E}\mathbb{1}_{\{F_j \subset G\}}.$$

Then, direct computations show that  $\mathbb{E}\mathbb{1}_{\{F_j \subset G\}}$  does not depend on  $j$  (see Proposition 1), and that

$$\mathbb{E}\mathbb{1}_{\{F_j \subset G\}} = \mu_F(f) := \int_{[0,1]^{|F|}} \prod_{uv \in F} f(x_u, x_v) \prod_{u \in F} dx_u, \quad (1)$$

so that  $\mathbb{E}X_F(G) = X_F(K_G)\mu_F(f)$ . Observe that  $\mu_F(f)$  is a moment of the kernel  $f$ , as discussed above.

Similar computations for higher moments, aided by Lemma 1, enable us to use the Lindeberg–Feller central limit theorem and the Cramer–Wold device to prove the following:

**Theorem 1 (Statistical properties of subgraph counts).** Fix a set of graphs  $\mathcal{F}$ , a kernel  $f$  and a sequence  $n = (n_i)_{i \in \mathbb{N}}$  such that  $2 \max_{F \in \mathcal{F}} |F| \leq \min_{i \in \mathbb{N}} n_i$ . Let  $\mathcal{G} = (G_i)_{i \in [N]}$  be a network sample such that for all  $i$ ,  $G_i \sim G_{n_i}(f)$ . Set  $\hat{\mu}_F(\mathcal{G}) = N^{-1} \sum_{G \in \mathcal{G}} X_F(G)/X_F(K_G)$ ,  $\hat{\mu}_{\mathcal{F}}(\mathcal{G}) = (\hat{\mu}_F(\mathcal{G}))_{F \in \mathcal{F}}$ , and  $\mu_{\mathcal{F}}(f) = (\mu_F(f))_{F \in \mathcal{F}}$ . Then,  $\hat{\mu}_{\mathcal{F}}(\mathcal{G})$  is an unbiased,  $\sqrt{N}$ -consistent and asymptotically normal estimator of  $\mu_{\mathcal{F}}(f)$ ; that is,  $\mathbb{E}\hat{\mu}_{\mathcal{F}}(\mathcal{G}) = \mu_{\mathcal{F}}(f)$  and there exists a positive semidefinite  $\Sigma_{\mathcal{F}}(n, f)$  such that asymptotically in  $N$ :

$$\sqrt{N}(\hat{\mu}_{\mathcal{F}}(\mathcal{G}) - \mu_{\mathcal{F}}(f)) \rightarrow \text{Normal}(0, \Sigma_{\mathcal{F}}(n, f)).$$

Furthermore, for each  $F, F' \in \mathcal{F}$ ,

$$\text{cov}(\hat{\mu}_F(\mathcal{G}), \hat{\mu}_{F'}(\mathcal{G})) = \sum_{H \in \mathcal{H}_{F \sqcup F'} \setminus \{F \sqcup F'\}} \omega_H(n; N) (\mu_H(f) - \mu_F(f)\mu_{F'}(f)), \quad (2)$$

with  $F \sqcup F'$  the disjoint union of  $F, F'$ ,  $\omega_H(n; N) = \frac{1}{N} \sum_{i=1}^N \frac{c_H X_H(K_{G_i})}{X_F(K_{G_i})X_{F'}(K_{G_i})}$ , and  $c_H$  is defined in Lemma 1.

Crucial to the following is the covariance matrix  $\Sigma_{\mathcal{F}}(n, f)$ —obtainable by taking the limit in  $N$  in (2) for each  $F, F' \in \mathcal{F}$ —which will enable the computation of confidence regions. Interestingly, its elicitation is more involved than for the study of large networks, where only a few terms dominate. We refer to the

Appendix for the proof. The Appendix’s proof relies on understanding the moments of single network counts, which have been studied extensively in the Erdős–Rényi and the exchangeable cases (see, e.g., Rucinski 1988; Coulson, Gaunt, and Reinert 2016).

**Remark 1 (Nonrandom latent positions).** Because our proof builds on the foundation of Rucinski (1988), our results extend to the case where the latent  $x_i$ ’s are not random; say fixed to some values. The only modification would be that the sum in (2) should be restricted to graphs  $H$  formed by copies of  $F$  and  $F'$  overlapping over at least an edge (as opposed to overlapping over one node in (2)). This also means that all our methods apply to that case, with the caveat that the resulting test will be conservative (as the variance is overestimated).

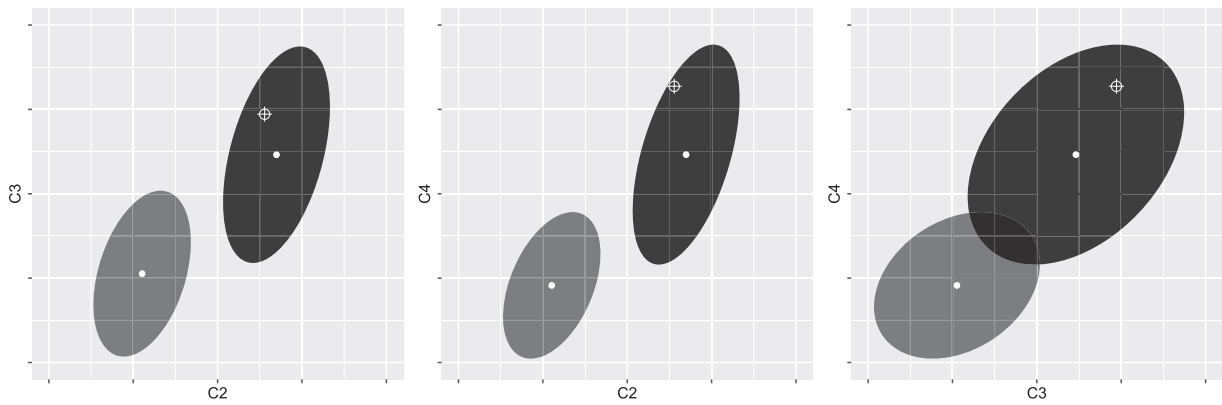
Theorem 1 enables testing against the null that all  $G_i$  are drawn from a given kernel. Further, the SI contains a simulation experiment exploring convergence in small samples. To make this concrete, we consider an example in Figure 2. There, we observe a graph sample  $\mathcal{G} = (G_1, \dots, G_{100})$ , and aim to compare it to two kernels  $f_a$  (in black) and  $f_b$  (in gray) using Theorem 1; that is, we assume that for  $i \in [N]$ ,  $G_i \sim G_{n_i}(f)$  and consider the null hypothesis  $H_0 : f = f_a$  and the alternative  $H_1 : f = f_b$ . We present as a white cross  $\hat{\mu}_{\mathcal{F}}(\mathcal{G})$ . The sizes of the networks in  $\mathcal{G}$ , the  $n_i$ , are nonrandom but not constant. We achieve this by using the sequence of digits of  $\pi$ .

First, since we have specified  $f_a$  and  $f_b$ , we can evaluate both  $\mu_{\mathcal{F}}(f_a)$  and  $\mu_{\mathcal{F}}(f_b)$  and include them in the figure (as white dots). Then, since  $n = (n_i)_{i \in \mathbb{N}}$  is observed, we can compute  $\Sigma_{\mathcal{F}}(n, f_a)$  and  $\Sigma_{\mathcal{F}}(n, f_b)$  using Theorem 1, which allows us to compute the confidence ellipse around  $\mu_{\mathcal{F}}(f_a)$  and  $\mu_{\mathcal{F}}(f_b)$  (in shaded black and gray, respectively). Finally, since we know the limit distribution and covariance under the null, we can use  $\Sigma_{\mathcal{F}}(n, f_a)$  and  $\mu_{\mathcal{F}}(f_a)$  to compute a  $p$ -value using Mahalanobis distance; for example, assuming  $\Sigma_{\mathcal{F}}(f_a)$  is full rank, the  $p$ -value is  $1 - F_{\chi^2_{|\mathcal{F}|}}((\hat{\mu}_{\mathcal{F}}(\mathcal{G}) - \mu_{\mathcal{F}}(f_a))' \Sigma_{\mathcal{F}}(f_a)^{-1} (\hat{\mu}_{\mathcal{F}}(\mathcal{G}) - \mu_{\mathcal{F}}(f_a)))$ , with  $F_{\chi^2_{|\mathcal{F}|}}$  the cumulative distribution function of the  $\chi^2$  distribution with  $|\mathcal{F}|$  degrees of freedom.

Figure 2 is useful to understand the power of the proposed test. We see that since not all the confidence ellipses overlap, the power is larger than 0.95. However, if we were to use only  $\mathfrak{H}$ , then the power would be less as these ellipses do overlap. For larger sample sizes, the radii of both ellipses will be smaller, so that eventually the power tends to 1 using any combination of subgraphs. In the supplementary materials, we explore the power of our test when the expected  $\bullet \rightarrow$  count is held fixed, but the number of blocks in the null and true models differ.

#### 3.1. Finding the Odd Connectome Out

Before proceeding to a more general sampling design, and therefore a less constrained null, we consider the following experiment. In this experiment, we study a diffusion MRI dataset consisting of 57 connectomes with between 750 and 1191 vertices each. Indeed, because each brain has a slightly different shape, the number of vertices that is captured will change for each subject. Graphs were estimated using the NeuroData’s MR



**Figure 2.** Testing for a kernel using  $\mathcal{F} = \{C_2, C_3, C_4\} = \{\bullet\rightarrow, \blacktriangle, \blacksquare\}$ . The sample  $\mathcal{G}$  is such that:  $N = 100$ ,  $n_i$  is the  $i$ th digit of  $\pi$  plus 30,  $G_i$  is drawn from a kernel  $f$  ( $G_i \sim G_{n_i}(f)$ ). The estimate  $\hat{\mu}_{\mathcal{F}}(\mathcal{G})$  is denoted by a white cross. Overlaid are the expected densities (white dots) and the confidence ellipse (shaded area) for two alternative kernels  $f_a$  and  $f_b$ .

Graphs pipeline (Kiar et al. 2018), with vertices representing subregions defined via spatial proximity (we use parcellation technique DS01216) and edges defined by tensor-based fiber streamlines connecting these regions.

During the data collection of one of the connectomes, a numerical error was reported by the system, however, the process completed properly. While one could conservatively discard the observation, securing a subject and performing the scan is time consuming and expensive, so that it would be preferable to keep the connectome unless it is proven to be too different to be processed.

Testing whether this odd connectome is too different is something that can be done from Theorem 1 as follows: (i) count in each of the connectomes the density of  $\bullet\rightarrow$ ,  $\blacktriangle$ , and  $\blacksquare$  and compute  $\Sigma$ , the covariance matrix of these densities; (ii) split the sample in two at random; (iii) compute the mean  $\mu_i$  of the counts in both samples, with  $i \in \{1, 2\}$ ; (iv) then, under the null that both samples are generated independently from the same kernel,  $q = (57/2)(\mu_2 - \mu_1)(2\Sigma)^{-1}(\mu_2 - \mu_1)^\top$  should be a draw from the  $\chi^2_3$  distribution.

Performing this exercise on our sample multiple times (1000 times) we find that we can reject the null of the odd connectome being from the same kernel as the others. Indeed, the  $p$ -value of observing a value at least as large as the observed  $q$  is 0.03 (after using Bonferroni correction). If we exclude the odd connectome, we fail to reject the null with the same experimental set-up. Closer examination shows that while the density of  $\bullet\rightarrow$  in the odd connectome is similar to the other connectomes, the densities of  $\blacktriangle$  and  $\blacksquare$  are larger in the odd connectome compared to the others.

**Remark 2 (Uniform convergence and bootstrap).** In Theorem 1, we obtain uniform convergence (since in the proof we verify the Lyapunov condition, which implies that the Berry-Esseen inequality applies). This means that resampling bootstrap can be applied, using standard arguments. In the application above we could have leveraged a bootstrap method, and for instance tested for the distribution of the bootstrapped  $p$ -values being significantly different from the uniform distribution. This was unnecessary here, but it can lead to a more powerful test.

#### 4. The General Case: Flexible Sampling Design

Here, we expand our results to cases where the observed networks may be generated from different kernels. Indeed, in many settings, the sampling mechanism may distort the structure of the underlying kernel; for example, although the network connecting brain regions can be satisfactorily modeled by a blockmodel (Koutra, Vogelstein, and Faloutsos 2013), the proportion of nodes of each block may be different in different experimental settings, so that each observation is drawn from a different blockmodel.

In this practically important and conceptually challenging new setting, the proof techniques developed for Theorem 1 yield the following.

**Theorem 2.** Fix a set of graphs  $\mathcal{F}$ , a sequence of kernels  $f = (f_i)_{i \in \mathbb{N}}$  and a sequence of integers  $n = (n_i)_{i \in \mathbb{N}}$  such that  $2 \max_{F \in \mathcal{F}} |F| \leq \min_{i \in \mathbb{N}} n_i$ . Let  $\mathcal{G} = (G_1, \dots, G_N)$  be a network sample such that for all  $i$ ,  $G_i \sim G_{n_i}(f_i)$ . Set  $\hat{\mu}_{\mathcal{F}}(\mathcal{G}) = N^{-1} \sum_{G \in \mathcal{G}} (X_F(G)/X_F(K_G))_{F \in \mathcal{F}}$  and  $\mu_{\mathcal{F}}(f; N) = N^{-1} \sum_{i=1}^N \mu_{\mathcal{F}}(f_i)$ . Then, asymptotically in  $N$ , and for some  $\Sigma_{\mathcal{F}}^*(n, f)$ , we have

$$\sqrt{N}(\hat{\mu}_{\mathcal{F}}(\mathcal{G}) - \mu_{\mathcal{F}}(f; N)) \rightarrow \text{Normal}(0, \Sigma_{\mathcal{F}}^*(n, f)).$$

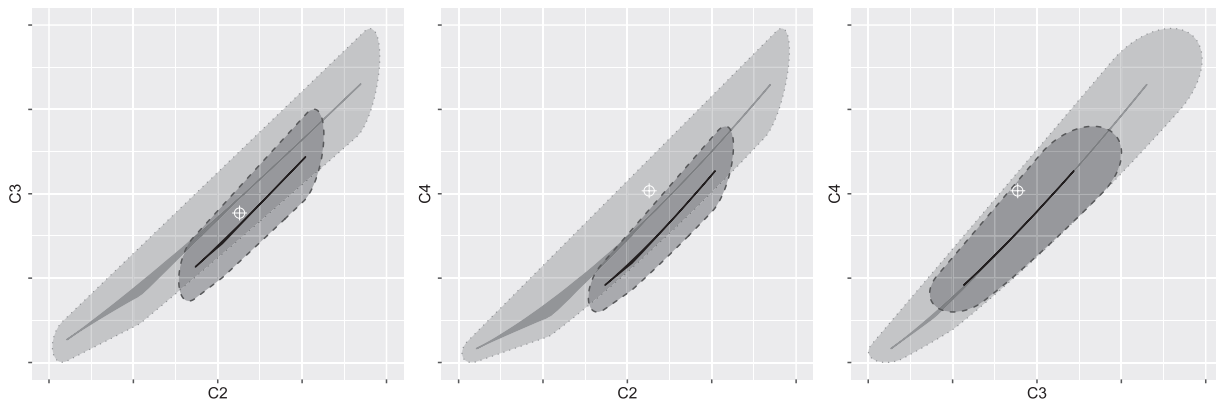
Therefore, even in this much more flexible setting, we can recover the barycenter of the  $\mu_{\mathcal{F}}(f_i)$ . However, the variance has now a more complex structure (see Appendix for details).

Following the intuition of our example of brain networks, and to make the usefulness of Theorem 2 concrete, we introduce the flexible stochastic blockmodel (FSBM).

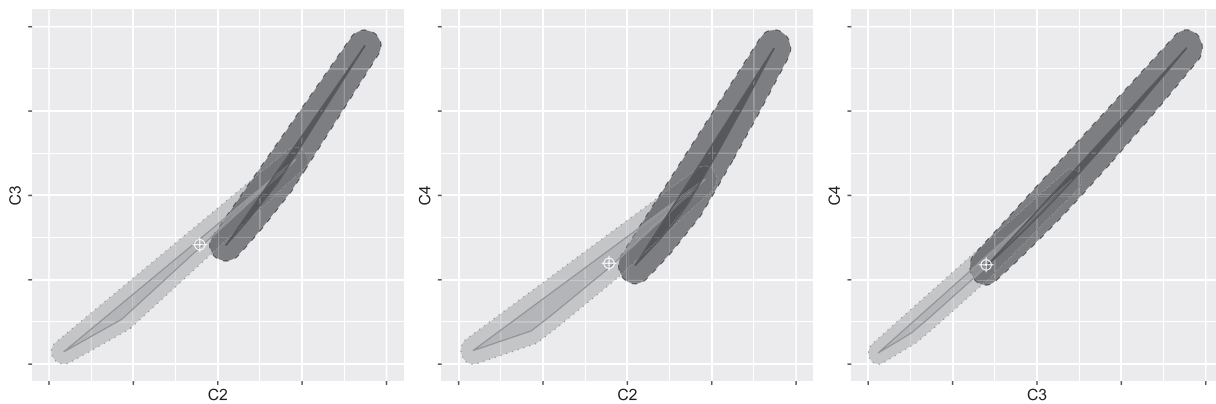
**Definition 4 (FSBM and embedding shape).** For a symmetric matrix  $\mathbf{B} \in [0, 1]^{K \times K}$  we call  $\mathcal{D}(\mathbf{B})$  the set of all possible kernels with the same block structure as  $\mathbf{B}$ ; that is, recalling that  $\mathcal{P}_K([0, 1])$  is the set of partitions of  $[0, 1]$  in  $K$  sets,

$$\mathcal{D}(\mathbf{B}) = \left\{ f : \exists (P_1, \dots, P_K) \in \mathcal{P}_K([0, 1]) \text{ s.t. } \forall x \in P_s, y \in P_t, \right. \\ \left. f(x, y) = \mathbf{B}_{st} \right\}.$$

For a set  $\mathcal{F}$  of graphs, we call the set  $\mu_{\mathcal{F}}(\mathbf{B}) = \{\mu_{\mathcal{F}}(f) \text{ for } f \in \mathcal{D}(\mathbf{B})\}$  the embedding shape.



**Figure 3.** Testing for a FSBM class. The sample  $\mathcal{G}$  is such that:  $N = 200$ ,  $n_i$  is the  $i$ th digit of  $\pi$  plus 30,  $G_i$  is drawn from a kernel  $f$  ( $G_i \sim G_{n_i}(f)$ ). With  $\mathcal{F} = \{C_2, C_3, C_4\}$ , we estimate  $\hat{\mu}_{\mathcal{F}}(\mathcal{G})$ , and plot it as a white cross. Then, we draw in solid color the embedding shapes  $\mu_{\mathcal{F}}(\mathbf{B}_a)$  and  $\mu_{\mathcal{F}}(\mathbf{B}_b)$ . In shaded color we draw the associated confidence regions;  $p$ -values can be obtained using the Mahalanobis distance associated with the closest point to  $\hat{\mu}_{\mathcal{F}}(\mathcal{G})$  in  $\mu_{\mathcal{F}}(\mathbf{B}_a)$  and  $\mu_{\mathcal{F}}(\mathbf{B}_b)$ .



**Figure 4.** Testing for a full FSBM class. The sample  $\mathcal{G}$  is such that:  $N = 10^3$ ,  $n_i$  is the  $i$ th digit of  $\pi$  plus 50,  $G_i$  is drawn from a kernel  $f_i$  ( $G_i \sim G_{n_i}(f_i)$ ). With  $\mathcal{F} = \{C_2, C_3, C_4\}$ , we estimate  $\hat{\mu}_{\mathcal{F}}(\mathcal{G})$ , and plot it as a white cross. Then, we draw in solid color the convex hulls of the embedding shapes  $\mu_{\mathcal{F}}(\mathbf{B}_a)$  and  $\mu_{\mathcal{F}}(\mathbf{B}_b)$ . In shaded color we draw the associated confidence regions; approximate (and conservative)  $p$ -values can be obtained by determining the confidence level at which the observation ceases to be in the confidence region.

For instance, with  $\mathbf{B} \in [0, 1]^{2 \times 2}$  and  $\mathcal{F} = \{\bullet \rightarrow, \blacktriangle\}$ , then:

$$\mu_{\mathcal{F}}(\mathbf{B}) = \left\{ \left( \pi^2 B_{11} + 2\pi(1 - \pi)B_{12} + (1 - \pi)^2 B_{22}, \right. \right. \\ \left. \left. \pi^3 B_{11}^3 + 3\pi^2(1 - \pi)B_{11}B_{12}^2 + 3\pi(1 - \pi)^2 B_{22}B_{12}^2 \right. \right. \\ \left. \left. + (1 - \pi)^3 B_{22}^3 \right) : \pi \in [0, 1] \right\}.$$

The most direct way of using the FSBM is to test for all the  $f_i$  being equal to any blockmodel instance in a class; that is, assume that all  $G_i$ 's are drawn from a kernel  $f$  and test for the null  $H_0 : f \in \mathcal{D}(\mathbf{B})$ . This is achieved by using a composite hypothesis test, and our results allow us to produce confidence regions and  $p$ -values using the same tools as before.

We present such an example in Figure 3. There, we observe  $\mathcal{G} = (G_1, \dots, G_{200})$ , and consider two FSBM classes generated from  $\mathbf{B}_a$  (in gray) and  $\mathbf{B}_b$  (in black). Then, we assume that all networks in the sample are drawn from a kernel  $f$  and test for the null  $H_0 : f \in \mathcal{D}(\mathbf{B}_a)$  and the alternative  $H_1 : f \in \mathcal{D}(\mathbf{B}_b)$ . We first represent  $\hat{\mu}_{\mathcal{F}}(\mathcal{G})$  as a white cross. Using Definition 4, we plot the embedding shapes  $\mu_{\mathcal{F}}(\mathbf{B}_a)$  and  $\mu_{\mathcal{F}}(\mathbf{B}_b)$  in solid gray and black, respectively. The confidence regions (in shaded gray with dotted contour and black with dashed contour) are the union of the confidence ellipses at all points in  $\mu_{\mathcal{F}}(\mathbf{B}_a)$  and  $\mu_{\mathcal{F}}(\mathbf{B}_b)$ .

A more general use of Theorem 2 is to test for all graphs in a sample being drawn from elements of a FSBM class; that

is, assume that the  $G_i$ 's are drawn from the  $f_i$ 's and test for the null  $H_0 : \forall i \in [N], f_i \in \mathcal{D}(\mathbf{B})$  for some  $\mathbf{B}$ . As before, we face a composite null, and we may compute the confidence region and the  $p$ -value by scanning all possible sequences  $f$ . This, however, is clearly computationally intractable. Nonetheless, the form of the variance and the structure of the FSBM allows us to propose conservative confidence regions and  $p$ -values that can be efficiently computed (we fully describe the method in the supplementary materials).

We present such an example in Figure 4. There we observe  $\mathcal{G} = (G_1, \dots, G_{10^3})$  and consider two FSBM classes generated by  $\mathbf{B}_a$  and  $\mathbf{B}_b$ . We first plot  $\hat{\mu}_{\mathcal{F}}(\mathcal{G})$  as a white cross. Then, using Definition 4, we plot the convex hull of the embedding shapes  $\mu_{\mathcal{F}}(\mathbf{B}_a)$  and  $\mu_{\mathcal{F}}(\mathbf{B}_b)$  (in solid gray and black, respectively) wherein—by Theorem 2— $\mu_{\mathcal{F}}(f; N)$  must lie. Finally, we use a method described in the SI to produce the confidence region around each shape (in shaded hue).

### 5. Are Creative Brains Different From Less Creative Ones?

We now consider a sample of brain networks  $\mathcal{G} = (G_1, \dots, G_{113})$  (Koutra, Vogelstein, and Faloutsos 2013). This sample was produced as follows: diffusion MRI (dMRI) and structural MRI (sMRI) scans from 113 individuals were collected

over 2 sessions from Beijing Normal University (Zuo et al. 2014). Graphs were estimated using the `ndmg` (Kiar et al. 2018) pipeline. The dMRI scans were preprocessed for eddy currents using FSL’s `eddy-correct` (Smith et al. 2004). FSL’s “standard” linear registration pipeline was used to register the sMRI and dMRI images to the MNI152 atlas (Mazziotta et al. 2001; Woolrich et al. 2009; Jenkinson et al. 2012). A tensor model is fit using DiPy (Garyfallidis et al. 2014) to obtain an estimated tensor at each voxel. A deterministic tractography algorithm is applied using DiPy’s `EuDX` (Garyfallidis et al. 2012) to obtain streamlines, which indicate the voxels connected by an axonal fiber tract. Simple graphs are formed by first contracting voxels into graph vertices according to the Desikan parcellation (Desikan et al. 2006) and then by placing an edge between two regions if a fiber tract is observed between any pair of voxels from each of 70 regions. Further, we have available a network covariate  $C = (c_1, \dots, c_{113})$  measuring subject creativity, which is related to the person’s performance on a series of creativity tasks.

To study this network sample and use the covariate  $C$ , we introduce a direct extension of our results for comparing two network samples:

**Corollary 1 (Two-sample test).** Fix a set of subgraphs  $\mathcal{F}$  and two network samples  $\mathcal{G}$  and  $\mathcal{G}'$  generated, respectively, by the kernels  $f$  and  $f'$  and the network size sequences  $n$  and  $n'$ . Then, as both  $|\mathcal{G}|$  and  $|\mathcal{G}'|$  tend to infinity, and if  $\min(n, n') \geq 2 \max_{F \in \mathcal{F}} |F|$ , we have that if  $f = f'$ , then

$$\frac{\sqrt{|\mathcal{G}||\mathcal{G}'|}}{\sqrt{|\mathcal{G}| + |\mathcal{G}'|}} (\hat{\mu}_{\mathcal{F}}(\mathcal{G}) - \hat{\mu}_{\mathcal{F}}(\mathcal{G}')) \rightarrow \text{Normal}(0, \Sigma_{\mathcal{F}}(n, n', f)),$$

and, as (2) applies, we can produce an unbiased  $\sqrt{|\mathcal{G}| + |\mathcal{G}'|}$ -consistent and asymptotically normal estimator of  $\text{vec}(\Sigma_{\mathcal{F}}(n, n', f))$ , this without specifying or estimating  $f$ .

In the following we choose to work with  $\mathcal{F} = \{\bullet\!\!\!\rightarrow, \blacktriangle, \blacksquare\}$ . We chose these for several reasons: in a simulation experiment (see the supplementary materials) we find that using these counts leads to power surpassing the best available alternative in the literature (Tang, Athreya, Sussman, Lyzinski, Park, et al. 2017; Tang, Athreya, Sussman, Lyzinski, Priebe, et al. 2017); the associated counts correspond to spectral moments of the underlying kernel (Maugis, Olhede, and Wolfe 2017b);  $\bullet\!\!\!\rightarrow$  and  $\blacktriangle$  are staples of graph analysis (related to conductance, clustering, transitivity, etc.) and are key network summaries throughout the literature;  $\blacksquare$  is one of the smallest graphs that is larger than  $\bullet\!\!\!\rightarrow$  and  $\blacktriangle$  but cannot be built from copies of  $\bullet\!\!\!\rightarrow$  and  $\blacktriangle$ , which reduces correlation between counts, and therefore improves power of the tests; all three are small, and there is extensive and available software to efficiently count them (e.g., Jha, Seshadhri, and Pinar 2015; Pinar, Seshadhri, and Vishal 2016; Maugis, Olhede, and Wolfe 2017a.)

We note that our estimator of  $\Sigma_{\mathcal{F}}(n, n', f)$  is entrywise Normal, not Wishart as in the classical setting. Thus, we have no guarantee that  $\hat{\Sigma}_{\mathcal{F}}(n, n', f)$  is positive definite, and cannot use the Hotelling’s  $T$ -squared distribution to compute  $p$ -values. If the estimate is positive definite, we recommend ignoring the variations in  $\hat{\Sigma}_{\mathcal{F}}(n, n', f)$  and using the  $\chi^2_{|\mathcal{F}|}$  distribution. If

the estimate is not positive definite, we recommend using the marginals.

Before analyzing  $\mathcal{G}$  using our results, we make the following test: we subsample uniformly at random and without replacement from  $\mathcal{G}$ , yielding  $\mathcal{G}_1$  and  $\mathcal{G}_2$  such that  $\mathcal{G}_1 \cup \mathcal{G}_2 = \mathcal{G}$ , and use Corollary 1 to test for  $\mathcal{G}_1$  and  $\mathcal{G}_2$  being drawn from the same kernel  $f$ . Unless  $\mathcal{G}$  presents characteristics that cannot be explained by our results,  $\mathcal{G}_1$  and  $\mathcal{G}_2$  should be indistinguishable, and we expect to see  $p$ -values that are uniformly distributed in  $[0, 1]$ .

We perform this experiment 100 times, and obtain a sample of  $p$ -values for which we fail to reject the null of a uniform distribution using the Kolmogorov–Smirnov test ( $D = 0.09$ ,  $p$ -value = 0.3). For this test we use  $\mathcal{F} = \{\blacksquare\}$  because of the small sample ( $|\mathcal{G}_1| + |\mathcal{G}_2| = 113$ ) size and a very high level of correlation; otherwise the correlation between the counts is so high that the correlation matrix appears singular.

We now use the creativity scores to split  $\mathcal{G}$  into two samples. To do so, we build a first subsample  $\mathcal{G}_1$  containing the less creative, and a second subsample  $\mathcal{G}_2$  containing the more creative. More precisely, for a quantile  $q$  and denoting  $Q_C$  the empirical quantile function of  $C$ :

$$\begin{aligned} \mathcal{G}_1^q &= \{G_i \in \mathcal{G} : c_i \leq Q_C(q)\} \quad \text{and} \\ \mathcal{G}_2^q &= \{G_i \in \mathcal{G} : c_i > Q_C(1 - q)\}. \end{aligned}$$

Interestingly, for  $q = 0.5$  and  $q = 0.4$ , we fail to reject the null that the networks in  $\mathcal{G}_1^q$  and  $\mathcal{G}_2^q$  come from the same kernel (see Table 1 for the  $p$ -values). However, for  $q = 0.3$  we can reject the null of the same kernel at the 5% confidence level using  $\blacktriangle$  or  $\blacksquare$ , but not  $\bullet\!\!\!\rightarrow$ .

Thus, we observe that individuals with a very high level of creativity present significantly more  $\blacktriangle$  and  $\blacksquare$  than those with a very low level of creativity. To further confirm this discovery, we undertake the following experiment: we estimate kernels (through the random-dot-product framework; Athreya et al. 2018) over the samples  $\mathcal{G}$ ,  $\mathcal{G}_1^{0.2}$ , and  $\mathcal{G}_2^{0.2}$ , that we call  $f$ ,  $f_1$ , and  $f_2$ , respectively; then, for several  $\gamma \in [0, 1]$ , we consider the power of Corollary 1’s test when the samples compared are of the same cardinality as  $\mathcal{G}_1^{0.2}$ , and  $\mathcal{G}_2^{0.2}$ , but drawn iid from  $G_{70}(\gamma f_1 + (1 - \gamma)f)$  and  $G_{70}(\gamma f_2 + (1 - \gamma)f)$ , respectively; we find that the proposed test is very powerful, as powerful as a semiparametric test relying on the random-dot-product structure of the model (Tang, Athreya, Sussman, Lyzinski, Park, et al. 2017). This provides comfort for the claim that the difference in samples is only observed for  $q \geq 0.3$ , and not for more central quantiles.

**Table 1.** Testing for differences between  $\mathcal{G}_1^q$  and  $\mathcal{G}_2^q$ .

Quantile ( $q$ )	p-value		
	$\bullet\!\!\!\rightarrow$	$\blacktriangle$	$\blacksquare$
0.5	0.126	0.110	0.115
0.4	0.077	0.051	0.050
0.3	0.062	0.042	0.040
0.2	0.014	0.011	0.012
0.1	0.046	0.047	0.061

NOTE: For each  $F \in \{\bullet\!\!\!\rightarrow, \blacktriangle, \blacksquare\}$  and  $q \in \{0.1, 0.2, \dots, 0.5\}$  we produce the  $p$ -value for the null  $H_0 : \mu_F(\mathcal{G}_1^q) = \mu_F(\mathcal{G}_2^q)$ . The  $p$ -values increase with  $q$ , except for  $q = 0.1$ , in which case  $|\mathcal{G}_1^q|$  and  $|\mathcal{G}_2^q|$  are too small for the test to be significant.



We now aim to understand whether the added  $\blacktriangle$  and  $\blacklozenge$  arise from a few edges completing partially present shapes or from fully new  $\blacktriangle$  and  $\blacklozenge$ . To do so, we first observe that if  $G \sim G_n(f)$ , then  $\bar{G} \sim G_n(1 - f)$ , where  $\bar{G}$  is the complement graph of  $G$ . Therefore, we may use our tests on  $\bar{G}$ , which can be understood as estimating  $\mu_{\mathcal{F}}(1 - f)$  instead of  $\mu_{\mathcal{F}}(f)$  to compare network samples.

Then, using the  $\bar{\mathcal{G}}_i^q = \{\bar{G} : G \in \mathcal{G}_i^q\}$ , we can test whether there are significantly more fully absent subgraphs in  $\mathcal{G}_1^q$  compared to  $\mathcal{G}_2^q$ . There, we find we cannot reject this null; that is, we cannot reject the null of the networks in  $\bar{\mathcal{G}}_1^q$  and  $\bar{\mathcal{G}}_2^q$  coming from the same kernel for  $q \geq 0.3$ . Therefore, we conclude that the added  $\blacktriangle$  and  $\blacklozenge$  in the highly creative arise from a few edges completing partially present  $\blacktriangle$  and  $\blacklozenge$ .

One key conclusion of our analysis is that we must use [Theorem 2](#) to study  $\mathcal{G}$ . Indeed, we have just shown that all networks generating  $\mathcal{G}$  cannot come from the same kernel. This allows us to write that using the full sample  $\mathcal{G}$ , a centered and consistent estimate of the average density of  $\blackrightarrow$ ,  $\blacktriangle$ ,  $\blacklozenge$  within human brains in our modality is (0.41, 0.13, 0.06). To conclude, we remark that since all the  $G_i$ 's in  $\mathcal{G}$  have the same number of nodes, one could produce a similar analysis using—instead of our results—the  $\mu_{\mathcal{F}}(G_i)$  as if they formed an independent and identically distributed sample. However, as we have just shown, the  $G_i$  are not identically distributed, and therefore such an analysis could lead to spurious conclusions.

## 6. Discussion

We provide the tools to perform statistical inference on a network sample using subgraph counts. Our two main results provide consistency and asymptotic normality of subgraph counts under very flexible conditions. Using these results, we show that subgraph counts are powerful statistics to test whether network samples come from a specified distribution, a specified model, or from the same model.

The key insight we provide is that statistical inference methods paralleling classical ones for standard samples may be obtained for network samples. From this perspective, our results may be seen as providing an analogue of a multivariate  $t$ -test for network samples. However, going beyond what our results directly imply, we expect that parallels to ANOVA, model selection, model ranking, and goodness of fit may be obtained for network samples using our proof techniques.

However, since our tests are analogous to the  $t$ -test, they rely on global summaries. Therefore, while we can reject the null of two network samples being drawn from the same model, our approach is unable to locate where in the network the difference is realized.

## Appendix A. Properties of Subgraph Counts

In the following, we formalize certain notions we use loosely in the main body (especially the notion of copy and the sets  $\mathcal{H}_{F'}$ , as well as the constants  $c_H$ ), prove all our results, and provide more details on the numerical examples we present.

We start by proving our first lemma, establishing the linearity of subgraph counts. We first define  $\mathcal{H}_{F_1 F_2}$  and  $c_H$ , generalizing definitions given in [Rucinski \(1988\)](#).

**Definition 5 (Overlapping copies).** For two graphs  $F_1$  and  $F_2$  we denote by  $\mathcal{H}_{F_1 F_2}$  the set of unlabeled graphs that can be formed by two copies of  $F_1$  and  $F_2$ , and  $c_H$  the number of ways a given  $H \in \mathcal{H}_{F_1 F_2}$  can be built from copies of  $F_1$  and  $F_2$ :

$$\begin{cases} \mathcal{H}_{F_1 F_2} &= \left\{ H \subset K_{|F_1|+|F_2|} : \begin{array}{l} \exists F'_1, F'_2 \subset K_{|F_1|+|F_2|} \text{ s.t.,} \\ F'_1 \equiv F_1, F'_2 \equiv F_2, \text{ and} \\ H = F'_1 \cup F'_2 \end{array} \right\} / \equiv \\ c_H &= \# \left\{ (F'_1, F'_2) \subset H : \begin{array}{l} F'_1 \equiv F_1, F'_2 \equiv \\ F_2 \\ \text{and } H = F'_1 \cup F'_2 \end{array} \right\}. \end{cases}$$

Finally, call  $\mathcal{H}_{F_1 F_2}^*$  the set  $\mathcal{H}_{F_1 F_2}$  removed of  $F_1 \sqcup F_2$ , the vertex disjoint union of  $F_1$  and  $F_2$ .

Note that  $\mathcal{H}_{F_1 F_2}$  is defined as a quotient set, sometimes called quotient space, through the equivalence relation  $\equiv$ . In the following, we identify the equivalence classes in  $\mathcal{H}_{F_1 F_2}$  with any of their element, and therefore will treat the elements of  $\mathcal{H}_{F_1 F_2}$  as graphs.

**Lemma 2 (Copies pairwise interaction).** Fix three graphs  $F_1, F_2$ , and  $G$ . Then,

$$X_{F_1}(G)X_{F_2}(G) = \sum_{H \in \mathcal{H}_{F_1 F_2}} c_H X_H(G).$$

*Proof.* We start by writing:

$$\begin{aligned} X_{F_1}(G)X_{F_2}(G) &= \sum_{F'_1 \subset G} 1_{\{F'_1 \equiv F_1\}} \sum_{F'_2 \subset G} 1_{\{F'_2 \equiv F_2\}} \quad (\text{A.1}) \\ &= \sum_{\substack{F'_1 \subset G \\ F'_2 \subset G}} 1_{\{F'_1 \equiv F_1\}} 1_{\{F'_2 \equiv F_2\}}. \end{aligned}$$

Now, from [Definition 5](#), we first note that by construction of  $\mathcal{H}_{F_1 F_2}$ , for each pair  $F'_1, F'_2$  in the sum,  $1_{\{F'_1 \equiv F_1\}} 1_{\{F'_2 \equiv F_2\}} = 1$  if and only if there exists  $H \in \mathcal{H}_{F_1 F_2}$  such that  $F'_1 \cup F'_2 \equiv H$ . Therefore, we can reindex the sum in (A.1) as follows:

$$\begin{aligned} X_{F_1}(G)X_{F_2}(G) &= \sum_{F'_1, F'_2 \subset G} 1_{\{\exists H \in \mathcal{H}_{F_1 F_2} : F'_1 \cup F'_2 \equiv H\}} \\ &= \sum_{H \in \mathcal{H}_{F_1 F_2}} \sum_{F'_1, F'_2 \subset G} 1_{\{F'_1 \cup F'_2 \equiv H\}}. \end{aligned}$$

We now note that by definition of  $c_H$ , for each copy of  $H$  in  $G$ , there will be  $c_H$  pairs  $(F'_1, F'_2)$  of copies of  $F_1$  and  $F_2$  in  $G$  such that  $F'_1 \cup F'_2 = H$ . Therefore we can simplify the sum above to obtain:

$$X_{F_1}(G)X_{F_2}(G) = \sum_{H \in \mathcal{H}_{F_1 F_2}} c_H \sum_{H' \subset G} 1_{\{H' \equiv H\}} = \sum_{H \in \mathcal{H}_{F_1 F_2}} c_H X_H(G),$$

yielding the desired result.  $\square$

With these tools in hand, we compute the first two moments of  $X_F(G)$  when  $G \sim G(n, f)$ .

**Proposition 1.** Fix two graphs  $F$  and  $F'$  and a random graph  $G \sim G(|G|, f)$  such that  $|F| + |F'| \leq |G|$ . Then, we have that

$$\begin{aligned} \mathbb{E}X_F(G) &= X_F(K_G)\mu_F(f) \\ \text{cov}(X_F(G), X_{F'}(G)) &= \sum_{H \in \mathcal{H}_{FF'}} c_H X_H(K_G)(\mu_H(f) - \mu_F(f)\mu_{F'}(f)). \end{aligned}$$

*Proof.* The proof is relegated to the supplementary materials.  $\square$

We now turn to the proofs of [Theorems 1](#) and [2](#). As the second generalizes the first, it is sufficient to prove the second.

*Proof.* We obtain the result by a joint application of the Lindeberg–Feller central limit theorem and the Cramer–Wold device. To do so, we fix  $\mathbf{a} \in \mathbb{R}^{|\mathcal{F}|}$  and compute the variance of our estimator projected along  $\mathbf{a}$ .

**Computing the variance:** First recall that  $\hat{\boldsymbol{\mu}}_{\mathcal{F}}(\mathcal{G}) = N^{-1} \sum_{i \in [N]} \left( \frac{X_F(G_i)}{X_F(K_{G_i})} \right)_{F \in \mathcal{F}}$ . Therefore, denoting by “ $\cdot$ ” the inner product and taking the expectation over  $\mathcal{G}$ , let

$$s_N^2 = \text{var}(\mathbf{a} \cdot \hat{\boldsymbol{\mu}}_{\mathcal{F}}(\mathcal{G})).$$

Then, using the independence of the  $G_i$ ’s and the bi-linearity of the covariance, we have

$$\begin{aligned} s_N^2 &= N^{-2} \sum_{i \in [N]} \text{var} \left( \mathbf{a} \cdot \left( \frac{X_F(G_i)}{X_F(K_{G_i})} \right)_{F \in \mathcal{F}} \right) \\ &= N^{-2} \sum_{i \in [N]} \sum_{F, F' \in \mathcal{F}} \frac{\mathbf{a}_F \mathbf{a}_{F'} \text{cov}(X_F(G_i), X_{F'}(G_i))}{X_F(K_{G_i}) X_{F'}(K_{G_i})}. \end{aligned}$$

To proceed, recall that for each  $i \leq N$ ,  $G_i \sim G(n_i, f_i)$ . Then, we may use [Proposition 1](#) to obtain

$$\begin{aligned} n^2 s_N^2 &= \sum_{i \in [N]} \sum_{F, F' \in \mathcal{F}} \frac{\mathbf{a}_F \mathbf{a}_{F'}}{X_F(K_{G_i}) X_{F'}(K_{G_i})} \\ &\quad \sum_{H \in \mathcal{H}_{FF'}^*} c_H X_H(K_{G_i}) (\mu_H(f_i) - \mu_F(f_i) \mu_{F'}(f_i)) \end{aligned}$$

Because all sums are finite, we can reorder the summations, leading to

$$\begin{aligned} s_N^2 &= N^{-1} \sum_{F, F' \in \mathcal{F}} \mathbf{a}_F \mathbf{a}_{F'} \sum_{H \in \mathcal{H}_{FF'}^*} \left( N^{-1} \sum_{i \in [N]} \frac{c_H X_H(K_{G_i})}{X_F(K_{G_i}) X_{F'}(K_{G_i})} \right. \\ &\quad \left. (\mu_H(f_i) - \mu_F(f_i) \mu_{F'}(f_i)) \right) \\ &= N^{-1} \sum_{F, F' \in \mathcal{F}} \mathbf{a}_F \mathbf{a}_{F'} \sum_{H \in \mathcal{H}_{FF'}^*} \omega_H(n, f; N), \end{aligned} \tag{A.2}$$

where  $\omega_H(n, f; N) = N^{-1} \sum_{i \in [N]} \frac{c_H X_H(K_{G_i})}{X_F(K_{G_i}) X_{F'}(K_{G_i})} (\mu_H(f_i) - \mu_F(f_i) \mu_{F'}(f_i))$ .

**Convergence of  $\omega_H(n, f; N)$ :** To proceed we must show that  $\omega_H(n, f; N)$  converges to a limit as  $N$  diverges. We will achieve this by showing that  $\omega_H(n, f; N)$  is Cauchy. To do so, observe that

$$\begin{aligned} &|\omega_H(n, f; N + 1) - \omega_H(n, f; N)| \\ &= \left| (N + 1)^{-1} \left( N \omega_H(n, f; N) - (N + 1) \omega_H(n, f; N) \right) \right. \\ &\quad \left. + \frac{c_H X_H(K_{G_{N+1}})}{X_F(K_{G_{N+1}}) X_{F'}(K_{G_{N+1}})} (\mu_H(f_{N+1}) - \mu_F(f_{N+1}) \mu_{F'}(f_{N+1})) \right) \\ &= \left| \frac{(N + 1)^{-1} c_H X_H(K_{G_{N+1}})}{X_F(K_{G_{N+1}}) X_{F'}(K_{G_{N+1}})} (\mu_H(f_{N+1}) \right. \\ &\quad \left. - \mu_F(f_{N+1}) \mu_{F'}(f_{N+1})) - \frac{\omega_H(n, f; N)}{N + 1} \right|. \end{aligned}$$

Then, we recall that  $X_F(K_G) = \binom{|G|}{|F|} \text{aut}(F)$  (see, e.g., [Bollobás and Riordan 2009](#)), where  $\text{aut}(F)$  is the number of automorphisms of  $F$

(the number of bijections from the vertex set of  $F$  to itself that preserve adjacency). Then, as  $|H| < |F| + |F'|$ , we have

$$\begin{aligned} \frac{X_H(K_{G_{N+1}})}{X_F(K_{G_{N+1}}) X_{F'}(K_{G_{N+1}})} &= \frac{\text{aut}(H)}{\text{aut}(F) \text{aut}(F')} \frac{\binom{n_{N+1}}{|H|}}{\binom{n_{N+1}}{|F|} \binom{n_{N+1}}{|F'|}} \\ &\leq \frac{\text{aut}(H)}{\text{aut}(F) \text{aut}(F')}. \end{aligned}$$

As furthermore  $f_{N+1}$  is bounded by 1, we have  $|(\mu_H(f_{N+1}) - \mu_F(f_{N+1}) \mu_{F'}(f_{N+1}))| \leq 1$ , leading to  $|\omega_H(n, f; N + 1) - \omega_H(n, f; N)| \leq \frac{2C}{N+1}$  for  $C = 1 + c_H \text{aut}(H) / \text{aut}(F) \text{aut}(F')$ . Thus, the sequence  $\omega_H(n, f; N)$  is Cauchy, and we may call  $\omega_H(n, f)$  its limit; that is,

$$\lim_{N \rightarrow \infty} \omega_H(n, f; N) = \omega_H(n, f).$$

Then, resuming from [\(A.2\)](#), and writing  $\boldsymbol{\Sigma}_{\mathcal{F}}$  the matrix indexed by  $\mathcal{F}$  such that

$$(\boldsymbol{\Sigma}_{\mathcal{F}}(n, f))_{FF'} = \sum_{H \in \mathcal{H}_{FF'}^*} \omega_H(n, f),$$

we have  $\lim_{N \rightarrow \infty} N s_N^2 = \mathbf{a}^\top \boldsymbol{\Sigma}_{\mathcal{F}}(n, f) \mathbf{a}$ .

**Satisfying the Lindeberg–Feller condition:** To invoke the Lindeberg–Feller central limit theorem, we must show that our sequence verifies the so called Lindeberg–Feller condition. Recall that the sequence under study is

$$Y_i := \mathbf{a} \cdot (X_F(G_i) / X_F(K_{G_i}))_{F \in \mathcal{F}} - \mathbf{a} \cdot \boldsymbol{\mu}_{\mathcal{F}}(f_i),$$

and that the variance of the partial sum is  $N^2 s_N^2$ . Therefore, the Lindeberg–Feller condition we need to satisfy is the following:

$$\forall \epsilon > 0 \lim_{N \rightarrow \infty} \frac{1}{N^2 s_N^2} \sum_{i \in [N]} \mathbb{E} [Y_i 1_{\{|Y_i| > \epsilon N s_N\}}] = 0. \tag{A.3}$$

To verify the condition we first fix  $\epsilon > 0$ . Then, observe that since for each  $i$  and  $F$  we have  $X_F(G_i) / X_F(K_{G_i}) \leq 1$  and  $\mu_F(f_i) \leq 1$ , we have that  $|Y_i| \leq 2 \|\mathbf{a}\|_1$  by the triangle inequality. Therefore, as  $N s_N \rightarrow \infty$  as  $N$  grows, we may fix an  $N$  such that for all  $N' > N$  we have  $\|\mathbf{a}\|_1 \leq \epsilon N s_N$ . In this setting, the sum in [\(A.3\)](#) is equal to zero for all  $N' > N$ , and the condition is verified. Therefore, we have that

$$\frac{1}{N s_N} \sum_{i \in [N]} Y_i \rightarrow \text{Normal}(0, 1).$$

Reverting to the notation of the statement of the theorem, we have that for any  $\mathbf{a}$

$$\sqrt{N}(\mathbf{a} \cdot \hat{\boldsymbol{\mu}}_{\mathcal{F}}(\mathcal{G}) - \mathbf{a} \cdot \boldsymbol{\mu}_{\mathcal{F}}(f)) \rightarrow \text{Normal}(0, \mathbf{a}^\top \boldsymbol{\Sigma}_{\mathcal{F}}(n, f) \mathbf{a}),$$

which is sufficient to obtain the claimed result for the limit in distribution.  $\square$

We now turn to the proof of [Corollary 1](#).

*Proof.* The result follows almost immediately from [Theorem 1](#) and Slutsky’s theorem.

First, since  $|\mathcal{G}|$  and  $|\mathcal{G}'|$  tend to infinity and both  $n$  and  $n'$  are large enough, we have

$$\begin{cases} \sqrt{|\mathcal{G}|}(\hat{\boldsymbol{\mu}}_{\mathcal{F}}(\mathcal{G}) - \boldsymbol{\mu}_{\mathcal{F}}(f)) \rightarrow \text{Normal}(0, \boldsymbol{\Sigma}_{\mathcal{F}}(n, f)), \\ \sqrt{|\mathcal{G}'|}(\hat{\boldsymbol{\mu}}_{\mathcal{F}'}(\mathcal{G}') - \boldsymbol{\mu}_{\mathcal{F}'}(f')) \rightarrow \text{Normal}(0, \boldsymbol{\Sigma}_{\mathcal{F}'}(n', f')). \end{cases}$$

As both samples are independent, linear combinations are also multivariate Gaussian. Therefore, multiplying the first line by  $\sqrt{|\mathcal{G}'|} / \sqrt{|\mathcal{G}| + |\mathcal{G}'|}$ , the second by  $\sqrt{|\mathcal{G}|} / \sqrt{|\mathcal{G}| + |\mathcal{G}'|}$  (both

ratios being in  $(0,1)$ , the limit in distribution is unaffected), and taking the difference, we obtain, with  $\Sigma_{\mathcal{F}}(n, n', f) = \lim_{|\mathcal{G}|, |\mathcal{G}'| \rightarrow \infty} \left\{ \frac{|\mathcal{G}'|}{|\mathcal{G}| + |\mathcal{G}'|} \Sigma_{\mathcal{F}}(n, f) + \frac{|\mathcal{G}|}{|\mathcal{G}| + |\mathcal{G}'|} \Sigma_{\mathcal{F}}(n', f) \right\}$

$$\frac{\sqrt{|\mathcal{G}||\mathcal{G}'|}}{\sqrt{|\mathcal{G}| + |\mathcal{G}'|}} (\hat{\mu}_{\mathcal{F}}(\mathcal{G}) - \hat{\mu}_{\mathcal{F}}(\mathcal{G}')) \rightarrow \text{Normal}(0, \Sigma_{\mathcal{F}}(n, n', f)),$$

which is the desired limit in distribution.

To obtain the consistent estimator of  $\Sigma_{\mathcal{F}}(n, n', f)$  we first recall that for  $F, F' \in \mathcal{F}$

$$(\Sigma_{\mathcal{F}}(n, f))_{FF'} = \sum_{H \in \mathcal{H}_{FF'}^*} \omega_H(n) (\mu_H(f) - \mu_{F'}(f) \mu_{F'}(f)).$$

Then, with  $\omega_H(n, n') = \lim_{|\mathcal{G}|, |\mathcal{G}'| \rightarrow \infty} \left( \frac{|\mathcal{G}'|}{|\mathcal{G}| + |\mathcal{G}'|} \omega_H(n) + \frac{|\mathcal{G}|}{|\mathcal{G}| + |\mathcal{G}'|} \omega_H(n') \right)$ , we have

$$(\Sigma_{\mathcal{F}}(n, n', f))_{FF'} = \sum_{H \in \mathcal{H}_{FF'}^*} \omega_H(n, n') (\mu_H(f) - \mu_{F'}(f) \mu_{F'}(f)).$$

There observe that:

- As  $|\mathcal{G}|$  and  $|\mathcal{G}'|$  grow we have that

$$\begin{aligned} \omega_H(n, n'; |\mathcal{G}|, |\mathcal{G}'|) &= \frac{|\mathcal{G}'|}{|\mathcal{G}| + |\mathcal{G}'|} \omega_H(n; |\mathcal{G}|) \\ &+ \frac{|\mathcal{G}|}{|\mathcal{G}| + |\mathcal{G}'|} \omega_H(n'; |\mathcal{G}'|) \rightarrow \omega_H(n, n'). \end{aligned}$$

- All the  $\mu_H(f)$  in  $H \in \mathcal{H}_{FF'}^*$  may be estimated by  $\hat{\mu}_H(\mathcal{G} \cup \mathcal{G}') = \frac{1}{|\mathcal{G} \cup \mathcal{G}'|} \sum_{G \in \mathcal{G} \cup \mathcal{G}'} \frac{X_H(G)}{X_H(K_G)}$ . Furthermore, both  $\mu_{F'}(f)$  and  $\mu_{F'}(f)$  may be estimated in the same way.

Then, by a direct application of the Slutsky's theorem, we have that

$$\begin{aligned} (\hat{\Sigma}_{\mathcal{F}}(n, n', f))_{FF'} &= \sum_{H \in \mathcal{H}_{FF'}^*} \omega_H(n, n'; |\mathcal{G}|, |\mathcal{G}'|) (\hat{\mu}_H(\mathcal{G} \cup \mathcal{G}') \\ &- \hat{\mu}_{F'}(\mathcal{G} \cup \mathcal{G}') \hat{\mu}_{F'}(\mathcal{G} \cup \mathcal{G}')) \end{aligned}$$

is an asymptotically normal estimator of  $(\Sigma_{\mathcal{F}}(n, n', f))_{FF'}$  converging at rate  $\sqrt{|\mathcal{G}| + |\mathcal{G}'|}$ , which yields the claimed result.  $\square$

## Supplementary Materials

The supplementary files contain figures summarizing computer experiments in line with our mathematical derivations; two experiments exploring the power of two of our proposed tests; detail of the procedures used to generate the figures in the main body of the article; the proof of Proposition 1.

## Acknowledgments

The authors thank Joshua T. Vogelstein and his team for connectome data and expertise.

## Funding

This work was supported in part by DARPA under SIMPLEX contract N66001-15-C-4041; by the US Army Research Office under

Multidisciplinary University Research Initiative Award 58153-MA-MUR; by the US Office of Naval Research under Award N00014-14-1-0819; by the UK Engineering and Physical Sciences Research Council under Mathematical Sciences Leadership Fellowship EP/I005250/1, Established Career Fellowship EP/K005413/1, Developing Leaders Award EP/L001519/1, and Award EP/N007336/1; by the UK Royal Society under a Wolfson Research Merit Award; and by Marie Curie FP7 Integration Grant PCIG12-GA-2012-334622 and the European Research Council under grant CoG 2015-682172NETS, both within the Seventh European Union Framework Program. The authors thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, for support and hospitality during the program Theoretical Foundations for Statistical Network Analysis (EPSRC grant no. EP/K032208/1) where a portion of the work on this article was undertaken.

## References

- Ali, T., Rito, T., Reinert, G., Sun, F., and Deane, C. M. (2014), "Alignment-Free Protein Interaction Network Comparison," *Bioinformatics*, 30, i430–i437. [455,456]
- Ali, W., Wegner, A. E., Gaunt, R. E., Deane, C. M., and Reinert, G. (2016), "Comparison of Large Networks With Sub-Sampling Strategies," *Scientific Reports*, 6, 28955. [455,456]
- Alon, N., Yuster, R., and Zwick, U. (1997), "Finding and Counting Given Length Cycles," *Algorithmica*, 17, 209–223. [456]
- Asta, D., and Shalizi, C. R. (2014), "Geometric Network Comparison," arXiv no. 1411.1350. [455,456]
- Athreya, A., Fishkind, D. E., Levin, K., Lyzinski, V., Park, Y., Qin, Y., Sussman, D. L., Tang, M., Vogelstein, J. T., and Priebe, C. E. (2018), "Statistical Inference on Random Dot Product Graphs: A Survey," *The Journal of Machine Learning Research*, 18, 1–92. [461]
- Banerjee, D., and Ma, Z. (2017), "Optimal Hypothesis Testing for Stochastic Block Models With Growing Degrees," arXiv no. 1705.05305. [455]
- Benson, A. R., Gleich, D. F., and Leskovec, J. (2016), "Higher-Order Organization of Complex Networks," *Science*, 353, 163–166. [456]
- Bhattacharyya, S., and Bickel, P. J. (2015), "Subsampling Bootstrap of Count Features of Networks," *The Annals of Statistics*, 43, 2384–2411. [455,456,457]
- Bickel, P. J., Chen, A., and Levina, E. (2012), "The Method of Moments and Degree Distributions for Network Models," *The Annals of Statistics*, 39, 2280–2301. [455,456]
- Birmele, E. (2012), "Detecting Local Network Motifs," *Electronic Journal of Statistics*, 6, 908–933. [455]
- Bollobás, B., and Riordan, O. (2009), "Metrics for Sparse Graphs," in *Surveys in Combinatorics 2009*, eds. S. Huczynska, J. D. Mitchell, and C. M. Roney-Dougal, Cambridge, UK: Cambridge University Press, pp. 211–287. [457,463]
- Chatterjee, S., and Diaconis, P. (2013), "Estimating and Understanding Exponential Random Graph Models," *The Annals of Statistics*, 41, 2428–2461. [456]
- Coulson, M., Gaunt, R. E., and Reinert, G. (2016), "Poisson Approximation of Subgraph Counts in Stochastic Block Models and a Graphon Model," arXiv no. 1509.07754. [455,458]
- Daianu, M., Jahanshad, N., Nir, T., Toga, A., Jack, C., Weiner, M., and Thompson, P. (2013), "Breakdown of Brain Connectivity Between Normal Aging and Alzheimer's Disease: A Structural  $k$ -Core Network Analysis," *Brain Connectivity*, 3, 407–422. [455]
- Desikan, R., Segonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B., Albert, M., and Killiany, R. (2006), "An Automated Labeling System for Subdividing the Human Cerebral Cortex on MRI Scans Into Gyral Based Regions of Interest," *NeuroImage*, 31, 968–980. [461]
- Diaconis, P., and Janson, S. (2008), "Graph Limits and Exchangeable Random Graphs," *Rendiconti di Matematica e delle sue Applicazioni*, 28, 33–61. [456]
- Durante, D., and Dunson, D. B. (2018), "Bayesian Inference and Testing of Group Differences in Brain Networks," *Bayesian Analysis*, 13, 29–58. [455]
- Fosdick, B., and Hoff, P. D. (2015), "Testing and Modeling Dependencies Between a Network and Nodal Attributes," *Journal of the American Statistical Association*, 110, 1047–1056. [455]

- Gao, C., and Lafferty, J. (2017), “Testing Network Structure Using Relations Between Small Subgraph Probabilities,” arXiv no. 1704.06742. [455,456]
- Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., and Nimmo-Smith, I. (2014), “Dipy, a Library for the Analysis of Diffusion MRI Data,” *Frontiers in Neuroinformatics*, 8, 8. [461]
- Garyfallidis, E., Brett, M., Correia, M., Williams, G., and Nimmo-Smith, I. (2012), “Quickbundles, a Method for Tractography Simplification,” *Frontiers in Neuroscience*, 6, 175. [461]
- Ghoshdastidar, D., von Luxburg, U., Gutzeit, M., and Carpentier, A. (2017), “Two-Sample Tests for Large Random Graphs Using Network Statistics,” arXiv no. 1705.06168. [455]
- Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., and Kolaczyk, E. D. (2017), “Hypothesis Testing for Network Data in Functional Neuroimaging,” *The Annals of Applied Statistics*, 11, 725–750. [455]
- Gray, W. R., Bogovic, J. A., Vogelstein, J. T., Landman, B. A., Prince, J. L., and Vogelstein, R. J. (2012), “Magnetic Resonance Connectome Automated Pipeline: An Overview,” *IEEE Pulse*, 3, 42–48. [455]
- Ho, Q., Parikh, A. P., and Xing, E. P. (2012), “A Multiscale Community Blockmodel for Network Exploration,” *Journal of the American Statistical Association*, 107, 916–934. [455]
- Hočevár, T., and Demšar, J. (2014), “A Combinatorial Approach to Graphlet Counting,” *Bioinformatics*, 30, 559–565. [456,457]
- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2012), “Latent Space Approaches to Social Network Analysis,” *Journal of the American Statistical Association*, 97, 1090–1098. [455,456,457]
- Jenkinson, M., Beckmann, C., Behrens, T., Woolrich, M., and Smith, S. (2012), “FSL,” *NeuroImage*, 62, 782–90. [461]
- Jha, M., Seshadhri, C., and Pinar, A. (2015), “Path Sampling: A Fast and Provable Method for Estimating 4-Vertex Subgraph Counts,” in *Proceedings of the 24th International Conference on World Wide Web, WWW '15*, New York, NY, USA, ACM, pp. 495–505. [461]
- Kiar, G., Bridgeford, E., Roncal, W., Consortium for Reliability and Reproducibility (CoRR), Chandrashekar, V., Mhembere, D., Ryman, S., Zuo, X., Margulies, D., Craddock, R., Priebe, C., Jung, R., Calhoun, V., Caffo, B., Burns, R., Milham, M., and Vogelstein, J. (2018), “A High-Throughput Pipeline Identifies Robust Connectomes but Troublesome Variability,” bioRxiv no. 188706. [459,461]
- Klopp, O., Tsybakov, A. B., and Verzelen, N. (2016), “Oracle Inequalities for Network Models and Sparse Graphon Estimation,” *The Annals of Statistics*, 45, 316–354. [455]
- Koutra, D., Vogelstein, J. T., and Faloutsos, C. (2013), “DELTA CON: A Principled Massive-Graph Similarity Function,” in *Proceedings of the SIAM International Conference in Data Mining*, SIAM, pp. 162–170. [459,460]
- Lovász, L. (2012), *Large Networks and Graph Limits*, Providence, RI: American Mathematical Society. [456,457]
- Maugis, P.-A. G., Olhede, S. C., and Wolfe, P. J. (2017a), “Fast Counting of Medium-Sized Rooted Subgraphs,” arXiv no. 1701.00177. [461]
- (2017b), “Topology Reveals Universal Features for Network Comparison,” arXiv no. 1705.056777. [461]
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Feidler, J., Smith, K., Boomsma, D., Hulshoff, P., Cannon, T., Kawashima, R., and Mazoyer, B. (2001), “A Four-Dimensional Probabilistic Atlas of the Human Brain,” *Journal of the American Medical Informatics Association*, 8, 401–430. [461]
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002), “Network Motifs: Simple Building Blocks of Complex Networks,” *Science*, 298, 824–827. [456]
- Olhede, S. C., and Wolfe, P. J. (2014), “Network Histograms and Universality of Blockmodel Approximation,” *Proceedings of the National Academy of Sciences of the United States of America*, 111, 14722–14727. [455]
- Ortmann, M., and Brandes, U. (2016), “Quad Census Computation: Simple, Efficient, and Orbit-Aware,” in *International Conference and School on Network Science*, Cham: Springer International Publishing, pp. 1–13. [456]
- Pinar, A., Seshadhri, C., and Vishal, V. (2016), “Escape: Efficiently Counting all 5-Vertex Subgraphs,” arXiv no. 1610.09411. [461]
- Rucinski, A. (1988), “When Are Small Subgraphs of a Random Graph Normally Distributed?,” *Probability Theory and Related Fields*, 78, 1–10. [456,457,458,462]
- Simpson, S., Moussa, M., and Laurienti, P. (2012), “An Exponential Random Graph Modeling Approach to Creating Group-Based Representative Whole-Brain Connectivity Networks,” *NeuroImage*, 60, 1117–1126. [455]
- Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., Bannister, P., De Luca, M., Drobnjak, I., Flitney, D., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J., and Matthews, P. (2004), “Advances in Functional and Structural MR Image Analysis and Implementation as FSL,” *NeuroImage*, 23, S208–S219. [461]
- Stoffers, D., Berendse, H., Olde Dubbelink, K., Hillebrand, A., Stam, C., Deijen, J., and Twisk, J. (2013), “Disrupted Brain Network Topology in Parkinson’s Disease: A Longitudinal Magnetoencephalography Study,” *Brain*, 137, 197–207. [455]
- Sussman, D. L., Tang, M., Fishkind, D. E., and Priebe, C. E. (2012), “A Consistent Adjacency Spectral Embedding for Stochastic Blockmodel Graphs,” *Journal of the American Statistical Association*, 107, 1119–1128. [455,457]
- Talukder, N., and Zaki, M. J. (2016), “A Distributed Approach for Graph Mining in Massive Networks,” *Data Mining and Knowledge Discovery*, 30, 1024–1052. [456]
- Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., Park, Y., and Priebe, C. E. (2017), “A Semiparametric Two-Sample Hypothesis Testing for Random Dot Product Graphs,” *Journal of Computational and Graphical Statistics*, 26, 344–354. [461]
- Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., and Priebe, C. E. (2017), “A Nonparametric Two-Sample Hypothesis Testing for Random Dot Product Graphs,” *Bernoulli*, 23, 1599–1630. [455,456,461]
- Tang, M., Sussman, D. L., and Priebe, C. E. (2013), “Universally Consistent Vertex Classification for Latent Positions Graphs,” *The Annals of Statistics*, 41, 1406–1430. [455]
- Wang, L., Zhang, Z., and Dunson, D. (2019), “Common and Individual Structure of Brain Networks,” *The Annals of Applied Statistics*, 13, 85–112. [455,456]
- Woolrich, M., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. (2009), “Bayesian Analysis of Neuroimaging Data in FSL,” *NeuroImage*, 45, S173–S186. [461]
- Zuo, X.-N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., Breitner, J. C., Buckner, R. L., Calhoun, V. D., Castellanos, F. X., and Chen, A. (2014), “An Open Science Resource for Establishing Reliability and Reproducibility in Functional Connectomics,” *Scientific Data*, 1, 140049. [461]