

RESEARCH ARTICLE

Nonparametric two-sample test for networks using joint graphon estimation

Benjamin Sischka¹ and Göran Kauermann² 

¹Allianz Private Krankenversicherungs-AG, Risikomanagementfunktion, Munich, Germany and ²Department of Statistics, Ludwig-Maximilians-Universität München, München, Germany

Corresponding author: Göran Kauermann; Email: goeran.kauermann@lmu.de

Abstract

This paper focuses on the comparison of networks on the basis of statistical inference. For that purpose, we rely on smooth graphon models as a nonparametric modeling strategy that is able to capture complex structural patterns. The graphon itself can be viewed more broadly as local density or intensity function on networks, making the model a natural choice for comparison purposes. More precisely, to gain information about the (dis-)similarity between networks, we extend graphon estimation towards modeling multiple networks simultaneously. In particular, fitting a single model implies aligning different networks with respect to the same graphon estimate. To do so, we employ an EM-type algorithm. Drawing on this network alignment consequently allows a comparison of the edge density at local level. Based on that, we construct a chi-squared-type test on equivalence of network structures. Simulation studies and real-world examples support the applicability of our network comparison strategy.

Keywords: Network comparison; EM algorithm; Gibbs sampler; B-spline regression; Chi-squared test

1. Introduction

The field of statistical modeling and analysis of complex networks has gained increasing interest over the last two decades. This is driven by the fact that different types of systems can be reasonably formalized as relationships between individuals or interactions between objects. Analyzing such structures consequently allows to uncover and describe the phenomena that affect these systems. Network-structured data arise in various fields, for example, social and political sciences, economics, biology, and neurosciences, to name but a few. In this regard, a connectivity pattern between entities might describe friendships among members of a social group (Eagle et al., 2009), the trading between nations (Bhattacharya et al., 2008), interactions of proteins (Schwikowski et al., 2000), or the functional coactivation within the human brain (Bassett et al., 2018, Crossley et al., 2013).

In many situations, uncovering the underlying connectivity structure is not the only concern but also the comparison of akin networks and the exploration of potential differences. For instance, this might be of interest in the context of brain coactivation, which we consider here as motivating example. Recently, a lot of work has been going on investigating how the functional connectivity in the brain differs when people are affected by cognitive disorders like Alzheimer's disease or autism spectrum disorder (Song et al., 2019, Subbaraju et al., 2017, Pascual-Belda et al., 2018). Two such functional coactivation networks—resulting from averaging over the measurements within two different subject groups, respectively—are illustrated in Figure 1. One inquiry which is then particularly relevant to the research question posed is whether a significant difference can be observed in these brain processes. More generally, this can be phrased as a hypothesis test on structural equivalence of two networks as exemplarily raised for the situation in Figure 1.

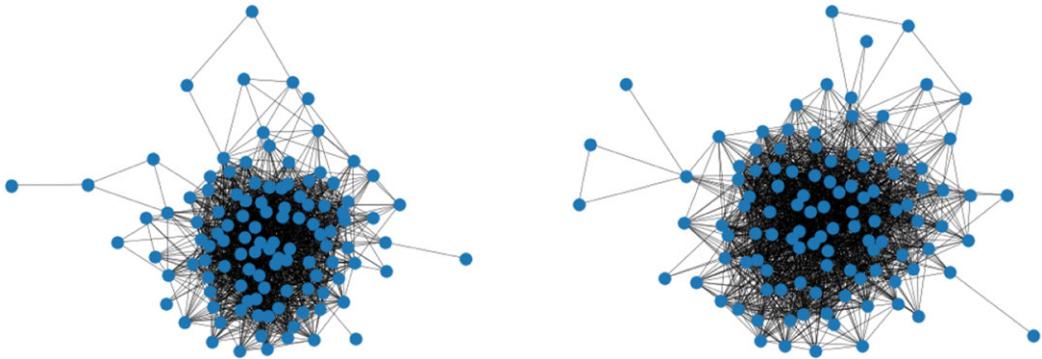


Figure 1. Functional coactivation networks of the human brain. The illustrated connectivity patterns result from averaging over multiple measurements for subjects with autism spectrum disorder (left) and typical development (right). Do these networks reveal a significant structural difference?.

To this end, we pursue constructing a statistical approach for network comparison that allows for a conclusion about significance. More precisely, we aim to test whether two networks can be considered as independent samples drawn from the same probability distribution. This is apparently in itself a technically difficult question since the two networks can have different sizes, and hence the two distributions need to be somehow different. Therefore, it is crucial that the applied distributional framework constitutes a rather universal probability measure. In fact, this is not a trivial property, and many network models entail conceptual issues that impede a direct comparison. For example, in the Exponential Random Graph Model (Robins et al., 2007), a concrete model parameterization has different implications for different network sizes, making corresponding coefficient estimates hardly comparable. Hence, for the given research question, statistical models seem advantageous where the specification of size and edge probabilities can be explicitly disentangled.

For such a comparative analysis of networks, we will demonstrate that Graphon Models (Lovász and Szegedy, 2006, Diaconis and Janson, 2008, Sischka and Kauermann, 2022) are a very useful tool. First, the graphon model is very flexible and able to capture complex network structures. Secondly, the graphon itself can be interpreted as local density or intensity function on networks of nonparametric fashion. Lastly, the model's design fulfills the above requirement of decoupling the network's structure and size, allowing for modeling multiple networks simultaneously (Navarro and Segarra, 2022). Hence, the graphon framework appears as a natural choice for comparison purposes.

The rest of the paper is organized as follows. In Section 2, we start with reviewing methods from the network comparison literature. A formalization of the test problem is elaborated and discussed in Section 3. Section 4 then deals with the joint graphon estimation, which is essential for the testing procedure to be applicable. Based on the joint graphon model, in Section 5, we develop a network comparison strategy for testing on equivalence of the underlying structures. The general applicability of the complete approach is demonstrated in Section 6, where we consider its performance on simulated and real-world data. The discussion and conclusion in Section 7 complete the paper.

2. Concepts for network comparison

When it comes to the comparison of networks, as a first step, one needs to decide on a concept for specifying network structures. In general, the various approaches can be broadly distinguished according to whether they rely on a descriptive or a model-based structural framework. Survey

articles in this field are given by Soundarajan et al., (2014), Yaveroğlu et al., (2015), Emmert-Streib et al., (2016), and Tantardini et al., (2019). A more general perspective on how complex data objects—such as adjacency matrices—might be compared is pointed out by Marron and Alonso (2014). Looking into the aforementioned compendia and examining the general literature on network comparison, however, reveals a lack of model-based comparison approaches. This can also be seen in the following, where we briefly review the different strategies that have been proposed so far.

Starting with approaches that are based on extracted network statistics, the most intuitive strategy is presumably to simply compare global characteristics such as the clustering coefficient or the average path length (Newman, 2018, pp. 364 ff., Butts 2008, p. 31). An extensive analysis on the capability of common network metrics in the setting of monitoring and identifying differences in dynamic network structures has been carried out by Flossdorf and Jentsch (2021). However, a single feature captures the overall network structure in most cases only very poorly since networks that are structured completely differently can apparently still possess a very similar value for that global statistic. Flossdorf et al., (2023) tackle this issue by using multiple metrics. Aiming in a similar direction, Wilson and Zhu (2008) consider the differences in the graph spectra, see also Gera et al., (2018). Yet, for the spectrum, it is often unclear which information in terms of local structural properties is extracted from the network and hence such approaches somehow ascribe only too little importance to the attributes of interest.

Another branch of the literature on descriptive network comparison relies on the concept of graphlets, i.e. prespecified subgraph patterns that are assumed to be sufficient for describing the present structure. Papers that, in one way or another, consider differences in the frequencies of graphlets are, among others, Pržulj et al., (2004), Pržulj (2007), Ali et al., (2014), and Faisal et al., (2017). Since the counting procedure is rather complex for larger graphlets, it is sort of a consensus to include only those that consist of no more than five nodes. Apparently, this seems somehow arbitrary and incomplete in terms of capturing all structural aspects. Moreover, Yaveroğlu et al., (2014) found high correlations among the graphlet-related statistics, including complete redundancies. On the other hand, a connection between subgraph frequencies and the concrete specification of the graphon model is exemplarily elaborated in Borgs et al., (2008), Bickel et al., (2011), and Latouche and Robin (2016).

Despite their wide dissemination, descriptive network statistics entail two general shortcomings. First, it is very difficult to assess which network segments accommodate the structural key differences. To be precise, the nodes or edges (present or absent) that contribute most to a quantified structural discrepancy can only hardly be detected. Secondly, drawing inference on the dissimilarity of two networks based on multiple network statistics is a complex inferential problem. For example, Butts (2008) aims to overcome this deficit by applying a simplistic conditional uniform graph distribution.

On the other hand, probabilistic models for network data induce distributions on structural patterns that extend to desired distributional assumptions on structural differences. As a consequence, these modeling frameworks might provide concepts for network comparison. Yet, due to a certain level of model complexity, which is required to adequately represent inherently complex network data, individual shortcomings are entailed that often prevent a direct comparison. While, for example, the Latent Distance Model (Hoff et al., 2002) does not yield any model-related key component to be compared, coefficient estimates from the exponential random graph model are not directly comparable across separated networks. The graphon model and the Stochastic Blockmodel (see Holland et al., 1983 and Snijders and Nowicki, 1997 and 2001) suffer from identifiability issues (Diaconis and Janson, 2008, Thm. 7.1) that make a comparison of corresponding individual estimates very complicated. The blockmodel's adaptivity is additionally highly dependent on the choice of the number of blocks. Onnela et al., (2012) tackle this issue by observing the networks' complete disintegration processes, which they describe by the profiles over well-known network statistics. Appropriately summarizing the differences in these profiles has been

demonstrated to provide a reasonable distance measure between networks that leads to good classification.

Overall, in view of the issues mentioned above, it seems little surprising that the literature on model-based strategies for network comparison is rather scant. More specifically, to the best of our knowledge, there exists no nonparametric statistical test on the equivalence of network structures. This shortcoming is what we aim to address in this paper. To do so, we formulate a joint smooth graphon framework that can subsequently be used for constructing an appropriate testing procedure.

3. Notation and formulation of the test problem

We consider the setting where two undirected networks of sizes $N^{(1)}$ and $N^{(2)}$ have been observed. Let $\mathbf{y}^{(g)} = [y_{ij}^{(g)}]_{i,j=1,\dots,N^{(g)}}$ for $g = 1, 2$ denote the two respective adjacency matrices, where $y_{ij}^{(g)} = 1$ if, in network g , there exists an edge between nodes i and j and $y_{ij}^{(g)} = 0$ otherwise. Since we assume the networks to be undirected, this specifically means that $y_{ij}^{(g)} = y_{ji}^{(g)}$, leading to a symmetric $\mathbf{y}^{(g)} \in \{0, 1\}^{N^{(g)} \times N^{(g)}}$. Additionally, for the diagonal elements, we set $y_{ii}^{(g)} = 0$, reflecting the absence of self-loops. In general, we consider $\mathbf{y}^{(g)}$ to be a realization of a random network $\mathbf{Y}^{(g)}$ of size $N^{(g)}$ which is subject to probability mass $\mathbb{P}(\mathbf{Y}^{(g)} = \mathbf{y}^{(g)}; N^{(g)})$. The question we aim to tackle is whether $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ are drawn from the same distribution. To suitably specify such a distribution, we rely on the smooth graphon model. The data-generating process is thereby as follows. Assume that we independently draw uniformly distributed random variables

$$U_i^{(g)} \sim \text{Uniform}[0, 1] \quad \text{for } i = 1, \dots, N^{(g)} \text{ and } g = 1, 2. \tag{1}$$

Conditional on $\mathbf{U}^{(g)} = (U_1^{(g)}, \dots, U_{N^{(g)}}^{(g)})$, we then draw the edges i.i.d. through

$$Y_{ij}^{(g)} \mid (\mathbf{U}^{(g)} = \mathbf{u}^{(g)}) \sim \text{Binomial}(1, w^{(g)}(u_i^{(g)}, u_j^{(g)})) \quad \text{for } j > i \tag{2}$$

with $\mathbf{u}^{(g)} = (u_1^{(g)}, \dots, u_{N^{(g)}}^{(g)}) \in [0, 1]^{N^{(g)}}$ and under the setting of $Y_{ij}^{(g)} \equiv Y_{ji}^{(g)}$ for $j < i$ and $Y_{ii}^{(g)} \equiv 0$. The function $w^{(g)}: [0, 1]^2 \rightarrow [0, 1]$ is called graphon (see Lovász and Szegedy, 2006 and Diaconis and Janson, 2008). We assume $w^{(g)}(\cdot, \cdot)$ to be smooth according to some Hölder or Lipschitz condition (cf. Wolfe and Olhede, 2013 or Chan and Airolidi, 2014). Relying on this data-generating process, the graphon-based probability model can be defined through

$$\mathbf{Y}^{(g)} \sim \mathbb{P}(\mathbf{Y}^{(g)} = \cdot; w^{(g)}(\cdot, \cdot), N^{(g)}). \tag{3}$$

In this distribution model, the network's size and structure are apparently dissociated, which therefore allows for a size-independent comparison of underlying structures. Hence, our goal is to develop a statistical test on the hypothesis

$$H_0: w^{(1)}(\cdot, \cdot) \equiv w^{(2)}(\cdot, \cdot). \tag{4}$$

In this context, we emphasize that data-generating process (2) is not unique because it is invariant to permutations of $w^{(g)}(\cdot, \cdot)$, as discussed in detail by Diaconis and Janson(2008, Sec. 7). Thus, the formulation of H_0 needs to be formally understood from the perspective of corresponding equivalence classes. For the concrete implementation of the test procedure, we employ a concrete representation of the two graphons as described below. Specifically, under the assumption of H_0 , we define the *joint graphon*

$$w^{\text{joint}}(u, v) := w^{(1)}(u, v) = w^{(2)}(u, v) \quad \text{for all } (u, v)^\top \in [0, 1]^2.$$

Since $w^{(1)}(\cdot, \cdot)$ and $w^{(2)}(\cdot, \cdot)$ are assumed to be smooth, this also holds for $w^{\text{joint}}(\cdot, \cdot)$. Given such joint graphon, the node position vectors $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$ referring to $w^{\text{joint}}(\cdot, \cdot)$ then provide a specific

type of network alignment. This is what we utilize for a direct comparison of $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$. Having said that, in real-world applications, neither $(\mathbf{u}^{(1)}, \mathbf{u}^{(2)})$ nor $w^{\text{joint}}(\cdot, \cdot)$ is observed. Thus, in order to achieve this alignment, we first need to formulate an appropriate estimation procedure for the joint graphon model. Note that, strictly speaking, according to formulation (3) this requires the presence of dense networks. Yet, the following modeling technique and the subsequent testing procedure are reasonable if the empirical densities are high enough and the observed networks are considered to be fixed in $N^{(g)}, g = 1, 2$.

4. EM-based joint graphon estimation

In this section, we present an iterative estimation procedure for the joint smooth graphon model under the assumption that null hypothesis (4) is true. To do so, we follow the EM-based estimation routine of Sischka and Kauermann (2022), extending it to the situation of two networks.

4.1 MCMC E-step

Starting with the E-step of our iterative algorithm, we assume the joint graphon $w^{\text{joint}}(\cdot, \cdot)$ to be known for the moment. Based on that, the latent positions of the networks can be separately determined using MCMC techniques. To be precise, we apply Gibbs sampling by formulating the full conditional distribution of $U_i^{(g)}$ through

$$f(u_i^{(g)} | u_1^{(g)}, \dots, u_{i-1}^{(g)}, u_{i+1}^{(g)}, \dots, u_{N^{(g)}}^{(g)}, \mathbf{y}^{(g)}) \propto \prod_{j \neq i} w^{\text{joint}}(u_i^{(g)}, u_j^{(g)})^{y_{ij}^{(g)}} [1 - w^{\text{joint}}(u_i^{(g)}, u_j^{(g)})]^{1-y_{ij}^{(g)}} \tag{5}$$

for all $i = 1, \dots, N^{(g)}$ and $g = 1, 2$. Details on the concrete implementation of the Gibbs sampler are given in Section A of the Appendix. The resulting MCMC sequence (after cutting the burn-in period and appropriate thinning) then reflects the joint conditional distribution $f(\mathbf{u}^{(g)} | \mathbf{y}^{(g)})$. Thus, the marginal conditional means of the latent positions, i.e. $\mathbb{E}(U_i^{(g)} | \mathbf{Y}^{(g)} = \mathbf{y}^{(g)})$ for $i = 1, \dots, N^{(g)}$, can be approximated by taking the mean over the MCMC samples, which we denote by $\bar{\mathbf{u}}^{(g)} = (\bar{u}_1^{(g)}, \dots, \bar{u}_{N^{(g)}}^{(g)})$. This posterior mean vector, however, requires further adjustment due to additional identifiability issues that cannot be coped with by the standard EM-type algorithm. To illustrate this, let model assumption (1) be more relaxed in the sense that the $U_i^{(g)}$'s are assumed to follow a continuous distribution $F^{(g)}(\cdot)$. Under this configuration, the model $(F^{(g)}(\cdot), w^{(g)}(\cdot, \cdot))$ is equivalent to any other model $(F^{(g)' }(\cdot), w^{(g)' }(\cdot, \cdot))$ constructed through

$$F^{(g)' } (u') := F^{(g)}(\varphi(u')) \quad \text{and} \quad w^{(g)' } (u', v') := w^{(g)}(\varphi(u'), \varphi(v'))$$

with $\varphi: [0, 1] \rightarrow [0, 1]$ being a strictly increasing continuous function (that is, in contrast to Diaconis and Janson, 2008, Sec. 7 not measure-preserving). Specifically, that means

$$\mathbb{P}(\mathbf{Y}^{(g)} = \cdot; F^{(g)}(\cdot), w^{(g)}(\cdot, \cdot), N^{(g)}) \equiv \mathbb{P}(\mathbf{Y}^{(g)} = \cdot; F^{(g)' }(\cdot), w^{(g)' }(\cdot, \cdot), N^{(g)})$$

for all $N^{(g)} \geq 2$. As a matter of conception, this issue cannot be solved by the EM algorithm since it aims at specifying a model that adapts optimally to the given data instead of perfectly recovering the underlying model structure. Consequently, the EM approach is not able to distinguish between the two conceptually equivalent model specifications $(F^{(g)}(\cdot), w^{(g)}(\cdot, \cdot))$ and $(F^{(g)' }(\cdot), w^{(g)' }(\cdot, \cdot))$. Nonetheless, this identifiability issue can simply be tackled by adjusting $\bar{\mathbf{u}}^{(g)}$ before estimating the graphon in the M-step. To do so, we just impose that the inferred node positions follow an ideal sample drawn from the standard uniform distribution and thus intend to sprawl out the posterior

means in an equidistant manner. More precisely, as an order-preserving adjusted estimate that is directly derived from the Gibbs sampler we utilize the “uniformized” posterior mean

$$\hat{u}_i^{(g)} = \frac{\text{rank}(\bar{u}_i^{(g)})}{N^{(g)} + 1},$$

where $\text{rank}(\bar{u}_i^{(g)})$ is the rank from the smallest to the largest of element $\bar{u}_i^{(g)}$ within $\bar{\mathbf{u}}^{(g)}$. In this context, note that the values $i/(N^{(g)} + 1)$ with $i = 1, \dots, N^{(g)}$ represent the expectations of $N^{(g)}$ ordered random variables that are independently drawn from the standard uniform distribution. As a result, with $\hat{\mathbf{u}}^{(g)} = (\hat{u}_1^{(g)}, \dots, \hat{u}_{N^{(g)}}^{(g)})$ we obtain a plausible realization of the node positions of network g . Apparently, this relies on the current joint graphon estimate $\hat{w}^{\text{joint}}(\cdot, \cdot)$, which is applied as substitute in conditional distribution (5). In the next step, we formulate the procedure for updating $\hat{w}^{\text{joint}}(\cdot, \cdot)$ given $\hat{\mathbf{u}}^{(1)}$ and $\hat{\mathbf{u}}^{(2)}$.

4.2 Spline-based M-step

For a semiparametric estimation of the joint smooth graphon, we choose a linear B-spline regression approach. To this end, we assume the joint graphon to be approximated through

$$w_{\theta}^{\text{joint}}(u, v) = \mathbf{B}(u, v) \boldsymbol{\theta} = [\mathbf{B}(u) \otimes \mathbf{B}(v)] \boldsymbol{\theta},$$

where \otimes is the Kronecker product, $\mathbf{B}(u) \in \mathbb{R}^{1 \times L}$ is a linear B-spline basis on $[0, 1]$, normalized to have a maximum value of one, and $\boldsymbol{\theta} \in \mathbb{R}^{L^2}$ is the parameter vector to be estimated. The inner B-spline knots are specified as lying equidistantly on a regular two-dimensional grid within $[0, 1]^2$, where $\boldsymbol{\theta}$ is indexed accordingly through $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1L}, \theta_{21}, \dots, \theta_{LL})^\top$. Based on this representation and given the node positions $\hat{\mathbf{u}}^{(1)}$ and $\hat{\mathbf{u}}^{(2)}$, we formulate the marginal log-likelihood over both networks as

$$\ell(\boldsymbol{\theta}) = \sum_g \sum_{\substack{ij \\ j \neq i}} \left[y_{ij}^{(g)} \log(\mathbf{B}_{ij}^{(g)} \boldsymbol{\theta}) + (1 - y_{ij}^{(g)}) \log(1 - \mathbf{B}_{ij}^{(g)} \boldsymbol{\theta}) \right], \quad (6)$$

where $\mathbf{B}_{ij}^{(g)} = \mathbf{B}(\hat{u}_i^{(g)}) \otimes \mathbf{B}(\hat{u}_j^{(g)})$. Furthermore, through standard calculations, we are able to derive the score function $\mathbf{s}(\boldsymbol{\theta})$ and the Fisher information $\mathbf{F}(\boldsymbol{\theta})$, as demonstrated in Section B of the Appendix. Fisher scoring can then be used to maximize $\ell(\boldsymbol{\theta})$. In addition, we include side constraints to ensure that $w_{\theta}^{\text{joint}}(\cdot, \cdot)$ is bounded to $[0, 1]$ and symmetric. In the linear B-spline setting, this means restricting the parameters by the conditions

$$\theta_{kl} \geq 0, \quad \theta_{kl} \leq 1, \quad \text{and} \quad \theta_{kl} - \theta_{lk} = 0$$

for all $l > k$. Apparently, all three conditions are of linear form and thus can be written in matrix format. Taken together, the Fisher scoring becomes a quadratic programming problem that can be solved using standard software (see Andersen et al., 2016 or Turlach and Weingessel, 2013).

Moreover, we intend to add penalization on the B-spline estimate. As outlined in Eilers and Marx (1996) and Ruppert et al., (2003, 2009), penalized spline estimation under the setting of a rather large spline basis yields a preferable outcome since it guarantees a functional fit that covers the data adequately but is still smooth. Thus, this approach enables to precisely capture the underlying structure while avoiding overfitting. To realize this, we add a first-order penalty, meaning that elements of $\boldsymbol{\theta}$ get penalized that are neighboring on the notional two-dimensional grid, i.e. θ_{kl} and $\theta_{(k+1)l}$ as well as θ_{kl} and $\theta_{k(l+1)}$. For the log-likelihood, the score function, and the Fisher

information, this leads to the penalized versions in the form of

$$\begin{aligned} \ell_p(\boldsymbol{\theta}, \lambda) &= \ell(\boldsymbol{\theta}) - \frac{1}{2} \lambda \boldsymbol{\theta}^\top \mathbf{P} \boldsymbol{\theta}, \quad \mathbf{s}_p(\boldsymbol{\theta}, \lambda) = \mathbf{s}(\boldsymbol{\theta}) - \lambda \mathbf{P} \boldsymbol{\theta}, \\ \text{and } \mathbf{F}_p(\boldsymbol{\theta}, \lambda) &= \mathbf{F}(\boldsymbol{\theta}) + \lambda \mathbf{P}, \end{aligned} \tag{7}$$

respectively, where \mathbf{P} is a penalization matrix of appropriate shape (see Section B of the Appendix). For an adequate choice of the penalty parameter λ in the two-dimensional spline regression, we follow Kauermann et al., (2013) and apply the corrected Akaike Information Criterion (AIC_c, see Hurvich and Tsai, 1989 and Burnham and Anderson, 2002). This is defined as

$$\text{AIC}_c(\lambda) = -2 \ell(\hat{\boldsymbol{\theta}}_p) + 2 \text{df}(\lambda) + \frac{2 \text{df}(\lambda) [\text{df}(\lambda) + 1]}{N(N - 1) - \text{df}(\lambda) - 1},$$

where $\hat{\boldsymbol{\theta}}_p$ is the corresponding penalized parameter estimate and $\text{df}(\lambda)$ specifies the degrees of freedom of the penalized B-spline function. More precisely, according to Wood (2017, pp. 211 ff.), the latter is defined through

$$\text{df}(\lambda) = \text{tr} \left\{ \mathbf{F}_p^{-1}(\hat{\boldsymbol{\theta}}_p, \lambda) \mathbf{F}(\hat{\boldsymbol{\theta}}_p) \right\}$$

with $\text{tr}\{\cdot\}$ being the trace of a matrix. A numerical optimization of the corrected AIC with respect to λ concludes the estimation of $\boldsymbol{\theta}$, resulting in the estimate $\hat{\boldsymbol{w}}^{\text{joint}}(\cdot, \cdot)$ of the current M-step. The convergence rate of $\hat{\boldsymbol{w}}^{\text{joint}}(\cdot, \cdot)$ towards $\boldsymbol{w}^{\text{joint}}(\cdot, \cdot)$ for growing $N^{(g)}, g = 1, 2$, and $L = \mathcal{O}(\min\{N^{(1)}, N^{(2)}\})$ and under the assumption of $(\hat{\boldsymbol{u}}^{(1)}, \hat{\boldsymbol{u}}^{(2)})$ being true then depends on the graphon’s smoothness, which has been postulated as a rather general assumption by the Hölder/Lipschitz condition in Section 3. Here, $\mathcal{O}(\cdot)$ refers to Landau’s small- \mathcal{O} notation.

Finally, the EM-type estimation procedure described above allows us to adequately estimate both the joint graphon $\boldsymbol{w}^{\text{joint}}(\cdot, \cdot)$ and the corresponding node positions $\boldsymbol{u}^{(1)}$ and $\boldsymbol{u}^{(2)}$ of the two networks. Specifically, given that $\boldsymbol{u}^{(1)}$ and $\boldsymbol{u}^{(2)}$ refer to the same graphon representation, this allows to formulate an appropriate test procedure based on local comparison.

5. Two-sample test on network structures

Returning to the test problem raised in Section 3, we now develop a statistical test procedure on hypothesis (4), i.e. whether $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(2)}$ are drawn from the same distribution. To do so, we utilize the network alignment resulting from the (inferred) joint smooth graphon model. More precisely, we exploit the fact that two edge variables $Y_{i_1 j_1}^{(1)}$ and $Y_{i_2 j_2}^{(2)}$ that have nearby positions—i.e. for which the distance between $(u_{i_1}^{(1)}, u_{j_1}^{(1)})^\top$ and $(u_{i_2}^{(2)}, u_{j_2}^{(2)})^\top$ is small—possess similar probabilities to form a connection. In a more formalized way, this means that, from $\|(u_{i_1}^{(1)}, u_{j_1}^{(1)})^\top - (u_{i_2}^{(2)}, u_{j_2}^{(2)})^\top\| \approx 0$, it follows that

$$\mathbb{P}(Y_{i_1 j_1}^{(1)} = 1 \mid U_{i_1}^{(1)} = u_{i_1}^{(1)}, U_{j_1}^{(1)} = u_{j_1}^{(1)}) \approx \mathbb{P}(Y_{i_2 j_2}^{(2)} = 1 \mid U_{i_2}^{(2)} = u_{i_2}^{(2)}, U_{j_2}^{(2)} = u_{j_2}^{(2)}),$$

where $\|\cdot\|$ is the Euclidean distance. Following this intuition, we divide the unit square into small segments and compare between networks the ratio of present versus absent edges occurring in these segments (see Figure 2 for an exemplary division). For that purpose, we choose a suitable $K \in \mathbb{N}$, specify a corresponding boundary sequence $a_0 = 0 < a_1 < \dots < a_K = 1$, and define the

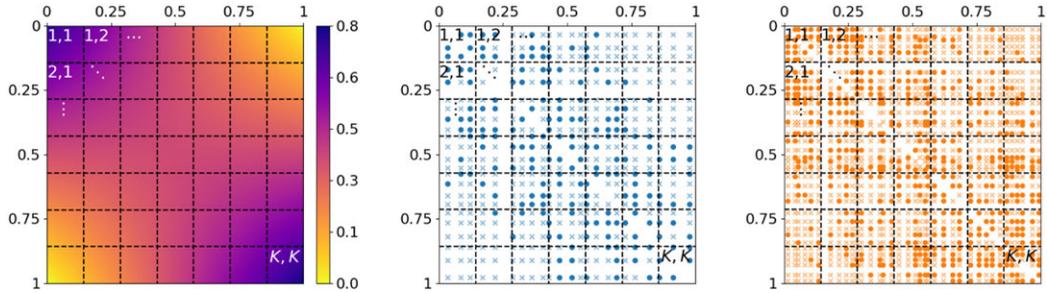


Figure 2. Dividing the unit square as domain of the graphon model into small segments for comparing network structure on a microscopic level. Left: division of $w^{\text{joint}}(\cdot, \cdot)$ into approximately piecewise-constant rectangles. Middle and right: edge positions $(u_i^{(g)}, u_j^{(g)})^\top$ of two simulated networks with respect to $w^{\text{joint}}(\cdot, \cdot)$; weakly colored crosses and intensively colored circles represent absent and present edges, respectively. The two networks can be compared by pairwise contrasting the edge proportions within the labeled rectangles.

following two quantities for $l, k = 1, \dots, K$ with $l \geq k$:

$$\begin{aligned}
 d_{kl}^{(g)} &= \sum_{\substack{i,j \\ j>i}} y_{ij}^{(g)} \mathbb{1}_{\{u_i^{(g)} \in [a_{k-1}, a_k]\}} \mathbb{1}_{\{u_j^{(g)} \in [a_{l-1}, a_l]\}} \\
 m_{kl}^{(g)} &= \sum_{\substack{i,j \\ j>i}} \mathbb{1}_{\{u_i^{(g)} \in [a_{k-1}, a_k]\}} \mathbb{1}_{\{u_j^{(g)} \in [a_{l-1}, a_l]\}}.
 \end{aligned} \tag{8}$$

This means $d_{kl}^{(g)}$ and $m_{kl}^{(g)}$ represent the number of present ($y_{ij}^{(g)} = 1$) and (a priori) possible edges of network g , respectively, within the constructed rectangle $[a_{k-1}, a_k] \times [a_{l-1}, a_l]$. The corresponding cross-network counts can be calculated by $d_{kl} = d_{kl}^{(1)} + d_{kl}^{(2)}$ and $m_{kl} = m_{kl}^{(1)} + m_{kl}^{(2)}$. Note that different network sizes are implicitly taken into account, since the sums in (8) run over the number of edges in network g , which is suppressed in the notation for simplicity of notation. Since $w^{\text{joint}}(\cdot, \cdot)$ is smooth, we further assume that the induced probability on edge variables within $[a_{k-1}, a_k] \times [a_{l-1}, a_l]$ is approximately constant. That allows for putting the observed ratios between present and absent edges in direct relation. In this light, we formulate the following contingency table to keep track of homogeneity between the networks within rectangle (k, l) :

$d_{kl}^{(1)}$	$d_{kl}^{(2)}$	d_{kl}
$m_{kl}^{(1)} - d_{kl}^{(1)}$	$m_{kl}^{(2)} - d_{kl}^{(2)}$	$m_{kl} - d_{kl}$
$m_{kl}^{(1)}$	$m_{kl}^{(2)}$	m_{kl}

Apparently, if H_0 is assumed to be true, we would expect the proportions of present edges, $d_{kl}^{(1)}/m_{kl}^{(1)}$ and $d_{kl}^{(2)}/m_{kl}^{(2)}$, to be similar. This can be assessed by contrasting the observed numbers of edges with their expectations conditional on the given margin totals, which is in line with the construction of Fisher’s exact test on 2×2 contingency tables. In this regard, the theoretical

random counterpart of $d_{kl}^{(1)}$ can be defined as

$$D_{kl}^{(1)} = \sum_{\substack{i,j \\ j>i}} Y_{ij}^{(1)} \mathbb{1}_{\{u_i^{(1)} \in [a_{k-1}, a_k]\}} \mathbb{1}_{\{u_j^{(1)} \in [a_{l-1}, a_l]\}},$$

for which under H_0 it approximately holds that

$$D_{kl}^{(1)} \mid d_{kl} \sim \text{Hyp} \left(m_{kl}, d_{kl}, m_{kl}^{(1)} \right) \quad \text{with} \quad E_{kl}^{(1)} := \mathbb{E}(D_{kl}^{(1)} \mid d_{kl}) = m_{kl}^{(1)} \frac{d_{kl}}{m_{kl}} \tag{9}$$

$$\text{and} \quad V_{kl}^{(1)} := \mathbb{V}(D_{kl}^{(1)} \mid d_{kl}) = m_{kl}^{(1)} \frac{d_{kl}}{m_{kl}} \frac{m_{kl} - d_{kl}}{m_{kl}} \frac{m_{kl} - m_{kl}^{(1)}}{m_{kl} - 1}.$$

Based on these specifications, we define our test statistic as

$$T = \sum_{\substack{k,l \\ l \geq k}} \frac{\left(D_{kl}^{(1)} - E_{kl}^{(1)} \right)^2}{V_{kl}^{(1)}} \quad \text{with realization} \quad t = \sum_{\substack{k,l \\ l \geq k}} \frac{\left(d_{kl}^{(1)} - E_{kl}^{(1)} \right)^2}{V_{kl}^{(1)}} \cdot \mathcal{U} \tag{10}$$

Note that here we only sum over the squared standardized deviations of the first network as the hypergeometric distribution is symmetric, making the same information for the second network redundant. Moreover, summands for which $V_{kl}^{(1)} = 0$ —resulting from $m_{kl}^{(1)}$, $m_{kl} - m_{kl}^{(1)}$, d_{kl} , or $m_{kl} - d_{kl}$ being zero—carry no relevant information and thus can simply be omitted from the calculation. We can now easily simulate a sample of the distribution under H_0 by drawing $D_{kl}^{(1)} \mid d_{kl}$ according to (9) and calculating T as in (10). From this, we can easily derive a critical value $c_{1-\alpha}$ or p -value by comparing the realization t of the simulated values. Moreover, asymptotic calculation is possible if $m_{kl}^{(1)}$ is large, m_{kl} and d_{kl} are large compared to $m_{kl}^{(1)}$, and d_{kl}/m_{kl} is not close to zero or one. Then $D_{kl}^{(1)}$ is approximately normally distributed and we can conclude that

$$T \stackrel{a}{\sim} \chi_{K(K+1)/2}^2. \tag{11}$$

Here, the test statistic is essentially the sum of squared (conditionally) independent random variables that approximately follow a standard normal distribution. This allows again to derive a critical value or a p -value, respectively. The choice of an appropriate K applied for these calculations is discussed in Section C of the Appendix. Note that altogether the presented test procedure follows a conception similar to the one underlying the log-rank test for time-to-event data, which also includes the construction of a conditional distribution as in (9). Regarding the latter, it seems worth noting that conditioning on the margins in the corresponding contingency table does not imply adding any assumptions on $w^{(1)}(\cdot, \cdot)$ or $w^{(2)}(\cdot, \cdot)$, or the structural relation between these two. It rather means to presume that the realization of these marginal totals carries no information about the relationship between $w^{(1)}(\cdot, \cdot)$ and $w^{(2)}(\cdot, \cdot)$, equivalently as for Fisher’s test on contingency tables (Yates 1984). We also refer to (Lang 1996) for further discussion in this line.

Apparently, when conducting the test procedure on real-world networks, we obtain the joint graphon and the corresponding alignment of the networks by applying the estimation procedure described in Section 4. In the end, this enables us to appropriately approximate test statistic (10). In this context, it is important to consider the general behavior of the joint graphon estimation under the alternative, that is, if hypothesis (4) does not hold. We stress that the intuition of the estimation procedure is to align the two networks as well as possible with respect to some suitable joint graphon model. Consequently, the expectation of T will be higher the more the true graphons $w^{(1)}(\cdot, \cdot)$ and $w^{(2)}(\cdot, \cdot)$ differ after “optimal” alignment. This clearly implies that the power of our test is higher for instances that deviate more strongly from the null hypothesis.

6. Applications

In this section, we showcase the applicability of the joint graphon estimation routine and the subsequent testing procedure. To give a comprehensive insight, this comprises both simulated and real-world networks. For an optimal estimation result and to best approximate test statistic (10), we repeat the estimation and testing procedure several times. In a modeling-oriented context, we would then typically pick the outcome with the lowest corrected AIC. However, since here the focus is on the statistical testing aspect, we choose the estimation result which leads to the highest p -value, assuming that this provides an optimal lower bound for the outcome under the (potentially existing) oracle network alignment.

Moreover, applying test statistic (10) requires to specify K and the sequence a_0, \dots, a_K . For the sake of simplicity and due to its minor impact, in the following applications we employ equidistant boundaries a_0, \dots, a_K as an intuitive approach, that is, setting $a_k = k/K$. In turn, under equidistant node positions, the consistent number of network entries per network and rectangle, $m_{kl}^{(g)}$, can be easily deduced, and we choose K such that this parameter is “not too small” for both networks. For more details, see Section C of the Appendix.

6.1 Simulation studies

6.1.1 Exemplary application to synthetic data

To demonstrate the general capability of the joint graphon estimation and the performance of the subsequent testing procedure, we consider the graphon in the top left plot of Figure 3. Its formation is inspired by and can be interpreted as a stochastic blockmodel with smooth transitions between communities. Based on this ground-truth model specification, we simulate two networks with $N^{(1)} = 200$ and $N^{(2)} = 300$ by making use of data-generating process (2). The underlying structure can then be recovered by applying the joint graphon estimation, where, for initialization, we make use of an uninformative random node positioning. This yields the graphon estimate at the top right, which apparently fully captures the structure of the ground-truth graphon. Based on the associated estimated node positions, we subsequently conduct the testing procedure on whether the underlying distributions are equivalent. To this end, we start with calculating the rectangle-wise differences according to the construction of test statistic (10). The results are depicted as a heat map at the bottom left plot of Figure 3. This reveals that the difference in the local edge density is rather low to moderate in most rectangles, whereas it is distinctly higher in a few others. Aggregating these differences yields a test statistic of 203.2 as depicted by the black solid vertical line at the bottom right. Contrasting this result with the simulated 95% quantile of the distribution of T under H_0 as the critical value (red dashed vertical line) yields no rejection. Hence, the underlying distributions of the two networks do not significantly differ with respect to a significance level of 5%. As a final remark with regard to the bottom right plot, the simulated distribution of T (black solid step function) and its theoretical approximation (blue dashed curve)—both relying on the assumption of H_0 being true—are very close to one another. Consequently, they also provide very similar critical values, namely 243.6 and 244.8, respectively. This demonstrates that asymptotic distribution (11) represents a good approximation.

6.1.2 Performance analysis under H_0

To evaluate the performance of the testing procedure in this example more profoundly, we repeat the above proceeding 400 times, with newly simulated networks in each trial (remaining with $N^{(1)} = 200$ and $N^{(2)} = 300$). Note that we run the estimation procedure always ten times (with different random node positions as varying initialization) and finally pick the highest p -value as the actual result for the given network pair. These repetitions already provide a broad insight into the method’s performance under the given setting. An even more extensive evaluation becomes

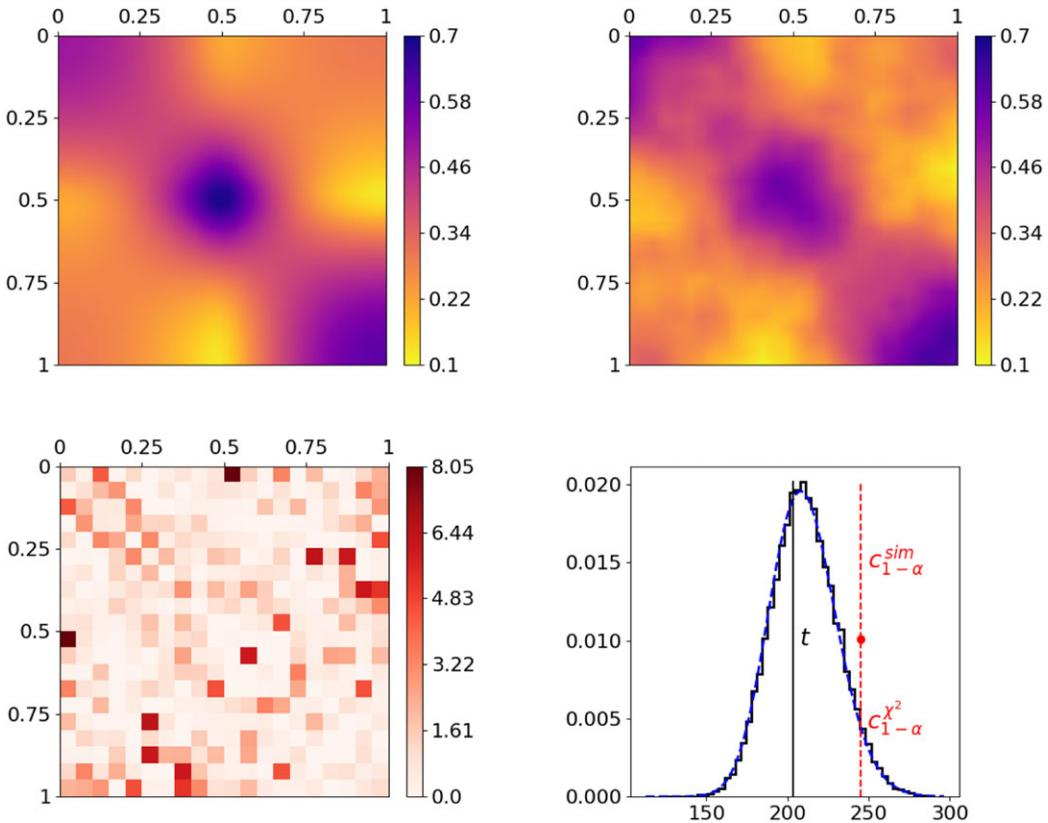


Figure 3. Joint graphon estimation for simulated networks with subsequent testing on equivalence of the underlying distribution models. The top row shows the true and the jointly estimated graphon on the left and right, respectively. The realizations of the terms of test statistic (10), representing the dissimilarities of the two networks per rectangle, are visualized at the bottom left, where $m_{kl}^{(g)} \geq 100$ for $k \neq l$ and ≥ 45 otherwise (cf. contingency table on page 16). The final result of the test statistic (black solid vertical line) as well as its distribution under H_0 are illustrated at the bottom right, where the black solid step function and the blue dashed curve depict the simulated and the asymptotic chi-squared distribution, respectively. The red dashed vertical lines (separated by red dot) visualize the critical values at a significance level of 5%, derived from the simulated (upper line) and the asymptotic distribution (lower line).

possible when, in contrast to the proceeding above, the testing procedure is performed on the basis of the oracle node positions. These positions already describe an optimal network alignment and can thus be used directly as a substitute for $(\hat{u}^{(1)}, \hat{u}^{(2)})$ that will even lead to the smallest test statistic possible. Moreover, it allows us to dramatically reduce the computational burden since it releases us from the preceding (computationally expensive) model estimation. As a consequence, we are able to increase the number of conducted tests to 10, 000. From these two repetition studies (using either $\hat{u}^{(g)}$ or $u^{(g)}$), we obtain rejection rates of 6.5% and 6.15% under the estimated and oracle node positioning, respectively. That means the test is slightly overconfident relative to the nominal significance level of 5%. The test is not exact in any way, as it is built upon the asymptotic statement (11) which explains the small bias of the test. The top row of Figure 4 shows additionally the empirical distributions of the observed p -values, illustrated as densities (left) and cumulative distribution functions (right). In accordance with the mildly inflated rejection rates, this exhibits a slight tendency to underestimate the p -value, i.e. interpreting the discrepancy as too high in distributional terms.

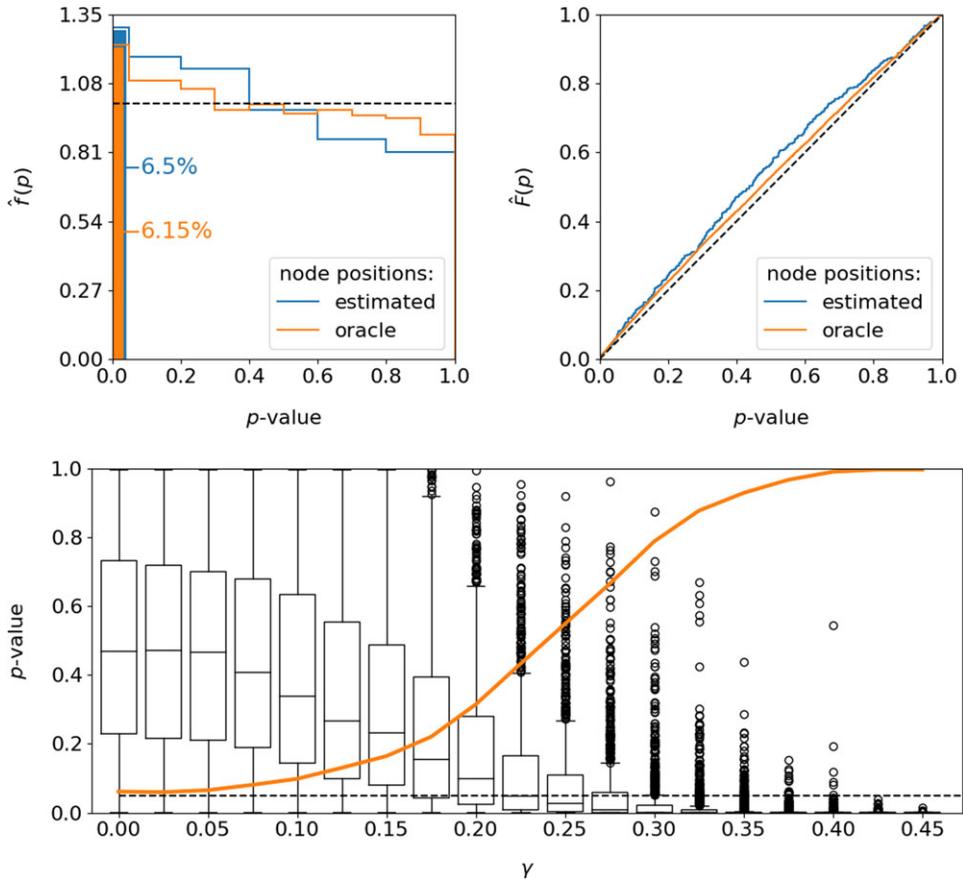


Figure 4. Performance of the testing procedure with regard to the resulting p -value; results are simulation-based. Top: empirical distribution of the p -value under H_0 , illustrated as density (left, including a depiction of rejection rates at significance level of 5%) and cumulative distribution function (right). The black dashed lines illustrate the desired distributional behavior of an optimal test. Number of repetitions for estimated / oracle node positions: 400 / 10,000. Bottom: distribution of the p -value under H_1 and the usage of oracle node positions (in box plot format); based on 1,000 repetitions each. The x -axis illustrates different settings according to formulation (12) (higher value of γ implies stronger deviation from H_0). The black dashed horizontal line represents the 5% significance level, and the orange curve illustrates the corresponding power.

6.1.3 Performance analysis under H_1

Conclusively, we are interested in evaluating the test performance under a false null hypothesis, which apparently requires formulating a suitable alternative. To this end, we “shrink” the heterogeneity within the graphon such that the present structure becomes less pronounced. The resulting graphon specification consequently tends more towards an Erdős–Rényi model, with the global density remaining unchanged. To be precise, based on $w^{(1)}(\cdot, \cdot)$, we formulate

$$w^{(2)}(u, v) := (1 - \gamma) w^{(1)}(u, v) + \gamma \bar{w}^{(1)} \tag{12}$$

with $\gamma \in [0, 1]$ and $\bar{w}^{(1)} = \int \int w^{(1)}(u, v) du dv$. Apparently, increasing the mixing parameter γ leads to a stronger deviation from H_0 . At the same time, this setting guarantees an optimal alignment of $w^{(1)}(\cdot, \cdot)$ and $w^{(2)}(\cdot, \cdot)$, meaning that there exists no rearrangement of $w^{(2)}(\cdot, \cdot)$ that is closer to $w^{(1)}(\cdot, \cdot)$ than specification (12). For this experiment, we again choose $N^{(1)} = 200$ and $N^{(2)} = 300$. Moreover, here we rely exclusively on the oracle node positions. This provides a lower bound of the power since the rejection rate can be expected to be higher when using estimated node positions instead (cf. previous analysis under H_0). The results for this setup are presented

in the bottom plot of Figure 4, where the distribution of the resulting p -value is illustrated for different settings of γ . The orange curve additionally visualizes the resulting power, i.e. the proportion of cases with $p < 0.05$. These results clearly show that the probability of detecting the false null hypothesis monotonically increases as the parameter γ gets larger, which underpins the appropriateness of our test procedure.

Overall, the obtained simulation results demonstrate that the elaborated estimation and testing procedure yields reasonable results for assessing structural differences between networks. Building upon these findings, we next want to investigate the method's performance on real-world data.

6.2 Real-world examples

6.2.1 Facebook ego networks

As a first real-world example, we consider two Facebook ego networks, with each describing the friendship behavior between all friends of one user ("ego"). These networks have been assembled by Leskovec and McAuley (2012) and are publicly available on the Stanford Large Network Dataset Collection (Leskovec and Krevl, 2014). The two ego networks consist of 333 and 168 individuals, respectively, where the ego nodes are not included. An illustration of these networks is given in the top row of Figure 5, with nodes being colored according to resulting positions. In both networks, these final estimated node positions appear to be in line with the given network structure in terms of reflecting the nodes' embedding within the network. Moreover, they seem to be aligned across networks. For example, in both networks, the reddish nodes represent the rather centric individuals, whereas the nodes from the dark blue spectrum constitute a moderately interconnected branch that is more detached from the rest of the network. However, the segment-wise differences depicted at the bottom left in Figure 5 exhibit some strong deviations. This can be clearly traced back to the blockwise division of the adjacency matrices as it results from partitioning the domain of edge positions (middle row). The aggregated differences ultimately result in a test statistic that is far from the sector of plausible values under the null hypothesis, as illustrated at the bottom right. Consequently, we can conclude that the structural behavior in the two networks differs significantly.

6.2.2 Human brain functional coactivation networks

In the second real-world application, we are concerned with differences in the human brain coactivation structure. To be precise, we compare two types of individuals, one with autism spectrum disorder (ASD) and the other with typical development (TD). In particular, we are interested in whether the functional connectivity within the brain significantly differs between these two groups (cf. the introductory example from Figure 1). For this analysis, we use resting-state functional magnetic resonance imaging data from the Autism Brain Imaging Data Exchange project (ABIDE I, 2013). More specifically, we employ preprocessed data provided by the Preprocessed Connectomes Project platform (PCP 2015). Based on these person-specific datasets, we first calculate correlations between brain regions with respect to concurrent activation over time. Aggregating the results of participants from the same clinical group and employing an appropriate threshold finally yields the network-structured data which we aim to compare. To be precise, by performing the described preprocessing, we achieve for both groups, ASD and TD, a global connectivity pattern between 116 prespecified relevant brain regions. Note that these regions are the same for both groups, which is why this could also be viewed as a comparison task under known node correspondence. However, we emphasize that in neurosciences, the transfer of competencies between brain regions is a well-known phenomenon, wherefore the general functional connectivity structure might be of greater relevance than the functional connection between specific regions. Further details on the acquisition and adequate transformation of the data are provided in Section D of the Appendix.

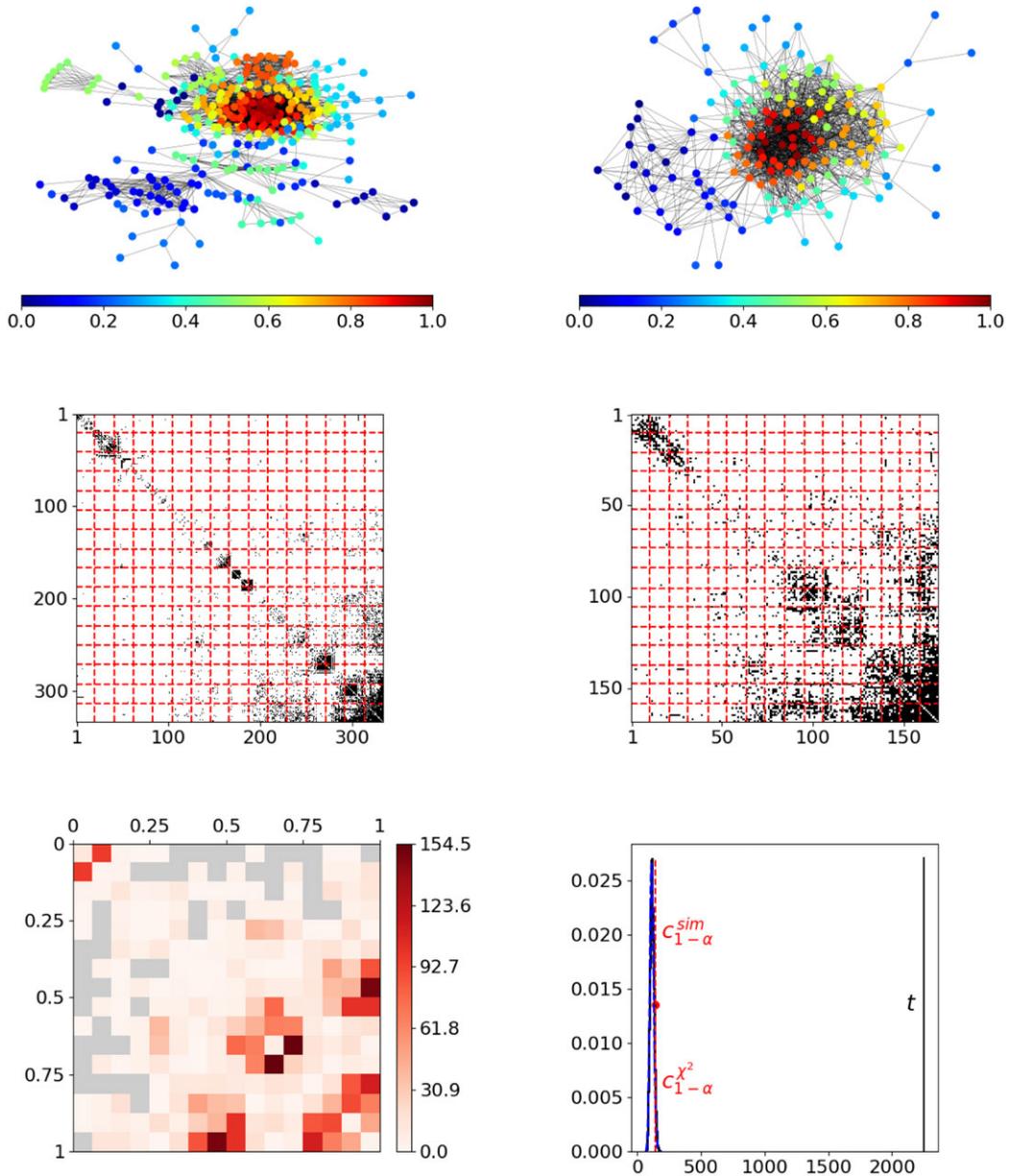


Figure 5. Comparison of two facebook ego networks. Top: illustration of networks with coloring referring to estimated node positions. Middle: ordered adjacency matrices divided into blockwise segments. Bottom left: segment-wise differences between the two networks with $m_{kl}^{(g)} \geq 100$ for $k \neq l$ and ≥ 45 otherwise (cf. contingency table on page 16); gray rectangles do not contain any observed edges ($d_{kl} = 0$) and thus provide no information. Bottom right: realization of test statistic (black solid vertical line) plus corresponding distribution under H_0 (black solid step function and blue dashed curve represent simulated and asymptotic chi-squared distribution, respectively); critical values derived from the two types of distributions are represented by the upper and the lower red dashed vertical line (separated by red dot).

For analyzing the differences in the brain coactivation structure between the two diagnostic groups, we again start with appropriately aligning the two networks. This is apparently done by employing the joint graphon estimation routine. The resulting node positions in relation to the embedding of nodes within the networks are illustrated in the top row of Figure 6. Again this

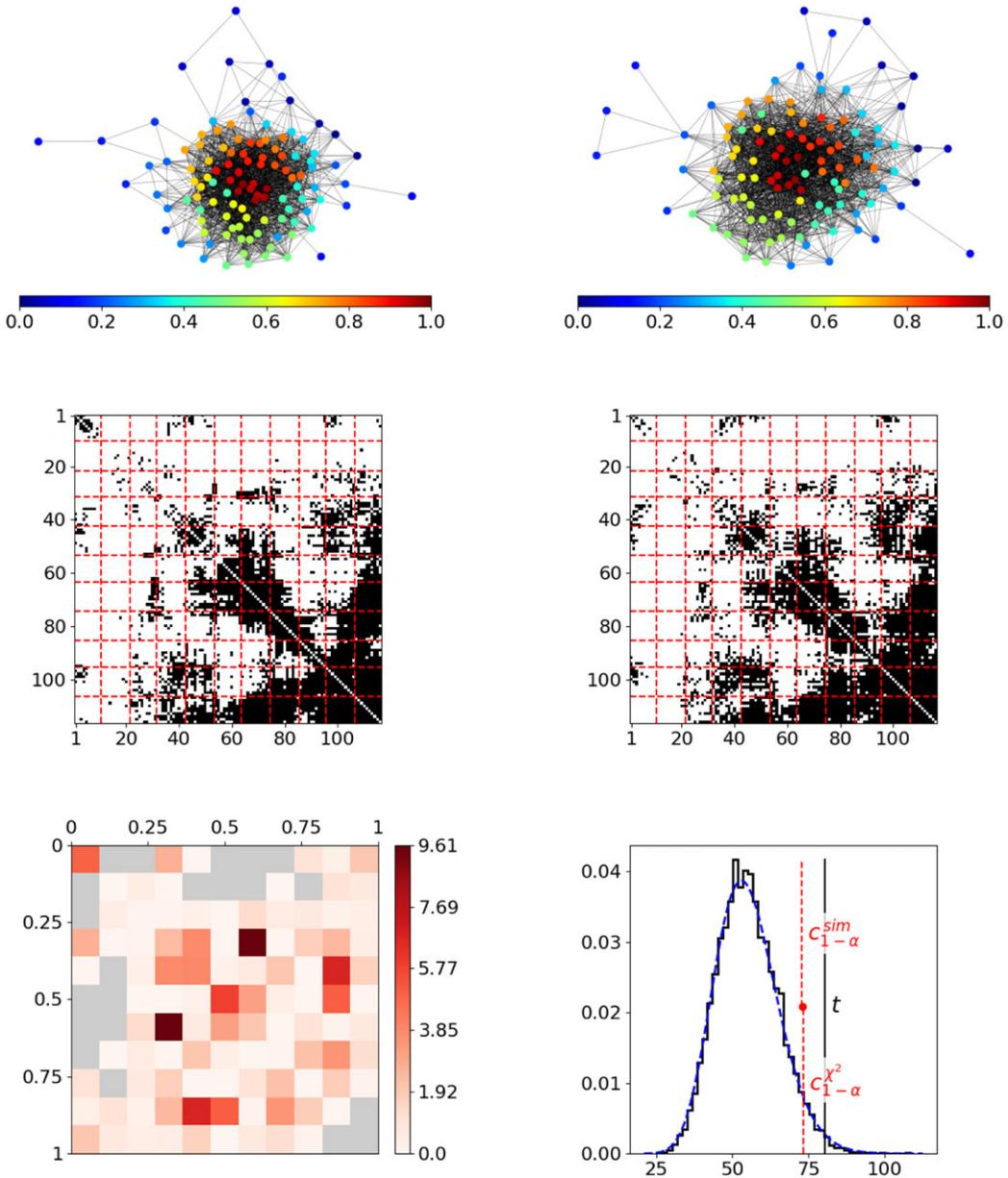


Figure 6. Comparison of functional coactivation in the human brain between groups of subjects with autism spectrum disorder and with typical development. The top row shows the networks of the ASD and the TD group on the left- and right-hand side, respectively. All illustration aspects are equivalent to the representation in Figure 5. The number of nodes per rectangle is again given by $m_{kl}^{(g)} \geq 100$ for $k \neq l$ and ≥ 45 otherwise, where N/A's in the blockwise differences result from d_{kl} or $m_{kl} - d_{kl}$ being zero.

reveals a plausible allocation of the nodes in the joint graphon model. The structural evolution can further be evaluated by consulting the correspondingly ordered adjacency matrices (see middle row), where the red dashed lines represent the blockwise division resulting from the assignment of the edge positions to the rectangles in $[0, 1]$. At first view, the formed structure looks quite similar in both matrices. Yet, on closer inspection, some blocks can be found where the density seems considerably different. This is also observed in the rectangle-wise differences

depicted in the bottom left plot. In the end, aggregating these differences leads to a p -value of 0.013 which gives a strong indication against the null hypothesis.

Given that this outcome does not support an utterly unambiguous decision of the conducted test procedure, one might additionally be interested in the nature of the inferred differences. To address this, in Section 1.2 of the Supplementary Material, we localize different behavior between the networks on the microscopic scale. Note that this can also be derived more or less directly from the joint graphon model. Moreover, for comparison reasons, we repeated the above analysis for two randomly selected disjoint subgroups of the TD group. According to the results illustrated in Figure 6 of the Supplementary Material, in this scenario, we do not observe a significant overall deviation. This further underlines the findings about the dissimilarity between the ASD and the TD group.

7. Discussion and conclusion

In the literature on network comparison, the task of drawing statistical inference appears to be an open challenge up to now. In this paper, we addressed this shortcoming by developing a non-parametric test on the equivalence of network structures. As a basis, we utilized joint smooth graphon estimation in order to arrange the networks in a meaningful way. In turn, this allowed us to compare the networks' densities at a local level. Appropriately quantifying and aggregating the differences between these densities finally enabled us to construct a nonparametric testing procedure on structural equivalence of networks. Given that graphons induce dense networks, our approach is restricted to such networks and an extension to sparse networks is questionable.

The proposed strategy of ordering networks in common relation is very similar to what is often referred to in the literature as “network alignment” (Kuchaiev et al., 2010). However, as an essential distinction to classical network alignment strategies, our method does not seek to find a node-wise one-to-one mapping. Instead, it implies a mapping of local components, meaning small fuzzy groups of nodes with similar structural roles in their respective domains.

Specifically, this also allows the networks to be of different sizes. In fact, the graphon model's property of decoupling structure and size is crucial for the proposed approach to work. In the same line, one could think of further decoupling the global density by following the approach of Bickel and Chen (2009), i.e. by introducing network-related quantities $\rho^{(g)}$ that serve as individual density coefficients. Specifically, this means modifying formulation (2) by employing $\rho^{(g)} w^{(g)}(\cdot, \cdot)$, where $\hat{\rho}^{(g)} = [N^{(g)}(N^{(g)} - 1)]^{-1} \sum_{i,j} y_{ij}^{(g)}$ could serve as an estimate that is independent of the rest of the structure. Such a framework sequentially might lead to a more balanced comparison strategy.

Beyond the applications presented in the previous section, which all refer to the situation with two networks, the method could easily be extended to cases with multiple or even a single network. For example, in the one-sample setting, to test whether a given network follows a hypothetical distribution $\mathbb{P}(Y = \cdot; w(\cdot, \cdot), N)$, we could first align the network with the theoretical graphon. That is, applying the E-step based on $w(\cdot, \cdot)$. Given this alignment, we could then turn distributional assumption (9) into a binomial distribution with the rectangle-specific mean over $w(\cdot, \cdot)$ as success probability, and, based on that, calculate the test statistic as in (10).

Besides the testing aspect, our approach could further be used to uncover relevant differences between networks on the microscopic scale. To be precise, determining the nodes or edges (present or absent) that contribute most to a quantified structural discrepancy between networks is possibly interesting in many situations. This is further elaborated in Section 1.1 of the Supplementary Material.

As a last application case, the joint graphon estimation could further be used to predict edges *between* separated networks by considering the cross-sample probabilities $w^{\text{joint}}(u_{i_1}^{(1)}, u_{i_2}^{(2)})$. This might be of interest when two (or more) networks are assumed to be samples of a larger global

network. To the best of our knowledge, this has not been pursued by any other approach so far and hence constitutes a novel perspective. As a particular hurdle in this framework, the sampling strategy that is supposed for the drawing of subnetworks needs to be taken into account in the estimation routine. As far as this is not the simple Induced Subgraph Sampling, where one selects a simple random sample of nodes within which all edges are observed, further adaptations are required. Hence, this lies beyond the scope of the paper.

Supplementary material

- (i) Description for deriving differences between networks at the microscopic level. As exemplary application, the two brain networks from Section 6.2.2 are considered.
- (ii) Replication of the test on functional coactivation networks for two subgroups of the typical-development group (confer Section 6.2.2).
- (iii) Python-package containing the code to perform the comparison methods described in the paper. The package also contains the preprocessed data of the human brain functional coactivation networks (see Section 6.2.2). (GNU zipped tar file).

Competing interests. None.

Data availability statement. All data used in the paper are publicly available. References are given in the paper.

Funding statement. This work was partially funded by the BERD@NFDI Consortium (<https://www.berd-nfdi.de>).

References

- Abide, I. (2013). Autism Brain Imaging Data Exchange I. Available online:- http://fcon_1000.projects.nitrc.org/indi/abide/ [accessed 06-21-2022]
- Ali, W., Rito, T., Reinert, G., Sun, F., & Deane, C. M. (2014). Alignment-free protein interaction network comparison. *Bioinformatics*, 30(17), i430–i437.
- Andersen, M., Dahl, J., & Vandenberghe, L. (2016). CvxOpt: Open source software for convex optimization (Python). Available online:-<https://cvxopt.org> [accessed 11-03-2022]. Version 1.2.7.
- Bassett, D. S., Zurn, P., & Gold, J. I. (2018). On the nature and use of models in network neuroscience. *Nature Reviews Neuroscience*, 19(9), 566–578.
- Bhattacharya, K., Mukherjee, G., Saramäki, J., Kaski, K., & Manna, S. S. (2008). The international trade network: weighted network analysis and modelling. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(02), P02002.
- Bickel, P. J., & Chen, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50), 21068–21073.
- Bickel, P. J., Chen, A., & Levina, E. (2011). The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5), 2280–2301.
- Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., & Vesztegombi, K. (2008). Convergent sequences of dense graphs I: subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6), 1801–1851.
- Burnham, K., & Anderson, D. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.
- Butts, C. T. (2008). Social network analysis: a methodological introduction. *Asian Journal of Social Psychology*, 11(1), 13–41.
- Chan, S. H., & Airoldi, E. M. (2014). A consistent histogram estimator for exchangeable graph models. In 31st International Conference on Machine Learning, ICML 2014, pp. 208–216.
- Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., Khundrakpam, B. S., Lewis, J. D., Li, Q., Milham, M., et al. (2013). The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, 7, 27.
- Crossley, N. A., Mechelli, A., Vértes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., & Bullmore, E. T. (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences*, 110(28), 11583–11588.
- Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D. A., Gallagher, L., Kennedy, D. P., Keown, C. L., Keyzers, C., Lainhart, J. E., Lord, C., Luna, B., Menon, V., Minshew, N. J., Monk, C. S., Mueller, S., Müller, R.-A., Nebel, M. B., Nigg, J. T., O’Hearn, K., Pelphrey, K. A., Peltier, S. J., Rudie, J. D., Sunaert, S., Thioux, M., Tyszka, J. M., Uddin, L. Q., Verhoeven, J. S., Wenderoth, N., Wiggins, J. L., Mostofsky, S. H., & Milham, M. P. (2014). The autism brain imaging

- data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6), 659–667.
- Diaconis, P., & Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni*, 28, 33–61.
- Eagle, N., Pentland, A., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36), 15274–15278.
- Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89–102.
- Emmert-Streib, F., Dehmer, M., & Shi, Y. (2016). Fifty years of graph matching, network alignment and network comparison. *Information Sciences*, 346–347, 180–197.
- Faisal, F. E., Newaz, K., Chaney, J. L., Li, J., Emrich, S. J., Clark, P. L., & Milenković, T. (2017). GRAFENE: Graphlet-based alignment-free network approach integrates 3D structural and sequence (residue order) data to improve protein structural comparison. *Scientific Reports*, 7(1), 1–15.
- Flossdorf, J., Fried, R., & Jentsch, C. (2023). Online monitoring of dynamic networks using flexible multivariate control charts. *Social Network Analysis and Mining*, 13(1), 87.
- Flossdorf, J., & Jentsch, C. (2021). Change detection in dynamic networks using network characteristics. *IEEE Transactions on Signal and Information Processing over Networks*, 7, 451–464.
- Gera, R., Alonso, L., Crawford, B., House, J., Mendez-Bermudez, J. A., Knuth, T., & Miller, R. (2018). Identifying network structure similarity using spectral graph theory. *Applied Network Science*, 3(1), 1–15.
- Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460), 1090–1098.
- Holland, P. W., Laskey, K., & Leinhardt, S. (1983). Stochastic blockmodels: first steps. *Social Networks*, 5(2), 109–137.
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297–307.
- Kauermann, G., Schellhase, C., & Ruppert, D. (2013). Flexible copula density estimation with penalized hierarchical B-splines. *Scandinavian Journal of Statistics*, 40(4), 685–705.
- Kuchaiev, O., Milenković, T., Memišević, V., Hayes, W., & Pržulj, N. (2010). Topological network alignment uncovers biological function and phylogeny. *Journal of the Royal Society Interface*, 7(50), 1341–1354.
- Lang, J. B. (1996). On the comparison of multinomial and poisson log-linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 253–266.
- Latouche, P., & Robin, S. (2016). Variational Bayes model averaging for graphon functions and motif frequencies inference in W-graph models. *Statistics and Computing*, 26(6), 1173–1185.
- Leskovec, J., & Krevl, A. (2014). SNAP Datasets: Stanford large network dataset collection. Available online: <http://snap.stanford.edu/data> [accessed 05-27-2022].
- Leskovec, J., & McAuley, J. (2012). Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems (NIPS)*, Vol. 25, pp. 539–547.
- Lovász, L., & Szegedy, B. (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6), 933–957.
- Marron, J. S., & Alonso, A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal*, 56(5), 732–753.
- Navarro, M., & Segarra, S. (2022). Joint network topology inference via a shared graphon model. *IEEE Transactions on Signal Processing*, 70, 5549–5563. arXiv preprint arXiv: 2209.08223.
- Newman, M. E. (2018). *Networks* (2nd ed.). Oxford: Oxford University Press.
- Nowicki, K., & Snijders, T. A. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455), 1077–1087.
- Olhede, S. C., & Wolfe, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 111(41), 14722–14727.
- Onnela, J.-P., Fenn, D. J., Reid, S., Porter, M. A., Mucha, P. J., Fricker, M. D., & Jones, N. S. (2012). Taxonomies of networks from community structure. *Physical Review E*, 86(3), 036104.
- Pascual-Belda, A., Díaz-Parra, A., & Moratal, D. (2018). Evaluating functional connectivity alterations in autism spectrum disorder using network-based statistics. *Diagnostics*, 8(3), 51.
- PCP. (2015). Preprocessed Connectomes Project: ABIDE. Available online: <http://preprocessed-connectomes-project.org/abide/> [accessed 06-21-2022].
- Pržulj, N. (2007). Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2), e177–e183.
- Pržulj, N., Corneil, D. G., & Jurisica, I. (2004). Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18), 3508–3515.
- Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2), 173–191.
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2009). Semiparametric regression during 2003–2007. *Electronic Journal of Statistics*, 3, 1193–1256.
- Schwikowski, B., Uetz, P., & Fields, S. (2000). A network of protein–protein interactions in yeast. *Nature Biotechnology*, 18(12), 1257–1261.

- Sischka, B., & Kauermann, G. (2022). EM-based smooth graphon estimation using MCMC and spline-based approaches. *Social Networks*, 68, 279–295.
- Sischka, B., & Kauermann, G. (2024). Stochastic block smooth graphon model. *Journal of Computational and Graphical Statistics (to appear)*.
- Snijders, T. A., & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14(1), 75–100.
- Song, Y., Epalle, T. M., & Lu, H. (2019). Characterizing and predicting autism spectrum disorder by performing resting-state functional network community pattern analysis. *Frontiers in Human Neuroscience*, 13, 203.
- Sundarajan, S., Eliassi-Rad, T., & Gallagher, B. (2014). A guide to selecting a network similarity method. In SIAM International Conference on Data Mining 2014, SDM 2014, Vol. 2, pp. 1037–1045.
- Subbaraju, V., Suresh, M. B., Sundaram, S., & Narasimhan, S. (2017). Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: a spatial filtering approach. *Medical Image Analysis*, 35, 375–389.
- Tantardini, M., Ieva, F., Tajoli, L., & Piccardi, C. (2019). Comparing methods for comparing networks. *Scientific Reports*, 9(1), 1–19.
- Turlach, B. A., & Weingessel, A. (2013). quadprog: Functions to solve quadratic programming problems (R). Available online: <https://rdrr.io/cran/quadprog/> [accesses 25.04.2025]. Version 1.5-8.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1), 273–289.
- Wilson, R. C., & Zhu, P. (2008). A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, 41(9), 2833–2841.
- Wolfe, P. J., & Olhede, S. C. (2013). Nonparametric graphon estimation, arXiv preprint arXiv: 1309.5936.
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*. Boca Raton: CRC Press.
- Xu, T., Yang, Z., Jiang, L., Xing, X.-X., & Zuo, X.-N. (2015). A connectome computation system for discovery science of brain. *Science Bulletin*, 60(1), 86–95.
- Yates, F. (1984). Tests of significance for 2×2 contingency tables. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 147(3), 426–449.
- Yaveroglu, Ö. N., Malod-Dognin, N., Davis, D., Levnajic, Z., Janjic, V., Karapandza, R., Stojmirovic, A., & Pržulj, N. (2014). Revealing the hidden language of complex networks. *Scientific Reports*, 4(1), 1–9.
- Yaveroglu, Ö. N., Milenković, T., & Pržulj, N. (2015). Proper evaluation of alignment-free network comparison methods. *Bioinformatics*, 31(16), 2697–2704.

Appendix

A. Implementation of the Gibbs sampler

In an iterative joint graphon estimation procedure, the joint posterior distribution of the node positions given $w^{\text{joint}}(\cdot, \cdot)$ and the networks $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ can be simulated by constructing a Gibbs sampler. We stress that the node positions are independent between networks and thus the Gibbs sampling procedure can be conducted for each network separately. The MCMC framework is then build upon full conditional distribution (5) and can be formulated as follows. For the successive updating procedure, we consider $\mathbf{u}^{(g), \langle t \rangle} = (u_1^{(g), \langle t \rangle}, \dots, u_{N^{(g)}}^{(g), \langle t \rangle})$ to be the current state of the Markov chain. In the $(t + 1)$ -th step, component i is then updated according to (5), where all other components remain unchanged, i.e. $u_j^{(g), \langle t+1 \rangle} := u_j^{(g), \langle t \rangle}$ for $j \neq i$. To do so, we propose a new position $u_i^{(g), *}$ by drawing from a normal distribution under the application of a logit link. To be precise, we first calculate

$$v_i^{(g), \langle t \rangle} = \text{logit}(u_i^{(g), \langle t \rangle}) = \log \left(\frac{u_i^{(g), \langle t \rangle}}{1 - u_i^{(g), \langle t \rangle}} \right),$$

then we add a normal term in the form of $v_i^{(g),*} = v_i^{(g), <t>} + \text{Normal}(0, \sigma_v^2)$, and finally we accomplish the retransformation through

$$u_i^{(g),*} = \text{logit}^{-1}(v_i^{(g),*}) = \frac{\exp(v_i^{(g),*})}{1 + \exp(v_i^{(g),*})}.$$

In this setting, the variance σ_v^2 should be chosen such that a balance between a wide-ranging exploration and a high acceptance rate is achieved, where the relation to the latter also strongly depends on the graphon’s smoothness. Given these formulations, the proposal density can be written as

$$\begin{aligned} q(u_i^{(g),*} | u_i^{(g), <t>}) &= \frac{\partial v_i^{(g),*}}{\partial u_i^{(g),*}} \phi(v_i^{(g),*} | v_i^{(g), <t>}) \\ &\propto \frac{1}{u_i^{(g),*} (1 - u_i^{(g),*})} \\ &\quad \cdot \exp\left(-\frac{1}{2} \frac{(\text{logit}(u_i^{(g),*}) - \text{logit}(u_i^{(g), <t>}))^2}{\sigma_v^2}\right), \end{aligned}$$

which leads to a proposal ratio of

$$\frac{q(u_i^{(g), <t>} | u_i^{(g),*})}{q(u_i^{(g),*} | u_i^{(g), <t>})} = \frac{u_i^{(g),*} (1 - u_i^{(g),*})}{u_i^{(g), <t>} (1 - u_i^{(g), <t>})}.$$

In combination with the likelihood ratio, the acceptance probability of the proposal, i.e. the probability for setting $u_i^{(g), <t+1>} := u_i^{(g),*}$, can be calculated through

$$\min \left\{ 1, \prod_{j \neq i} \left[\left(\frac{w(u_i^{(g),*}, u_j^{(g), <t>})}{w(u_i^{(g), <t>}, u_j^{(g), <t>})} \right)^{y_{ij}} \cdot \left(\frac{1 - w(u_i^{(g),*}, u_j^{(g), <t>})}{1 - w(u_i^{(g), <t>}, u_j^{(g), <t>})} \right)^{1-y_{ij}} \right] \frac{u_i^{(g),*} (1 - u_i^{(g),*})}{u_i^{(g), <t>} (1 - u_i^{(g), <t>})} \right\}.$$

In case the decision leads to a rejection of the proposal, we set $u_i^{(g), <t+1>} := u_i^{(g), <t>}$. To give an intuition about the proposal strategy and how it is affected by σ_v^2 , Figure 7 shows the proposal density $q(\cdot | u_i^{(g), <t>})$ and its standard deviation as a function of $u_i^{(g), <t>}$ for different values of σ_v . In our studies, we have observed a favorable behavior for the setting $\sigma_v^2 = 2^2$.

Applying the above updating strategy to all $i = 1, \dots, N^{(g)}$ completes one overall update. Finally, we achieve a proper Gibbs sampling routine through consecutively repeating this scheme. After cutting the burn-in period and applying an appropriate thinning, this approach yields a sample of the desired joint posterior distribution of the node positions.

B. Derivative and penalization of the B-spline function

As has been show in Section 4.2, the log-likelihood of a B-spline function can be straightforwardly extended towards the situation with multiple datasets. Given the formulation from (6), the score

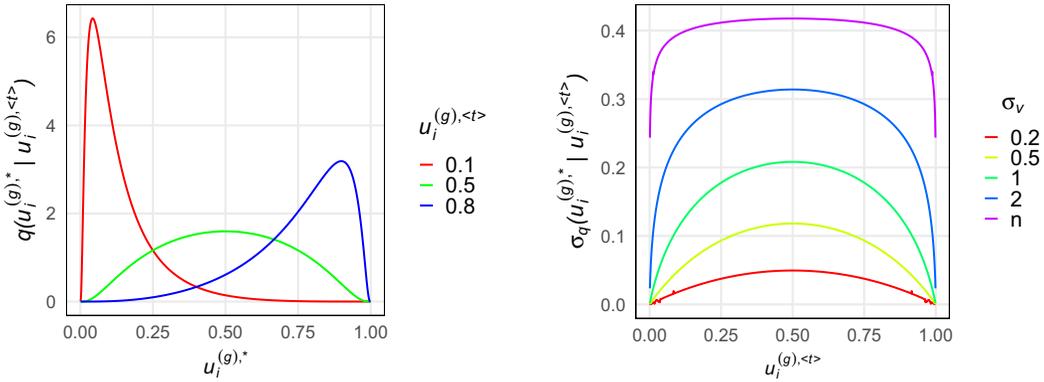


Figure 7. Left: proposal density for different current states $u_i^{(g), <t>}$ with $\sigma_v = 1$. Right: standard deviation of the proposal density against current state $u_i^{(g), <t>}$; different profiles represent the behavior for different settings of σ_v .

function can be calculated as

$$\begin{aligned}
 \mathbf{s}(\boldsymbol{\theta}) &= \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^\top \\
 &= \sum_g \sum_{\substack{i,j \\ j \neq i}} [\mathbf{B}_{ij}^{(g)}]^\top \left(\frac{\gamma_{ij}^{(g)}}{w_{\boldsymbol{\theta}}^{\text{joint}}(\hat{u}_i^{(g)}, \hat{u}_j^{(g)})} - \frac{1 - \gamma_{ij}^{(g)}}{1 - w_{\boldsymbol{\theta}}^{\text{joint}}(\hat{u}_i^{(g)}, \hat{u}_j^{(g)})} \right)
 \end{aligned}$$

where the dependence of the graphon on the parameters is expressed as subscript, i.e. $w_{\boldsymbol{\theta}}(\cdot, \cdot)$. This, in turn, leads to the Fisher information in the form of

$$\begin{aligned}
 \mathbf{F}(\boldsymbol{\theta}) &= -\mathbb{E} \left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \\
 &= \sum_g \sum_{\substack{i,j \\ j \neq i}} [\mathbf{B}_{ij}^{(g)}]^\top \mathbf{B}_{ij}^{(g)} \left[w_{\boldsymbol{\theta}}^{\text{joint}}(\hat{u}_i^{(g)}, \hat{u}_j^{(g)}) \cdot \left(1 - w_{\boldsymbol{\theta}}^{\text{joint}}(\hat{u}_i^{(g)}, \hat{u}_j^{(g)}) \right) \right]^{-1}.
 \end{aligned}$$

These results can then be used to implement the Fisher scoring procedure, where in (7) we additionally add a penalization term to guarantee a smooth estimation result. For penalizing “neighborhood” elements of the parameter vector $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1L}, \theta_{21}, \dots, \theta_{LL})^\top$, the penalization matrix can be formulated through

$$\mathbf{P} = (\mathbf{J}_L \otimes \mathbf{I}_L)^\top (\mathbf{J}_L \otimes \mathbf{I}_L) + (\mathbf{I}_L \otimes \mathbf{J}_L)^\top (\mathbf{I}_L \otimes \mathbf{J}_L),$$

where \mathbf{I}_L is the identity matrix of size L and

$$\mathbf{J}_L = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \ddots & & \ddots & \\ 0 & \cdots & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(L-1) \times L}.$$

C. Choosing the number and extent of rectangles

In order to appropriately test null hypothesis (4), in Section 5, we have developed an approach that relies on the partition of the graphon’s domain. According to formulation (8), that involves

the number of rectangles, K , as well as their concrete specification in the form of $[a_{k-1}, a_k) \times [a_{l-1}, a_l)$. In this regard, we emphasize that two aspects need to be observed. On the one hand, the joint graphon should be approximately constant within rectangles, requiring $[a_{k-1}, a_k) \times [a_{l-1}, a_l)$ to be not too expansive. On the other hand, the amount of edge variables per network falling into these blocks should be high, which needs rather broad rectangles. Thus, a trade-off between these two opposed requirements should be reached. In general, we choose K to grow more slowly than both network dimensions, e.g. scaling as $\min_g \sqrt{N^{(g)}}$. Note that choosing $K = 1$ would imply to test whether the two networks possess the same global density under the assumption of a joint Erdős–Rényi model. Having determined a suitable value for K , we then simply specify the boundaries of the rectangles through $a_k = k/K$ for $k = 0, \dots, K$. In combination with the subsequent adjustment of the latent positions as described in Section 4.1, which leads to equidistance of the estimates $\hat{u}_i^{(g)}$, a general lower bound for the amount of contained nodes per interval, $N_k^{(g)}$, can be derived. To be precise, we can formulate

$$N_k^{(g)} = \left| \left\{ i \in \{1, \dots, N^{(g)}\} : \frac{i}{N^{(g)} + 1} \in [a_{k-1}, a_k) \right\} \right| \geq \left\lfloor \frac{1}{K} (N^{(g)} + 1) \right\rfloor,$$

where $\lfloor x \rfloor$ returns the largest integer smaller than or equal to x . Given that, K could also be chosen such that, per network, a prescribed minimum amount of edge variables per rectangle ($N_k^{(g)} N_l^{(g)}$ for $l > k$ and $N_k^{(g)} (N_k^{(g)} - 1)/2$ for $l = k$) is guaranteed.

As a final remark, we emphasize that with regard to the rectangle-based test statistic, it seems natural to alternatively apply a histogram estimator (Chan and Airolidi, 2014 or Olhede and Wolfe, 2014). However, the smooth graphon estimation adapted from Sischka and Kauermann (2024) considers a global node ordering which refers not only to separated intervals but to the entire domain of $[0, 1]$. This consequently facilitates the iterative estimation procedure and thus leads to a more plausible and faster converging node positioning.

D. Acquiring and processing of brain functional activation data

The data we use for analyzing differences in the functional brain activation are originally provided by the Autism Brain Imaging Data Exchange project (ABIDE I, 2013, Di Martino et al., 2014). However, we make use of preprocessed data that are directly accessible through the Preprocessed Connectomes Project platform (PCP 2015, Craddock et al., 2013). To be precise, we here apply the Connectome Computation System pipeline (Xu et al., 2015) and the reduction to the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al., 2002). For each participant, this yields a dataset that consists of activity measurements over time for 116 prespecified brain regions (a.k.a. regions of interest). Based on these temporal activity measurements, we calculate Pearson's correlation coefficient between all pairs of brain regions which, per participant, leads to the corresponding functional connectivity matrix (Song et al., 2019, Subbaraju et al., 2017). For this analysis, we rely on the data from New York University, comprising 73 ASD patients and 98 TD subjects. For aggregating these connectivity patterns per clinical group, we apply Fisher's transformation to the pairwise correlation coefficients, calculate their mean for all pairs of brain regions, and finally retransform these means (Pascual-Belda et al., 2018). This yields for both diagnostic groups a 116×116 weighted connectivity matrix which we binarize by employing a threshold of 0.4. Based on that, the two final networks we obtain both possess a global density of about 30%. With regard to the choice of the threshold, Song et al. (2019) have found that this has only minor effects when comparing the networks.