Uncertainty Quantification in Graphon Estimation using Generalized Fiducial Inference

Yi Su, Jan Hannig, and Thomas C. M. Lee, *Senior Member, IEEE*

***Abstract*—Network data can be modeled as an exchangeable graph model (ExGM), and graphon is a two-dimensional function that generates an ExGM. The problem of graphon estimation has been popular in recent years, and several consistent estimation methods have been proposed. However, statistical inference on graphon has not been intensively studied. In this paper, we propose applying the generalized fiducial inference (GFI) methodology to the framework of graphon and perform the uncertainty quantification task. GFI is a branch of inference methods that utilizes the "switching principle" of the parameter and the data, and it seeks for a distribution estimator of the parameters without the need of a prior. We propose an easy-to-implement algorithm to generate fiducial samples of a graphon, which are then used to construct confidence sets. We establish theoretical guarantees of the GFI confidence intervals, and use synthetic graphons to demonstrate its empirical performance for finite sample size. When the labels are unknown, we extend our algorithm and discuss its asymptotic properties. We also apply the proposed method to Facebook social network data and unveil some interesting patterns.**

***Index Terms*—confidence intervals, exchangeable graph model, generalized fiducial inference, network analysis, statistical inference.**

I. Introduction

RELATIONAL datasets have become popular and more available. The exchangeable graph model (ExGM) is a tool for analyzing network data when the nodes are exchangeable [1, 2, 3, 4, 5, 6]. Behind the ExGM is a two dimensional symmetric function termed *graphon*, which defines the probability of connection between two nodes given their latent labels (can be understood as positions in the graph). Graphon is a generative model and can be viewed as a limit of finite-size graphs as the number of nodes grows to infinity [5, 7], and it can also be considered as the population object that any observed graphs are sampled from.

In general, graphon offers a long-sought and unified framework for network modeling. For example, a parametric piecewise-constant model of graphon is commonly used to approximate the community structures in social networks. In addition, graphon opens up the flexibility of nonparametric modeling, which is often instrumental in discovering interesting patterns in the corresponding network generation process [e.g., 8, 9, 10]. Applications of graphons include epidemics, graph neural networks, quantum physics, signal processing [e.g., 11, 12, 13, 14, 15, 16].

Therefore, given an observed network, which is usually realized as an adjacency matrix, it is important to estimate and make inference on the underlying graphon. Consequently, there has been a surge of interest in graphon estimation, and many of these methods can provide consistent estimators. There are mainly two categories of estimation methods. One is to use a histogram-like function with diminishing bin-size to approximate the graphon [8, 9, 17, 18, 19, 20, 21, 22, 23, 24, 25, e.g.,]. The other major category is based on variational Bayesian models and expectation maximization for SBM [e.g., 26, 27, 28, 29, 30, 31, 32].

To a much lesser extent, random graphs and graphons have also been used as tools to answer some uncertainty questions in the observed networks. For example, several bootstrapping methods were proposed to quantify the uncertainty in network summaries such as degree distributions, count features [e.g., 33, 34, 35, 36]. [37] proposed to first estimate the underlying graphon, then use the estimated graphon to re-generate graph samples in order to simulate the sampling distribution of motif densities (a normalized count of the occurrences of certain subgraphs). [38] proposed "multi-graphon" to incorporate the dynamics in a series of networks and construct confidence intervals for network summary statistics (e.g., average path length) with multiple observed networks. Another work [39] proposed a hypothesis test for testing if the estimated graphon is equal to a target graphon. This testing statistic is based on the L_2 distance in function space, and the rejection region is obtained by a Monte Carlo simulation. However, these works do not quantify the uncertainty (e.g., build confidence intervals) for the graphon itself.

In this work, we adopt the idea of generalized fiducial inference (GFI) [40] to construct block-wise confidence intervals for a graphon. GFI is a generalization of Fisher's fiducial argument [41, 42, 43], and it seeks for a "distribution estimator" of the parameter. The fiducial distribution of a parameter is similar to a Bayesian posterior distribution while it does not require a prior. The idea behind GFI is similar to that for the celebrated likelihood function — switching the roles of data and the parameter. By properly inverting the so-called data generating equation, one can generate fiducial samples from the parameter space. Analytical solutions to simple models (e.g., Gaussian distribution, linear regression) are available [40, 44]. For more complex problems, like in the graphon framework, the fiducial samples can be generated via Markov Chain Monte Carlo (MCMC). We will introduce GFI in a more detailed manner in Section III. To the best of our

Yi Su and Thomas C. M. Lee are with the Department of Statistics, University of California at Davis. E-mail: njusu@ucdavis.edu, tcmlee@ucdavis.edu. Jan Hannig is with the Department of Statistics and Operations Research, University of North Carolina at Chapel Hill. E-mail: jan.hannig@unc.edu.
The authors are most grateful to the reviewers and the associate editor for their most useful and constructive comments that led to a much improved version of the paper.

knowledge, the proposed work is the first that performs the uncertainty quantification task for graphons.

The rest of this paper is organized as follow. In Sections II and III, we provide necessary background for graphon and GFI. We formalize the graphon problem using GFI in Section 3, and derive the procedure of generating fiducial samples for graphon. In Section 4, we provide theoretical guarantees of the proposed method, followed by an illustration of its finite-sample performance when labels are known in Section VI. We also present its empirical performance with unknown labels and analyze these results from a theoretical aspect. Finally, we apply our method to Facebook social network data in Section VII. Technical details are deferred to Section IX.

II. A Review on Graphon

Let $\boldsymbol{A} \in \{0,1\}^{n\times n}$ be the adjacency matrix of a undirected simple graph with n nodes (n can be infinity). Then, according to Aldous-Hoover Theorem [45, 46], every undirected ExGM can be represented by a graphon — a symmetric measurable function $g : [0,1]^2 \to [0,1]$. Thereby, a network of size n can be generated by the following two-step sampling scheme:

$$w_i \overset{iid}{\sim} \text{Uniform}(0,1), \quad i = 1,\dots,n;$$
$$A_{ij}|w_i, w_j \overset{ind}{\sim} \text{Bernoulli}(g(w_i, w_j)), \quad i < j.$$

Since the nodes are exchangeable and so are the latent labels w_i, it is well-known that graphon is not identifiable. A widely used condition that guarantees a unique representation is the *strict monotonicity of degree condition*, and the unique representation is called the canonical graphon [39, 47]. Moreover, [48] proved that while not all graphons have a cannonical form, any strictly increasing graphon is unique.

Definition 1. (Canonical graphon) A graphon g is said to has a canonical form g^{can}, if there exists a measure preserving transformation φ, and $g^{can}(u,v) := g(\varphi(u),\varphi(v))$ satisfies the *strict monotonicity of degree condition* — the degree function $d(\cdot) = \int_0^1 g^{can}(\cdot, v)dv$ is strictly increasing.

In the graphon estimation literature, there are two categories of methods that focus on different layers of estimation. Some research groups are only interested in estimating the linkage probabilities, i.e. $\hat{g}(w_i, w_j)$. Therefore, they can skip the estimation of labels to bypass the identifiability issue. Related works include [8, 17, 23, 25, 49]. Based on some smoothness assumptions, they define a proxy distance based on the observed adjacency matrix to group nodes with similar labels together, and average over their neighbors to estimate the probability of connection. The other category of methods attempts to recover the complete graphon function, which usually relies on the strict monotonicity of degree assumption to ensure the graphon to estimate is identifiable [e.g., 19, 39]. They first sort the adjacency matrix according to the empirical node degrees, then apply a histogram estimator. [39] also discussed a hypothesis testing method based on the estimated graphon.

Since the main goal of this paper is to construct confidence intervals for graphon, we will focus on canonical graphons per Definition 1.

III. Generalized Fiducial Inference

A. Fiducial inference

The origin of fiducial inference can be traced back to 1930's when R. A. Fisher first proposed this idea in an attempt to overcome what he saw as a drawback of Bayesian inference — the use of a subjective prior distribution [41, 42, 43]. The main goal of fiducial inference is to construct a distribution for parameters of interest. The fiducial distribution can then be used for statistical inferences, for instance, confidence sets. Like Bayesian posterior distribution, the fiducial distribution is data-dependent, but the key distinction is that the fiducial approach does not demand a priori of the parameter.

Fisher showed that in simple settings, especially for one-parameter families of distributions, fiducial intervals coincide with classical confidence intervals. In multiple-parameter families of distributions, the fiducial distribution provides confidence sets whose coverage was close to the target confidence levels. However, controversies had rose because in multi-parameter settings, fiducial inference often led to procedures that were not exact in the frequentist sense. Also, there is often no unique way to define a fiducial distribution. Interested readers can find a detailed discussion on the controversies regarding fiducial inference in [50].

Because of the non-exactness and non-uniqueness of the fiducial distributions, Fisher's fiducial argument was not widely accepted among mainstream statisticians until a recent resurgence of interest that facilitated a bunch of modern modifications of the original proposal after the year 2000 [e.g., 51, 52, 53, 54, 55].

B. From FI to GFI

Generalized fiducial inference (GFI) [40, 44, 56] has been at the front line of these efforts. The development of generalized fiducial inference is essentially inspired by authors' understanding of Fisher's fiducial argument. GFI starts with describing the relationship between the data $\boldsymbol{Y}$ and the parameter $\boldsymbol{\theta}$ using a structural equation called the data generating equation (DGE):

$$\boldsymbol{Y} = \boldsymbol{G}(\boldsymbol{\theta}, \boldsymbol{U}). \tag{1}$$

Here $\boldsymbol{G}$ is a deterministic function, $\boldsymbol{\theta}$ is the parameter, and $\boldsymbol{U}$ is the random component whose distribution is completely known. For example, for a univariate normal distribution $\mathcal{N}(\mu, \sigma^2)$, one can write $\boldsymbol{Y} = \mu + \sigma\boldsymbol{U}$ with $\boldsymbol{U} \sim \mathcal{N}(0,1)$, and $\boldsymbol{\theta} = (\mu, \sigma^2)$. The key idea behind GFI is the "switching principle" between $\boldsymbol{Y}$ and $\boldsymbol{\theta}$. Given observed data $\boldsymbol{y}$, we define the following "inverse" of DGE:

$$Q_{\boldsymbol{y}}(\boldsymbol{U}) = \{\boldsymbol{\theta} : \boldsymbol{G}(\boldsymbol{\theta}, \boldsymbol{U}) = \boldsymbol{y}\}. \tag{2}$$

We will discuss this inverting procedure in detail in Section IV. The big picture is that, once given a realization $\boldsymbol{u}$ of $\boldsymbol{U}$, $Q_{\boldsymbol{y}}(\boldsymbol{u})$ is a set of $\boldsymbol{\theta}$'s in the parameter space such that $\boldsymbol{G}(\boldsymbol{\theta}, \boldsymbol{u})$ happens to equal the observed data $\boldsymbol{y}$. If the above inverse mapping exists, one can generate a fiducial sample of $\boldsymbol{\theta}$ by first generating a series of independent $\{\boldsymbol{U}_i\}_{i=1}^m$ from $\boldsymbol{U}$'s distribution and letting $\{\boldsymbol{\theta}_1 : \boldsymbol{y} = \boldsymbol{G}(\boldsymbol{\theta}_1, \boldsymbol{U}_1)\}, \dots, \{\boldsymbol{\theta}_m : \boldsymbol{y} = \boldsymbol{G}(\boldsymbol{\theta}_m, \boldsymbol{U}_m)\}$.

We need to point out that neither the existence nor the uniqueness of $Q_{\boldsymbol{y}}(\boldsymbol{U})$ can be guaranteed for all $\boldsymbol{y}$ and $\boldsymbol{u}$. If $Q_{\boldsymbol{y}}(\boldsymbol{u})$ contains more than one $\boldsymbol{\theta}$'s, one can simply select one of the several solutions using a possibly random rule — denoted as $V(\cdot)$, for instance, taking a random vertex of the polyhedron. Some guidance of such selection can be found in [56]. In fact, the uncertainty due to multiple solutions will only introduce a second-order effect on the statistical inference in many parametric problems [44]. Thus, the fiducial distribution is not sensitive to the choice of V as n grows. For the existence, one can condition on those $\boldsymbol{u}$ that does not make $Q_{\boldsymbol{y}}(\boldsymbol{u})$ an empty set. The rationale is that if the DGE is correct, i.e., $\boldsymbol{y} = G(\boldsymbol{\theta}_0, \boldsymbol{u}_0)$ for some $\boldsymbol{\theta}_0$ and $\boldsymbol{u}_0$, the values of $\boldsymbol{u}$ that result in no solution cannot be the true $\boldsymbol{u}_0$. Therefore, the generalized fiducial distribution (GFD) of $\boldsymbol{\theta}$ is the conditional distribution

$$V(Q_{\boldsymbol{y}}(\boldsymbol{U}^*))|\{Q_{\boldsymbol{y}}(\boldsymbol{U}^*) \neq \emptyset\},$$

where $\boldsymbol{U}^*$ is an independent copy of $\boldsymbol{U}$ in the DGE (1). The above conditional distribution is ill-defined for absolutely continuous random variables because $P(Q_{\boldsymbol{y}}(\boldsymbol{U}^*) \neq \emptyset) = 0$. Since this is out of the scope of this paper, interested readers can read [40, 44] for a detailed solution and theoretical results.

The theoretical properties of GFI have been better understood and the asymptotics were established in [40, 56, 57]. GFI has been applied to various applications and showed promising results, such as wavelet curve estimation [58] and ultrahigh-dimensional regression [59], amongst others. In this paper, we further extend GFI's applications to the framework of graphon.

IV. Methodology

We formally derive the GFI formulation for graphon inference problem, and develop the algorithm of simulating the fiducial distribution and constructing block-wise confidence intervals for a graphon.

A. Setup

Let $\boldsymbol{A} \in \{0,1\}^{n\times n}$ be the adjacency matrix, $A_{kl}|w_k, w_l \overset{ind}{\sim}$ Bernoulli(g_{kl}) with $g_{kl} := g(w_k, w_l)$ $(k > l)$, and $A_{kl} = A_{lk}$. In this section, all the derivation is conditional on the latent labels $\{w_i\}$.

We start with the assumptions on the underlying graphon g. To ensure its identifiability (up to a measure preserving mapping), we assume that g is a canonical graphon that satisfies the strict monotonicity degree constraint. We also need some smoothness assumption on it — g is Lipschitz continuous with constant L; see Definition 2. In fact, this assumption can be relaxed to piecewise Lipschitz as long as these pieces do not vanish too fast.

Definition 2. (Piecewise Lipschitz continuous graphon) A graphon g is piecewise Lipschitz continuous with constant $L > 0$ if there exist partitions $I : [0, x_1], [x_1, x_2], \ldots, [x_{k-1}, 1]$ and $J : [0, y_1], [y_1, y_2], \ldots, [y_{k-1}, 1]$ such that for any $\boldsymbol{w} = (w_1, w_2)$, $\boldsymbol{w}' = (w_1', w_2') \in [x_i, x_{i+1}] \cap [y_i, y_{i+1}]$,

$$|g(\boldsymbol{w}) - g(\boldsymbol{w}')| \leq L \|\boldsymbol{w} - \boldsymbol{w}'\|.$$

Second, we cut the unit square $[0,1]^2$ into $K \times K$ blocks — $b_1 = [0, \frac{1}{K}]$, $b_k = (\frac{k-1}{K}, \frac{k}{K}]$ for $k = 2, \ldots, K$, and $B_{ij} := b_i \times b_j$. Using block models to approximate a graphon has been a commonly used strategy for graphon estimation because one cannot make inference on a single observation, unless multiple graph realizations were observed [e.g., 1, 8, 19, 20, 60, 61, 62]. We assume that within each block, g is approximately constant. Thereby, $\{A_{kl}\}_{k>l}$ are independent and almost identically distributed in each block, with a common parameter

$$p_{ij} := \frac{\int_{B_{ij}} g(u,v)dudv}{\int_{B_{ij}} 1 dudv},$$

which represents the average of graphon in block B_{ij}. This assumption is reasonable when the number of blocks K grows with n, and we will justify this in Section V.

For $i, j = 1, \ldots, K$. Let

$$X_{ij} = \sum_{(w_k,w_l)\in B_{ij}} A_{kl} \text{ and } n_{ij} = |\{(w_k, w_l) : (w_k, w_l) \in B_{ij}\}|.$$

Notice that $X_{ij} = X_{ji}$. When $i = j$, due to the symmetry of $\boldsymbol{A}$, we replace X_{ii} and n_{ii} with $X_{ii}/2$ and $(n_{ii} - \sqrt{n_{ii}})/2$. Then X_{ij} follows an approximate Binomial(n_{ij}, p_{ij}).

B. Inverse mapping derivation

Let us first consider estimation on each block separately. Let $F_{n_{ij},p_{ij}}(x_{ij})$ be the CDF of Binomial(n_{ij}, p_{ij}) and denote $F^{-1}_{n,p}(u) := \inf\{x : F_{n,p}(x) \geq u\}$. Then we can write

$$X_{ij} = F^{-1}_{n_{ij},p_{ij}}(U_{ij}), \qquad 1 \leq i \leq j \leq K, \tag{3}$$

where p_{ij} is the parameter of interest, and U_{ij} are i.i.d. Uniform$(0,1)$. We will take (3) as our data generating equation (DGE). In the sequel, we put n_{ij} and p_{ij} in the subscript just to specify which Binomial CDF we use.

Next, we need to properly invert the DGE. First, since X_{ij} is discrete, (3) is equivalent to

$$F_{n_{ij},p_{ij}}(X_{ij} - 1) < U_{ij} \leq F_{n_{ij},p_{ij}}(X_{ij}). \tag{4}$$

Making use of the fact that $u = F_{n,p}(x)$ is decreasing in p, we have

$$p^L_{ij} := Q_{n_{ij},X_{ij}-1}(U_{ij}) < p_{ij} \leq Q_{n_{ij},X_{ij}}(U_{ij}) =: p^U_{ij}, \tag{5}$$

where the inverse $Q_{n,x}(u) = \{p : F_{n,p}(x) = u\}$ is computed with respect to p with x being fixed. Note that $Q_{n,x}(u)$, as a function of u is strictly decreasing. When $X_{ij} = 0$, the left hand side of (4) and (5) becomes 0. Further when $n_{ij} = X_{ij} = 0$, the right hand side of (4) and (5) is set to 1.

Thus, we have reversed the roles of data X_{ij} and the parameter p_{ij}, and we are able to generate a fiducial sample p^{L*}_{ij} of p^L_{ij} given the observed X_{ij} and a plugging in a new realization U^*_{ij} of the random component U_{ij} into $Q_{n_{ij},X_{ij}-1}(U^*_{ij})$. Similarly $p^{U*}_{ij} = Q_{n_{ij},X_{ij}}(U^*_{ij})$. We will call the distribution of p^{L*}_{ij} the lower fiducial distribution and p^{U*}_{ij} the upper fiducial distribution. Any distribution stochastically larger than the distribution of p^{L*}_{ij} and stochastically smaller than the distribution of p^{U*}_{ij} can be called GFD.

C. Linking blocks

Clearly, (5) only provides a range for p_{ij}. According to [40, 44], the selection of a specific value only brings in a second-order effect on the statistical inference. Therefore, any value of p^*_{ij} between p^L_{ij} and p^U_{ij} will not affect the asymptotic properties of this method.

A solution (5) is feasible for the graphon if it satisfies the strict monotonicity of degree assumption. Let $\boldsymbol{P} = (p_{ij}) \in [0,1]^{K\times K}$, then the parameter space for $\boldsymbol{P}$ is

$$\{\boldsymbol{P} \in [0,1]^{K\times K} : \text{ symmetric, and } \sum_{k=1}^{K} p_{ik} - \sum_{k=1}^{K} p_{i-1,k} > 0 \text{ for } i = 2,\dots,K\}. \tag{6}$$

Therefore, we need to generate $U^*_{ij},\ 1 \le i \le j \le K$ independent Uniform(0, 1) random variables such that there exists at least one $p^*_{ij},\ 1 \le i \le j \le K$ satisfying both (5) and (6). In other words, we need to generate $\boldsymbol{U}^*$ from the uniform distribution on the set

$$\mathcal{U} = \{U_{ij} \in (0,1),\ 1 \le i \le j \le K\ : \text{ both (5) and (6) are satisfied for some } \boldsymbol{P}\}.$$

While the joint distribution of $U^*_{ij}, 1 \le i \le j \le K$ can be complicated, the conditional distribution of a single U^*_{ij} given all the others is relatively straightforward; a uniform distribution on an interval described below in Section IV-D2.

D. Algorithm

Next we describe the MCMC algorithm for simulating samples from the uniform distribution on $\mathcal{U}$.

1) Initialization: We find starting values for MCMC in 3 steps.

A) First, we do an initial graphon estimation $\boldsymbol{P}^*$ by solving

$$\min_{\boldsymbol{P}^*} \sum_{i,j} \left(p^*_{ij} - \frac{X_{ij} + 1/2}{n_{ij} + 1} \right)^2 \tag{7}$$

subject to the constraint (6).

B) We compute plausible U^*_{ij} as

$$U^*_{ij} = \frac{F(n_{ij}, p^*_{ij}, X_{ij} - 1) + F(n_{ij}, p^*_{ij}, X_{ij})}{2}$$

for $i, j = 1, \dots, K$.

C) Initialize p^L_{ij}, p^U_{ij} by

$$p^L_{ij} = Q(n_{ij}, X_{ij} - 1, U^*_{ij}), \quad p^U_{ij} = Q(n_{ij}, X_{ij}, U^*_{ij})$$

for $i, j = 1, \dots, K$.

2) Gibbs sampler: We do a random scan Gibbs sampler. That means that every scan traverses all the (i, j)'s in a random order (different each time). For each fixed (i, j), we update U^*_{ij} via the following 3 steps:

A) Find temporary q^L_{ij} and q^U_{ij} by solving

$$q^L_{ij} = \min_{\boldsymbol{P}^*} p^*_{ij}, \quad q^U_{ij} = \max_{\boldsymbol{P}^*} p^*_{ij}$$

subject to constraint (6) and $p^L_{i'j'} \le p^*_{i'j'} \le p^U_{i'j'},\ \forall (i', j') \ne (i, j)$.

B) Generate new U^*_{ij} by sampling from the uniform distribution

$$U^*_{ij} \sim \text{Uniform}\left(F(n_{ij}, q^U_{ij}, X_{ij} - 1), F(n_{ij}, q^L_{ij}, X_{ij})\right).$$

C) Update

$$p^L_{ij} = Q(n_{ij}, X_{ij} - 1, U^*_{ij}), \quad p^U_{ij} = Q(n_{ij}, X_{ij}, U^*_{ij}).$$

3) GFI sample generation: Once the Gibbs scan is completed, we generate a sample graphon $\boldsymbol{P}^*$ to be saved for later (e.g., computing confidence intervals). Recall, that any $\boldsymbol{P}^*$ is valid as long as it satisfies constraint (6) and is between $\boldsymbol{p}^L$ and $\boldsymbol{p}^U$. Here we propose to select $\boldsymbol{P}^*$ using a randomization procedure to generate more diverse samples.

This is done by:

A) Flip a coin (0 or 1) with 50:50 probability, i.e. $\mathcal{C}$ is Bernoulli(1/2).

B) Generate $\tilde{\boldsymbol{P}}$ from $\tilde{P}_{ij} \sim \text{Beta}(X_{ij}+\mathcal{C}, n_{ij}+1-X_{ij}-\mathcal{C})$, $1 \le i \le j \le K$.

C) Sample $\boldsymbol{P}^*$ by solving (7) with $\tilde{P}_{ij}$ replacing $\frac{X_{ij}+1/2}{n_{ij}+1}$, subject to constraint (6) and $p^L_{ij} \le p^*_{ij} \le p^U_{ij},\ 1 \le i \le j \le K$. Other objective functions such as $\min \sum_{ij} p^*_{ij}$ could also be used.

Once we have $\boldsymbol{P}^*$'s, the empirical GFD of $\boldsymbol{P}$, we can construct $(1-\alpha)\%$ block-wise confidence intervals for p_{ij} by their $(\alpha/2)100\%$ and $(1-\alpha/2)100\%$ quantiles $(p^*_{ij,\alpha/2}, p^*_{ij,1-\alpha/2})$, $i, j = 1, \dots, K$.

V. Theoretical Results

In this section, we discuss the theoretical guarantees of our fiducial confidence interval. In the following theorems, we presume that g is Lipschitz continuous per Definition 2. The complete proofs can be found in Section IX. These theorems can also be extended to Hölder continuity, which requires a slightly different requirement of the relation between K and n. Please see Section IX-D for a remark.

Lemma 1. Under the Binomial assumption of X_{ij},

$$p^U_{ij}|X_{ij} \sim \text{Beta}(X_{ij} + 1, n_{ij} - X_{ij}),$$
$$p^L_{ij}|X_{ij} \sim \text{Beta}(X_{ij}, n_{ij} - X_{ij} + 1).$$

The proof of this lemma follows a verification of the CDF of Beta distribution.

Theorem 1. Assume that there exists some constant $\eta > 0$ such that

$$\frac{1}{K}\sum_{j=1}^{K} p_{ij} - \frac{1}{K}\sum_{j=1}^{K} p_{i-1,j} \ge \eta\frac{\sqrt{K}}{n}, \quad i = 2, \dots, K. \tag{8}$$

Define

$$R_i = \left\{ \frac{1}{K}\sum_{j=1}^{K} p^U_{ij} < \frac{1}{K}\sum_{j=1}^{K} p^L_{i-1,j} \right\},$$

and $\bigcup_{i=2}^{K} R_i$ is the event that we reject the fiducial sample as not compliant with the constraint. Then,

$$\mathbb{P}\left(\bigcup_{i=2}^{K} R_i\right) \to 0 \tag{9}$$

as $n \to \infty$, $K \to \infty$, $n/K \to \infty$.

The assumption (8) essentially says that there is no "flat part" in graphon g. It is equivalent to that the derivative of the degree function $d(\cdot)$ (see Definition 1 is bounded away from zero. Theorem 1 proves that, the constraint $\sum_{k=1}^{K} p_{ik} - \sum_{k=1}^{K} p_{i-1,k} > 0$ for $i = 2, \ldots, K$ does not affect the fiducial procedure in an asymptotic sense, as long as both K and the average number of nodes in each block grow with n.

Theorem 2. Assume graphon g is Lipschitz continuous (Definition 2). For $\forall (u_0, v_0) \in (0,1)^2$ and $g_0 = g(u_0, v_0)$, let B_{ij} be the block that contains (u_0, v_0). As $n \to \infty$, $K \to \infty$, $n/K \to \infty$, $n/K^2 \to 0$, we have

(i)
$$\frac{X_{ij}}{n_{ij}} \to g_0 \text{ a.s. (almost surely),}$$
and
$$\frac{\sqrt{n_{ij}}\left(\frac{X_{ij}}{n_{ij}} - g_0\right)}{\sqrt{g_0(1-g_0)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1);$$

(ii) Conditional on X_{ij},
$$\frac{\sqrt{n_{ij}}\left(p_{ij}^U - \frac{X_{ij}+1}{n_{ij}+1}\right)}{\sqrt{g_0(1-g_0)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \text{ a.s. in } X_{ij},$$
and
$$\frac{\sqrt{n_{ij}}\left(p_{ij}^L - \frac{X_{ij}}{n_{ij}+1}\right)}{\sqrt{g_0(1-g_0)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \text{ a.s. in } X_{ij}.$$

The first part of Theorem 2 implies that the the center of the GFD, which is coincidentally the usual point estimator, is asymptotically normal centered on the true value g_0, where randomness is the usual sample to sample variability. The second part of Theorem 2 shows that conditionally on the observed data both the upper and lower fiducial distribution are asymptotically normal with the same variance but centered on the point estimator. This means that the fiducial distribution is a good estimator of the sampling variability of the data. Consequently these theorems indicate that the GFI procedure will produce asymptotically correct confidence intervals, despite the fact that the Binomial assumption used in the derivation of the fiducial distribution is not strictly speaking correct.

VI. Numerical Experiments

In this section, we provide simulation results of the proposed GFI confidence intervals. We study both the bias of the GFD and the coverage of the GFI confidence intervals. We also compare the width of the GFI intervals with the "oracle" width $2z_{\alpha/2}\sqrt{\frac{p_{ij}(1-p_{ij})}{n_{ij}}}$, which is based on the asymptotic variance. With known labels, we explain how the simulation results support our theories in Section V. With unknown labels, we propose a Metropolis-Hastings-within-Gibbs sampler to take into account the randomness of latent labels. Experiments results are satisfying while exhibiting some boundary effects. We will explain these interesting patterns and discuss the theoretical reason behind. We use the following 4 canonical graphons in Table I, and their visualizations and degree functions are presented in Figure 1 and Figure 2 respectively.

TABLE I: Four canonical graphons (per Definition 1) used in the experiments.

g_1	$g(u,v) = \frac{u+v}{2}$
g_2	$g(u,v) = \log(1 + 0.5\max\{u,v\})$
g_3	$g(u,v) = \frac{1}{1+\exp(-10(u^2+v^2))} - 0.2$
g_4	$g(u,v) = 1 - \exp\left(-0.5(\min\{u,v\} + u^{1/2} + v^{1/2})\right)$

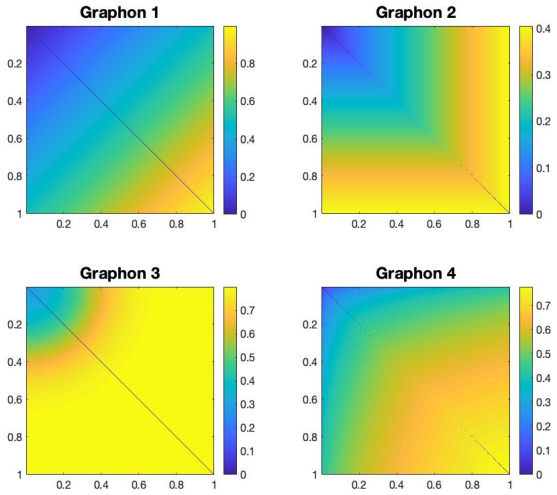

Fig. 1: Four canonical graphons used in the experiments.

A. Finite sample study with known labels

For each graphon, we randomly generate 200 graphs (i.e., realizations) with n nodes, with the same set of fixed labels. For each of these 200 replications, we use the proposed Gibbs sampler to generate an MCMC chain to obtain fiducial samples of $\boldsymbol{P}$. Then, the confidence interval for each p_{ij}, $i, j = 1, \ldots, K$, is constructed by the quantiles of the empirical fiducial distribution.

We present the results for $n = 100$, 400, 900, and we set $K = \sqrt{n}$, i.e., 10, 20, 30. To evaluate their performance, we report the average coverage of the GFI confidence intervals (CIs) across the $K \times K$ blocks, and compare with the target confidence level. We also look at the ratio of the average widths of the GFI CIs and the oracle ones. A ratio less than 1 means that the GFI CIs are narrower. The results are summarized in Table II.

We can see that the results for all 4 graphons are consistently good. When n and K are small, the coverage of the GFI confidence intervals is slightly lower than the target value, and the GFI CIs are also narrower. As n and K increase, the coverage moves closer to the target level, and the width of the GFI CI also tends to the oracle width. This observation is consistent with our theories in Section V, as they possess the same asymptotic variance. In practice, one still needs the constraint (6) to make the algorithm work properly, and it will result in more conservative intervals.

In addition, we also investigate how consistent the proposed GFI method performs over the blocks. For each position (i, j),

TABLE II: Coverage of 95% fiducial confidence intervals for finite sample sizes ($K = \sqrt{n}$).

Metric	Graphon	$n = 100$	$n = 400$	$n = 900$
GFI CI coverage	g_1	93.14%	94.52%	94.53%
	g_2	93.06%	94.67%	94.58%
	g_3	93.19%	94.39%	94.53%
	g_4	93.35%	94.63%	94.69%
Width ratio (GFI:oracle)	g_1	0.9237	0.9978	0.9981
	g_2	0.9033	0.9831	0.9905
	g_3	0.9006	0.9795	0.9856
	g_4	0.9191	0.9895	0.9930

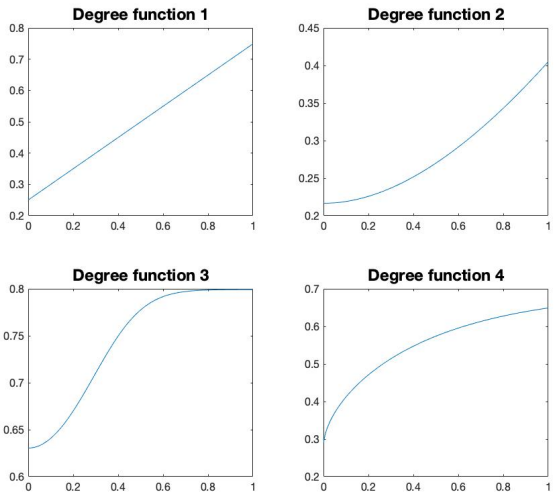

Fig. 2: The degree functions $d(u) = \int_0^1 g(u, v)dv$ of g_1 to g_4.

TABLE III: Coverage of GFI confidence intervals and the ratio of average width of GFI CIs and the oracle width ($n = 400, K = 20$).

Graphon	90%	95%	Width ratio (GFI:oracle)
g_1	98.04%	99.43%	1.45
g_2	94.24%	97.04%	1.30
g_3	94.16%	97.48%	1.34
g_4	98.34%	99.60%	1.33

we compute the bias of the fiducial distribution of p_{ij}, and the coverage of the GFI CI. The block-wise results are visualized in Figure 3. We can see that both the bias and the coverage of GFI CI perform uniformly well across all blocks.

B. Empirical performance with unknown labels

1) Metropolis-Hastings sampler: When the true labels are unknown, we need to treat the labels as part of the random component as both $\boldsymbol{X}$ and $\boldsymbol{N}$ depend on them. We need to generate $\{w_1, \ldots, w_n\}$ conditional on $\boldsymbol{U}$ such that there are feasible solutions for $\boldsymbol{G}$. Therefore, we should use a Metropolis-Hastings-within-Gibbs strategy. In each iteration, we perform the following Metropolis-Hastings sampling before Gibbs sampler in Section IV-D2.

A) Generate $w_i^* \overset{i.i.d}{\sim} \text{Uniform}(0, 1)$, $i = 1, \ldots, n$, and sort them according to the empirical node degrees of the adjacency matrix $\boldsymbol{A}$. Let σ denote such permutation, then the proposed labels are $\boldsymbol{w}^* = \{w^*_{\sigma(1)}, \ldots, w^*_{\sigma(n)}\}$.

B) Obtain temporary $\boldsymbol{X}^*$ and $\boldsymbol{N}^*$ based on the proposed labels.

C) Conditional on the current $\boldsymbol{U}^*$, compute temporary

$$\tilde{p}^L_{ij} = Q(n^*_{ij}, X^*_{ij} - 1, U^*_{ij}), \;\; \tilde{p}^U_{ij} = Q(n^*_{ij}, X^*_{ij}, U^*_{ij})$$

for $i, j = 1, \ldots, K$.

D) If there exists feasible $\boldsymbol{P}^*$ subject to constraint (6) and $\tilde{p}^L_{ij} \le p^*_{ij} \le \tilde{p}^U_{ij}$, $\forall i, j$, accept $\boldsymbol{w}^*$ and update $\boldsymbol{X}^*$, $\boldsymbol{N}^*$; otherwise, reject $\boldsymbol{w}^*$, $\boldsymbol{X}^*$ and $\boldsymbol{N}^*$.

2) Results: Here we present the results with $n = 400$ and $K = 20$. Although the average coverage of 95% GFI CI is over 97%, we observe a boundary effect as shown in Figure 4. Take g_2 for an example, from the coverage plot on the bottom we can see that the coverage on the boundary blocks is much lower than the target level, while the coverage in the middle part is somewhat higher than the target. The main cause for this phenomenon is the sorting labels step, which results in (i) an unavoidable bias on the boundary, (ii) wider GFI CIs.

From the plot of biases (left column in Figure 4) we can see that the top and left margins pf g_2 show large biases. For instance, let us consider the first segment $[0, b_1]$. When we sort nodes according to their empirical degrees, it will make X_{ij}/n_{ij}'s in those blocks near the top/left boundary smaller than they should be due to more zeros, which creates a negative bias in part (ii) of Theorem 2. Since our fiducial intervals are conditional on X_{ij}/n_{ij}, the biases are carried over to our fiducial samples. In addition, since the degree function of g_2 at $w \approx 0$ is much flatter (see Figure 1), the coverage is most worsened near the top left margins. Likewise, we observe big positive biases around the bottom right corner of g_3.

On the other hand, the GFI intervals become much wider compared to the known label case. In Table II we saw that the width of the GFI CIs is less than the oracle width when the labels were known. However, with unknown labels and the sorting step in MH sampling, the GFI intervals are on average 1.3 times wider than the oracle. From a theoretical perspective, sorting nodes according to their empirical degrees will pull observations from different distributions together and create a

mixture of Binomials in each block. Therefore, the variance of X_{ij} becomes higher and transfers to that of the generalized fiducial distribution (corresponding to Theorem 2). As a result, with relatively small bias, the coverage in the middle part is higher, i.e., $\mathbb{P}\left(p^*_{ij,\alpha/2} \leq g_0 \leq p^*_{ij,1-\alpha/2}\right) > 1-\alpha$. For the boundary, on the other hand, the lower coverage is mainly caused by the bias. By comparison, these patterns of bias and coverage are not present when labels are known.

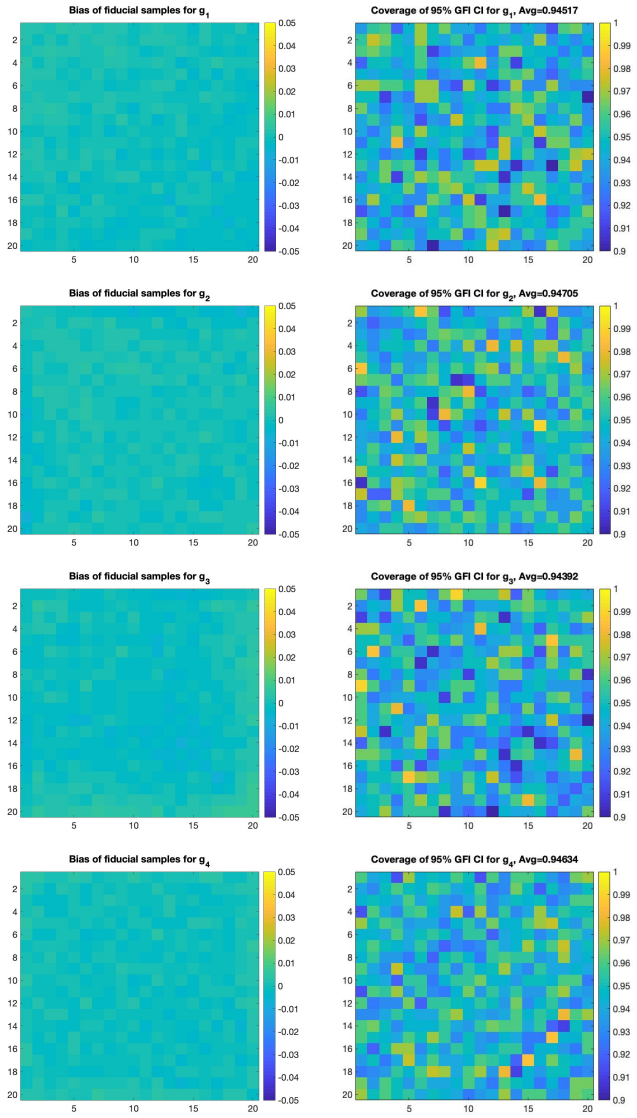

Fig. 3: Bias and coverage plots for g_1 to g_4 with known labels ($n = 400, K = 20$). Left: block-wise bias of fiducial samples. Right: block-wise GFI CI coverage.

C. Stochastic block model

In order to investigate behavior of the proposed method on a graphon that does not satisfy our assumptions, we also did an experiment with a stochastic block model (SBM) with 5 blocks.

When the labels are known (Figure 5) the performance is very good. In particular, when $K = 5$ matches the number

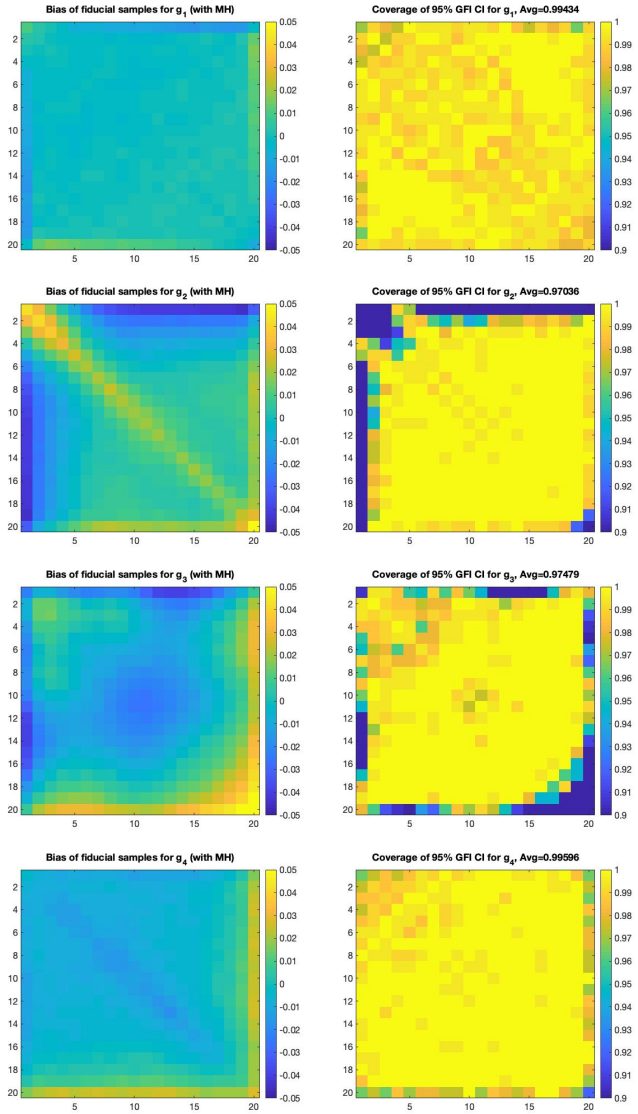

Fig. 4: Bias and coverage plots for g_1 to g_4 with unknown labels ($n = 400, K = 20$). Left: block-wise bias of fiducial samples. Right: block-wise GFI CI coverage.

of blocks the performance of GFD is perfect as expected. When $K = 20$ the performance deteriorates slightly which is expected because the graphon has flat sections while the GFI enforces strict monotonicity.

When labels and true blocks are both unknown (Figure 6), the performance of the fiducial sample varies in different areas of the graphon. This is expected behavior since SBM violates our assumption on the true graphon. In particular the SBM graphon is both discontinuous and does not have strictly increasing degree distribution. Therefore we would expect poor performance near the edges of each block where the SBM graphon is discontinuous. However, the performance of the GFI confidence intervals still performs better than a Wald's confidence interval on each block. The average coverage of the GFI 95% confidence intervals is 73%, while that for Wald's is only 57%.

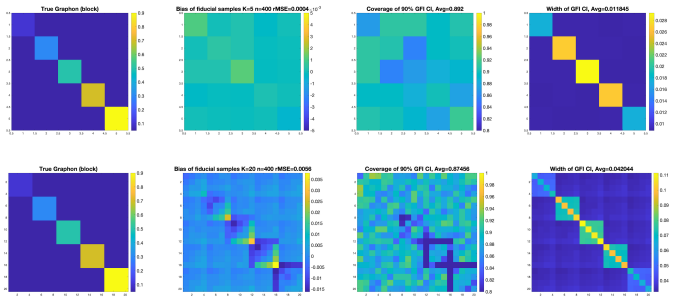

Fig. 5: Bias and coverage plots of fiducial distribution with known labels for SBM with 5 communities. Gibbs sampling is used to generate GFI samples.

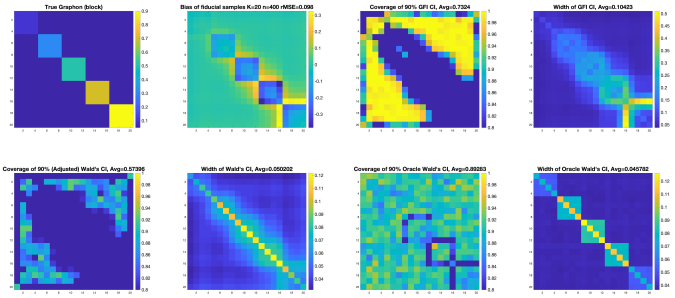

Fig. 6: Bias and coverage plots of fiducial distribution with $K = 20$ and unknown labels for SBM with 5 communities. Metropolis-Hastings-within-Gibbs sampling is used to generate GFI samples.

D. Graphon point estimation comparison

While the main contribution of this paper is to propose a new uncertainty quantification method for graphons, one can use the GFD to also find point estimators. In this section we compare the GFD estimator with several other graphon estimation methods. The competitors include sorting-and-smoothing (SAS) [19], neighborhood smoothing (NBS) [25], stochastic blockmodel approximation (SBA) [8], and universal singular value thresholding (USVT) [63]. The results are summarized in Table IV.

Among these methods, SAS algorithm requires the same strict monotonicity of degree assumption as the proposed method, while NBS, SBA and USVT do not rely on this assumption. To ensure the comparison is fair for all methods, the four graphons in Table I are used. For other methods, we average their point estimate over the $K \times K$ blocks, and compare the averaged performance based on the average graphon value on each block. The RMSE of these $K \times K$ blocks obtained from 200 replications is reported. It is worth noting that all other methods only estimate the link probabilities, i.e., $\hat{P}_{ij}$, without knowing the order of the nodes and value of the labels. Therefore, when aggregating competing methods into blocks, we are assuming their labels are known, which makes these errors over-optimistic.

VII. Real data application

In this section, we apply the proposed method to a Facebook social network [64]. The network consists of $n = 796$ Facebook users with a density of 0.0564. We report both a GFI point estimate and block-wise confidence intervals for the underlying graphon. We apply the proposed MH-within-Gibbs algorithm to generate 200 fiducial samples for $\boldsymbol{P}$. The GFI point estimate is the average of them, and the confidence intervals are constructed using the quantiles of the empirical GFD. Furthermore, we apply the proposed method to another dataset and demonstrate their different patterns.

A. Choice of K

The practical choice of K can be guided by the above theoretical results. That is, K has to be at least of order $\sqrt{n}$ (28 for this dataset). It is in fact a bias-variance trade-off — smaller K gives larger bias but smaller variance, and bigger K reduces bias at the cost of higher variance. We tried $K = 30, 50, 80$, and found that they all reveal similar patterns, although $K = 30$ shows slightly less informative structures compare to $K = 50$ and $K = 80$. Therefore, we set $K = 50$.

B. GFI estimate and confidence intervals

We visualize the estimated graphon in a 3-D plot in Figure 7. We can see that the surface is climbing due to the strict monotonicity of degree constraint. The nodes near $(1, 1)$ correspond to the people who have the most connections. We notice that there is a bump near the intersection of 0.4-0.5 and 1. This reveals an interesting phenomenon that although the nodes near 0.4-0.5 do not have high degrees, they are particularly highly connected to those "celebrities" in the network. In addition, there are some potential clusters in the flat area on the left although the estimated graphon is nearly zero. If there were more attributes on these nodes available, one could better understand their connection behaviors.

The GFI point-wise 95% confidence intervals are added to the graphon estimate in the bottom plot in Figure 7. We observe that the intervals also capture the bump and become wider as label increases due to the higher variance. On the diagonal, there are some ridges because of the fact that we only have half observations on those blocks.

C. Comparison with another network

We consider another network formed by a different group of Facebook users ($n = 962$). The GFI graphon estimate and the point-wise 95% confidence intervals are plotted in Figure 8. A substantial difference is that the second graphon decreases much faster than the first one. Also, the bump in the first network is not present in the second network.

With the proposed uncertainty quantification method, we are able to quantitatively evaluate the difference in these two graphons by a non-parametric block-wise two-sample test, using their empirical generalized fiducial distributions. For each block B_{ij}, let $p_{ij}^{(1)}$ and $p_{ij}^{(2)}$ denote the average graphon value for the two networks respectively. We carry out a non-parametric permutation test for H_0: $p_{ij}^{(1)} - p_{ij}^{(2)} = 0$ vs H_a: $p_{ij}^{(1)} - p_{ij}^{(2)} \neq 0$. We choose the significance level to be $\alpha = 0.01$. The results are plotted in Figure 9. We can see that the results also support our finding that the first graphon

TABLE IV: Graphon estimation method comparison of the 5 methods on g_1 to g_4 with $n = 400$ and $K = 20$. All values are RMSE based on 200 replications.

Graphon	Proposed GFI (unknown labels)	SAS	NBS	SBA	USVT
g_1	0.0234	0.0142	0.0140	0.0788	0.0686
g_2	0.0222	0.0232	0.0170	0.0758	0.2366
g_3	0.0214	0.0192	0.0154	0.0765	0.0335
g_4	0.0244	0.0150	0.0142	0.0819	0.0715

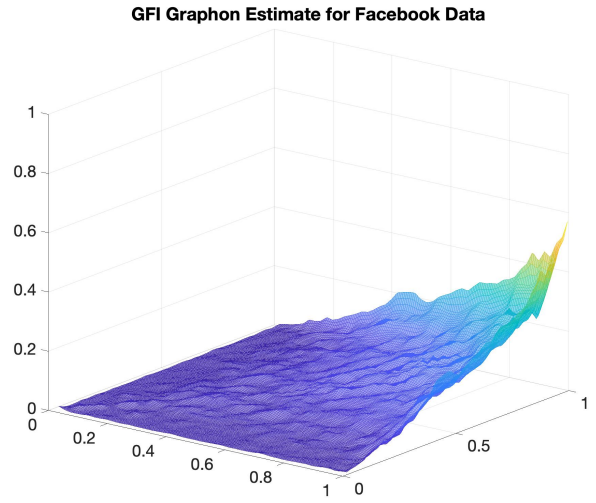

Fig. 7: GFI graphon point estimate and block-wise confidence intervals for Facebook data.

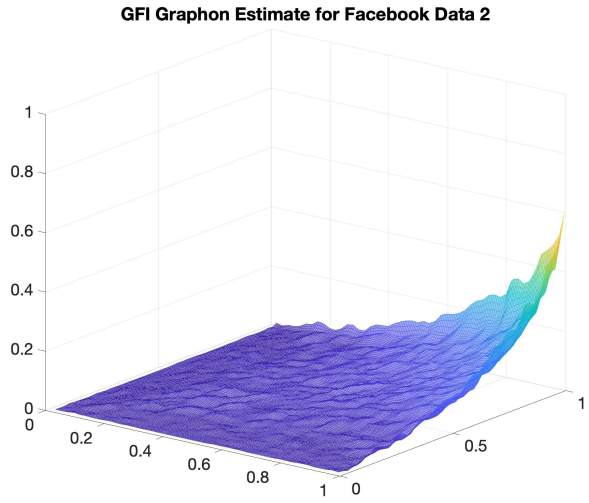

Fig. 8: GFI graphon point estimate and block-wise confidence intervals for the second Facebook network.

is significantly higher than the second one except for the area where the connections are near zero.

In addition, with the fiducial samples of $\boldsymbol{P}$, we can also construct block-wise confidence intervals for the degree distribution. We visualize the confidence bands for these two networks in Figure 10. The results also indicate that the first graphon has an overall larger degree function. However, the degree distribution does not capture the bump pattern in the first network.

VIII. Concluding Remarks

We have adapted the generalized fiducial inference in the framework of graphon, and developed an algorithm to simulate from the generalized fiducial distribution and carry out uncertainty quantification for a graphon. We proved that the block approximation and the fiducial procedure provide

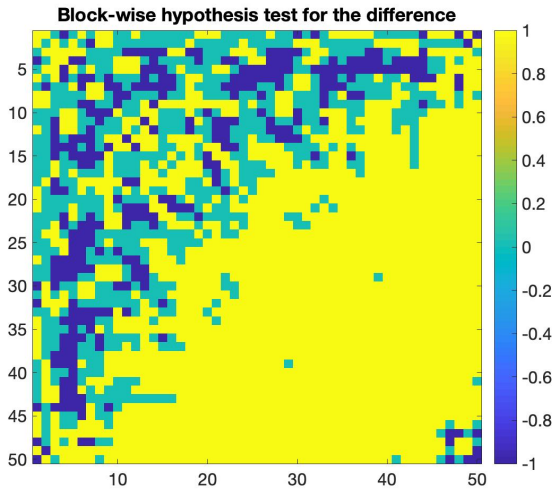

Fig. 9: Block-wise two-sample test results. The value is 0 if the test is not statistically significant at $\alpha = 0.01$. When the test is significant, the value is set to 1 if $p_{ij}^{(1)} > p_{ij}^{(2)}$, and -1 if $p_{ij}^{(1)} < p_{ij}^{(2)}$.

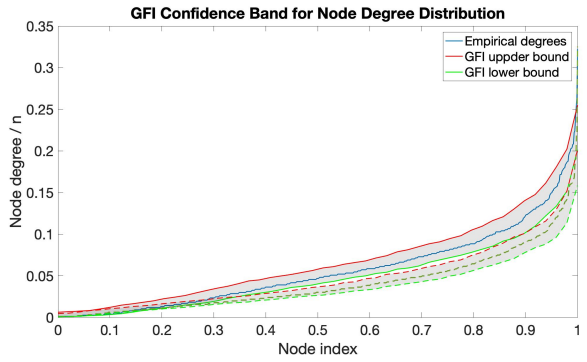

Fig. 10: Confidence bands for degree distribution. Node degrees are normalized to $[0, 1]$. Solid lines: network 1; dashed lines: network 2.

asymptotically true confidence intervals and used numerical experiments to demonstrate it. Although the circumstance of unknown labels remains an open challenge, the empirical performance of the GFI is promising despite some edge effects. It is naturally our next step to extend our methodology to mitigate the boundary issues by better estimating the latent labels. We applied the proposed method to two Facebook networks and reveal some interesting patterns. In particular, we not only visualized their differences, but also carried out a block-wise two-sample test. This work opens up the door to more extensive studies of statistical inference problems for network modeling such as simultaneous confidence intervals.

IX. Technical Details

A. Proof of Lemma 1

First of all, if $X_{ij} = 0$, $p_{ij}^L \equiv 0$; if $X_{ij} = n_{ij}$, $p_{ij}^U \equiv 1$. Otherwise, recall that $p_{ij}^L = \{p : F(n_{ij}, p, X_{ij}-1) = U_{ij}\}$ and $p_{ij}^U = \{p : F(n_{ij}, p, X_{ij}) = U_{ij}\}$ where $U_{ij} \sim \text{Uniform}(0, 1)$. The rest of the proof follows [40].

We need to prove that for fixed n and k, if $F(n, p, k)$ is the CDF of Binomial(n, p) and $U \sim \text{Uniform}(0, 1)$, then

$$p = \{p : F(n, p, k) = U\} \sim \text{Beta}(k + 1,\ n - k).$$

The CDF of p is $F_p(u) = \mathbb{P}(p \le u) = \mathbb{P}(U \ge F(n, u, k)) = 1 - F(n, u, k) = \sum_{j=k+1}^{n} \binom{n}{j} u^j (1-u)^{n-j}$.

Consider the order statistics $U_{(1)}, \dots U_{(n)}$ of Uniform(0, 1). Since $U_{(k+1)} \sim \text{Beta}(k + 1, n - k)$, its CDF is

$$F_{(k+1)}(u) = \sum_{j=k+1}^{n} \binom{n}{j} u^j (1-u)^{n-j}$$

which equals to $F_p(u)$. Therefore, $p = \{p : F(n, p, k) = U\} \sim \text{Beta}(k + 1,\ n - k)$. Thus,

$$p_{ij}^U | X_{ij} \sim \text{Beta}(X_{ij} + 1, n_{ij} - X_{ij})$$

and similarly

$$p_{ij}^L | X_{ij} \sim \text{Beta}(X_{ij}, n_{ij} - X_{ij} + 1).$$

■

B. Proof of Theorem 1

Denote $h = n/K$, and $n_i = |\{w_k \in b_i\}|$. Then $n_i \sim \text{Binomial}(n, 1/K)$ and $n_i/h \to 1$, a.s. (almost surely). Similarly, let $n_{ij} = |\{(w_k, w_l) \in B_{ij} \text{ with } k < l\}|$, then $n_{ij}/h^2 \to 1$, a.s. for $i \ne j$, and $n_{ii}/h^2 \to 1/2$, a.s.

To simplify our presentation, we write $n_i = h$, and $n_{ij} = h^2$ for $i \ne j$; $h(h-1)/2$ when $i = j$. This approximation only brings in smaller order effects when $n \to \infty$, $n/K \to \infty$ and $n/K^2 \to 0$.

Recall that

$$R_i = \left\{ \frac{1}{K} \sum_{j=1}^{K} \left(p_{ij}^U - p_{i-1,j}^L\right) \le 0 \right\},$$

and it suffices to show that

$$\mathbb{P}\left(\min_{i=2,\dots,K} \frac{1}{K} \sum_{j=1}^{K} \left(p_{ij}^U - p_{i-1,j}^L\right) \le 0 \right) \to 0.$$

We start with writing the sum in R_i as the following 5 summations.

$$\begin{aligned}
\frac{1}{K} \sum_{j=1}^{K} \left(p_{ij}^U - p_{i-1,j}^L\right) = {} & \frac{1}{K} \sum_{j=1}^{K} \left(p_{ij}^U - \frac{X_{ij}+1}{n_{ij}+1}\right) \\
& + \frac{1}{K} \sum_{j=1}^{K} \left(\frac{X_{ij}+1}{n_{ij}+1} - p_{ij}\right) \\
& + \frac{1}{K} \sum_{j=1}^{K} \left(p_{ij} - p_{i-1,j}\right) \\
& + \frac{1}{K} \sum_{j=1}^{K} \left(p_{i-1,j} - \frac{X_{i-1,j}}{n_{i-1,j}+1}\right) \\
& + \frac{1}{K} \sum_{j=1}^{K} \left(\frac{X_{i-1,j}}{n_{i-1,j}+1} - p_{i-1,j}^L\right) \\
= {} & \mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3 + \mathcal{S}_4 + \mathcal{S}_5.
\end{aligned}$$

(i) First, $\mathcal{S}_3 \geq \eta\frac{\sqrt{K}}{n}$ for all i by assumption.

(ii) For $\mathcal{S}_2$ and $\mathcal{S}_4$, we have

$$\begin{aligned}|\mathcal{S}_2| &= \frac{1}{K}\left|\sum_{j=1}^{K}\left(\frac{X_{ij}+1}{n_{ij}+1} - p_{ij}\right)\right| \\ &\leq \frac{2}{h^2} + \frac{1}{K}\left|\sum_{j=1}^{K}\left(\frac{X_{ij}}{n_{ij}} - p_{ij}\right)\right| \\ &\leq \frac{2}{h^2} + \frac{2}{Kh^2}\left|\sum_{j=1}^{K}\sum_{(w_k,w_l)\in B_{ij}}(A_{kl} - p_{ij})\right|\end{aligned}$$

where $A_{kl} \overset{ind.}{\sim} \text{Bernoulli}(g_{kl})$. By Hoeffding's inequality,

$$\mathbb{P}\left(\frac{1}{Kh^2}\left|\sum_{j,k,l}(A_{kl} - p_{ij})\right| > \varepsilon/2\right) \leq 2\exp\left\{-Kh^2\varepsilon^2/2\right\}.$$

Taking a union bound over $i = 2, \ldots, K$, and setting $\varepsilon = \sqrt{\frac{2\log Kh^2}{Kh^2}}$, we have

$$\begin{aligned}&\mathbb{P}\left(\max_{i=2,\ldots,K}\frac{1}{Kh^2}\left|\sum_{j,k,l}(A_{kl} - p_{ij})\right| > \varepsilon/2\right) \\ &\leq 2K\exp\left\{-Kh^2\varepsilon^2/2\right\} = \frac{2}{h^2} \to 0\end{aligned}$$

as $h = n/K \to \infty$.

Thus,

$$\mathbb{P}\left(\max_{i=2,\ldots,K}|\mathcal{S}_2| > \sqrt{\frac{\log Kh^2}{2Kh^2}}\right) \to 0 \text{ as } n/K \to \infty.$$

Same statement holds for $\mathcal{S}_4$.

(iii) For $\mathcal{S}_1$ and $\mathcal{S}_5$, recall that $p_{ij}^U|X_{ij} \sim \text{Beta}(X_{ij}+1, n_{ij}-X_{ij})$ and $\mathbb{E}(p_{ij}^U|X_{ij}) = \frac{X_{ij}+1}{n_{ij}+1}$.

According to [65], Beta distribution p_{ij}^U is sub-Gaussian with parameter $\sigma^2 \leq \frac{1}{4(n_{ij}+2)}$. Therefore, by Hoeffding's inequality,

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{K}\left|\sum_{j=1}^{K}\left(p_{ij}^U - \frac{X_{ij}+1}{n_{ij}+1}\right)\right| > \varepsilon\middle| X_{ij}\right) \\ &\leq 2\exp\left\{-\frac{(K\varepsilon)^2}{2\sum_{j=1}^{K}\frac{1}{4(n_{ij}+2)}}\right\} \\ &\leq 2\exp\left\{-Kh^2\varepsilon^2\right\}.\end{aligned}$$

Taking a union bound over $i = 2, \ldots, K$ and letting $\varepsilon = \sqrt{\frac{\log Kh^2}{Kh^2}}$, we have

$$\begin{aligned}&\mathbb{P}\left(\max_{i=2,\ldots,K}\frac{1}{K}\left|\sum_{j=1}^{K}\left(p_{ij}^U - \frac{X_{ij}+1}{n_{ij}+1}\right)\right| > \varepsilon\right) \\ &\leq 2K\exp\left\{-Kh^2\varepsilon^2\right\} = \frac{2}{h^2} \to 0\end{aligned}$$

as $h = n/K \to \infty$.

Same statement holds for $\mathcal{S}_5$.

To sum up, for $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_4$, $\mathcal{S}_5$, the union bound $\varepsilon^* = o\left(\sqrt{1/Kh^2}\right) = o\left(K^{1/2}/n\right)$, while $\mathcal{S}_3 \geq \eta\frac{\sqrt{K}}{n}$ for all i. Therefore, we have

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=2}^{K}R_i\right) &\leq \mathbb{P}\left(\min_i\frac{1}{K}\sum_{j=1}^{K}\left(p_{ij}^U - p_{i-1,j}^L\right) \leq 0\right) \\ &\leq \mathbb{P}\left(\min_i\mathcal{S}_3 \leq \max_i|\mathcal{S}_1| + |\mathcal{S}_2| + |\mathcal{S}_4| + |\mathcal{S}_5|\right) \to 0\end{aligned}$$

as $n/K \to \infty$. ■

C. Proof of Theorem 2

For any $(u_0, v_0) \in (0,1)^2$ and $g_0 = g(u_0, v_0)$, let B_{ij} be the block that contains (u_0, v_0). Define

$$g_{B_{ij}} := \frac{1}{n_{ij}}\sum_{(w_k,w_l)\in B_{ij}} g_{kl}.$$

As in the proof of Theorem 1, $n_i \sim h = n/K$ and $n_{ij} \sim h^2$ for $i \neq j$ and $h^2/2$ for $i = j$.

(i) $\frac{X_{ij}}{n_{ij}} = \frac{1}{n_{ij}}\sum_{(w_k,w_l)\in B_{ij}}A_{kl}$ with $A_{kl} \overset{ind.}{\sim} \text{Bernoulli}(g_{kl})$. By Kolmogorov's strong law of large numbers,

$$\frac{1}{n_{ij}}\sum_{(w_k,w_l)\in B_{ij}}A_{kl} \overset{a.s.}{\to} g_{B_{ij}} \text{ as } h = n/K \to \infty.$$

Also, by the Lipschitz property of g (with Lipschitz constant L), we have

$$\left|g_{B_{ij}} - g_0\right| = \left|\frac{1}{n_{ij}}\sum_{(w_k,w_l)\in B_{ij}}(g_{kl} - g_0)\right| \leq \frac{2L}{K} \to 0$$

as $K \to \infty$.

Therefore, by continuity,

$$\frac{X_{ij}}{n_{ij}} \overset{a.s.}{\to} g_0 \text{ as } K \to \infty, n/K \to \infty.$$

Moreover, by Lindeberg central limit theorem,

$$\begin{aligned}&\frac{X_{ij} - \sum_{(w_k,w_l)\in B_{ij}}g_{kl}}{\sqrt{\sum_{(w_k,w_l)\in B_{ij}}g_{kl}(1-g_{kl})}} \\ &= \frac{\sqrt{n_{ij}}\left(\frac{X_{ij}}{n_{ij}} - g_{B_{ij}}\right)}{\sqrt{\frac{1}{n_{ij}}\sum_{(w_k,w_l)\in B_{ij}}g_{kl}(1-g_{kl})}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)\end{aligned}$$

as $h = n/K \to \infty$. By continuity of g, the denominator $\frac{1}{n_{ij}}\sum_{(w_k,w_l)\in B_{ij}}g_{kl}(1-g_{kl}) \to g_0(1-g_0)$ as $K \to \infty$.

Note that

$$\sqrt{n_{ij}}\left(\frac{X_{ij}}{n_{ij}} - g_0\right) = \sqrt{n_{ij}}\left(\frac{X_{ij}}{n_{ij}} - g_{B_{ij}}\right) + \sqrt{n_{ij}}\left(g_{B_{ij}} - g_0\right),$$

and that

$$\sqrt{n_{ij}}\left|g_{B_{ij}} - g_0\right| \lesssim \frac{2hL}{K} = 2L\cdot\frac{n}{K^2} \to 0 \text{ as } \frac{n}{K^2} \to 0.$$

Thus, by Slutsky's lemma,

$$\frac{\sqrt{n_{ij}}\left(\frac{X_{ij}}{n_{ij}} - g_0\right)}{\sqrt{g_0(1-g_0)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

as $K \to \infty, n/K \to \infty, n/K^2 \to 0$.

(ii) It suffices to prove that for $b_{ij}|X_{ij} \sim \text{Beta}(X_{ij}, n_{ij} - X_{ij})$,

$$\frac{\sqrt{n_{ij}}\left(b_{ij} - \frac{X_{ij}}{n_{ij}}\right)}{\sqrt{g_0(1-g_0)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \text{ a.s. in } X_{ij}.$$

Since $\mathbb{E}(b_{ij}|X_{ij}) = \frac{X_{ij}}{n_{ij}}$, and $\text{Var}(b_{ij}|X_{ij}) = \frac{X_{ij}}{n_{ij}}\left(1 - \frac{X_{ij}}{n_{ij}}\right)\frac{1}{n_{ij}+1}$. Therefore, provided n_{ij} is large and $0 < a < X_{ij}/n_{ij} < b < 1$, the standardized b_{ij} is approximately normal [66]. In particular, conditional on X_{ij},

$$\frac{\sqrt{n_{ij}+1}\left(b_{ij} - \frac{X_{ij}}{n_{ij}}\right)}{\sqrt{\frac{X_{ij}}{n_{ij}}\left(1 - \frac{X_{ij}}{n_{ij}}\right)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \text{ a.s.},$$

as $n_{ij} \to \infty$ and $\frac{X_{ij}}{n_{ij}} \overset{a.s.}{\to} g_0 \in (0,1)$. Since $\sqrt{\frac{n_{ij}}{n_{ij}+1}} \to 1$, by Slutsky's lemma,

$$\frac{\sqrt{n_{ij}}\left(b_{ij} - \frac{X_{ij}}{n_{ij}}\right)}{\sqrt{g_0(1-g_0)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \text{ a.s. in } X_{ij}$$

as $K \to \infty, n/K \to \infty$. ■

D. Remark

These Theorems can be generalized to a more general class of graphons that are (piecewise) Hölder-continuous with constant L and α:

$$|g(\boldsymbol{w}) - g(\boldsymbol{w}')| \leq L \left\|\boldsymbol{w} - \boldsymbol{w}'\right\|^{\alpha}.$$

The result in Theorem 1 is not affected. In Theorem 2, we will need $\frac{2L}{K^{\alpha}} \to 0$ and $2L \cdot \frac{n}{K^{(1+\alpha)}} \to 0$.

Thus, we need to assume $K \to \infty$ to grow at least at the speed of $n^{1/(1+\alpha)}$ with $\alpha \in (0,1]$. When $\alpha = 1$, it reduces to the case of Lipschitz continuity. Recall that the relation between n and K under Lipschitz condition was $n/K \to \infty$, $n/K^2 \to 0$. With Hölder condition, the lower bound of K becomes $n/K^{(1+\alpha)} \to 0$, and the upper bound remains $n/K \to \infty$.

REFERENCES

[1] S. H. Chan, T. B. Costa, and E. M. Airoldi, "Estimation of exchangeable graph models by stochastic blockmodel approximation," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 293–296.

[2] P. Hoff, "Modeling homophily and stochastic equivalence in symmetric relational data," in *Advances in Neural Information Processing Systems*, 2008, pp. 657–664.

[3] O. Kallenberg, *Probabilistic symmetries and invariance principles*. New York: Springer Science & Business Media, 2006.

[4] J. Lloyd, P. Orbanz, Z. Ghahramani, and D. M. Roy, "Random function priors for exchangeable arrays with applications to graphs and relational data," in *Advances in Neural Information Processing Systems*, 2012, pp. 998–1006.

[5] L. Lovász, *Large networks and graph limits*. Rhode Island: American Mathematical Society Providence, 2012.

[6] P. Orbanz and D. M. Roy, "Bayesian models of graphs, arrays and other exchangeable random structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 437–461, 2015.

[7] P. Diaconis and S. Janson, "Graph limits and exchangeable random graphs," *Rendiconti di Matematica*, vol. 28, pp. 33–61, 2008.

[8] E. M. Airoldi, T. B. Costa, and S. H. Chan, "Stochastic blockmodel approximation of a graphon: Theory and consistent estimation," in *Advances in Neural Information Processing Systems*, 2013, pp. 692–700.

[9] C. Borgs, J. Chayes, A. Smith, and I. Zadik, "Revealing network structure, confidentially: Improved rates for node-private graphon estimation," in *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2018, pp. 533–543.

[10] P. E. Caines and M. Huang, "Graphon mean field games and the gmfg equations," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 4129–4134.

[11] A. Aurell, R. Carmona, G. Dayanıklı, and M. Laurière, "Finite state graphon games with applications to epidemics," *Dynamic Games and Applications*, pp. 1–33, 2022.

[12] R. Vizuete, P. Frasca, and F. Garin, "Graphon-based sensitivity analysis of SIS epidemics," *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 542–547, 2020.

[13] S. Maskey, R. Levie, and G. Kutyniok, "Transferability of graph neural networks: an extended graphon approach," *arXiv preprint arXiv:2109.10096*, 2021.

[14] A. Parada-Mayorga, L. Ruiz, and A. Ribeiro, "Graphon pooling in graph neural networks," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 860–864.

[15] A. SHOJAEI-FARD, "Graphon models in quantum physics," *IHES Preprint: IHES/M/20/03*, 2020.

[16] L. Ruiz, L. F. Chamon, and A. Ribeiro, "Graphon signal processing," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4961–4976, 2021.

[17] C. Borgs, J. Chayes, and A. Smith, "Private graphon estimation for sparse graphs," in *Advances in Neural Information Processing Systems*, 2015, pp. 1369–1377.

[18] E. Airoldi, T. Costa, and S. Chan, "A non-parametric perspective on network analysis: Theory and consistent estimation," *Advances in Neural Information Processing Systems (NIPS)*, vol. 26, pp. 692–700, 2013.

[19] S. Chan and E. Airoldi, "A consistent histogram estimator for exchangeable graph models," in *International Conference on Machine Learning*, 2014, pp. 208–216.

[20] C. Gao, Y. Lu, H. H. Zhou *et al.*, "Rate-optimal graphon estimation," *The Annals of Statistics*, vol. 43, pp. 2624–2652, 2015.

[21] R. He and T. Zheng, "Estimating exponential random graph models using sampled network data via graphon," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*

(ASONAM), 2016, pp. 112–119.

[22] O. Klopp, A. B. Tsybakov, N. Verzelen *et al.*, "Oracle inequalities for network models and sparse graphon estimation," *The Annals of Statistics*, vol. 45, pp. 316–354, 2017.

[23] Y. Su, R. K. W. Wong, and T. C. M. Lee, "Network estimation via graphon with node features," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, pp. 2078–2089, 2020.

[24] P. J. Wolfe and S. C. Olhede, "Nonparametric graphon estimation," *arXiv preprint arXiv:1309.5936*, 2013.

[25] Y. Zhang, E. Levina, and J. Zhu, "Estimating network edge probabilities by neighbourhood smoothing," *Biometrika*, vol. 104, pp. 771–783, 2017.

[26] P. Latouche and S. Robin, "Variational bayes model averaging for graphon functions and motif frequencies inference in w-graph models," *Statistics and Computing*, vol. 26, no. 6, pp. 1173–1185, 2016.

[27] P. Latouche, E. Birmele, and C. Ambroise, "Variational bayesian inference and complexity control for stochastic block models," *Statistical Modelling*, vol. 12, no. 1, pp. 93–115, 2012.

[28] A. Mele and L. Zhu, "Approximate variational estimation for a model of network formation," *The Review of Economics and Statistics*, pp. 1–30, 2017.

[29] J.-J. Daudin, F. Picard, and S. Robin, "A mixture model for random graphs," *Statistics and computing*, vol. 18, no. 2, pp. 173–183, 2008.

[30] V. Veitch and D. M. Roy, "Sampling and estimation for (sparse) exchangeable graphs," *The Annals of Statistics*, vol. 47, no. 6, pp. 3274–3299, 2019.

[31] P. Latouche, S. Robin, and S. Ouadah, "Goodness of fit of logistic regression models for random graphs," *Journal of Computational and Graphical Statistics*, vol. 27, no. 1, pp. 98–109, 2018.

[32] A. Mele, "A structural model of dense network formation," *Econometrica*, vol. 85, no. 3, pp. 825–850, 2017.

[33] S. Bhattacharyya, P. J. Bickel *et al.*, "Subsampling bootstrap of count features of networks," *The Annals of Statistics*, vol. 43, pp. 2384–2411, 2015.

[34] J. Chang, E. D. Kolaczyk, and Q. Yao, "Estimation of subgraph densities in noisy networks," *Journal of the American Statistical Association*, vol. 117, no. 537, pp. 361–374, 2020.

[35] Y. R. Gel, V. Lyubchich, and L. L. R. Ramirez, "Bootstrap quantification of estimation uncertainties in network degree distributions," *Scientific reports*, vol. 7, pp. 1–12, 2017.

[36] S. Ouadah, S. Robin, and P. Latouche, "Degree-based goodness-of-fit tests for heterogeneous random graph models: Independent and exchangeable cases," *Scandinavian Journal of Statistics*, vol. 47, no. 1, pp. 156–181, 2020.

[37] A. Green and C. R. Shalizi, "Bootstrapping exchangeable random graphs," *Electronic Journal of Statistics*, vol. 16, no. 1, pp. 1058–1095, 2022.

[38] S. Chandna and P.-A. Maugis, "Nonparametric regression for multiple heterogeneous networks," *arXiv preprint arXiv:2001.04938*, 2020.

[39] J. Yang, C. Han, and E. Airoldi, "Nonparametric estimation and testing of exchangeable graph models," in *Artificial Intelligence and Statistics*, 2014, pp. 1060–1067.

[40] J. Hannig, "On generalized fiducial inference," *Statistica Sinica*, vol. 19, pp. 491–544, 2009.

[41] R. A. Fisher, "Inverse probability," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 26. Cambridge University Press, 1930, pp. 528–535.

[42] ——, "The concepts of inverse probability and fiducial probability referring to unknown parameters," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 139, pp. 343–348, 1933.

[43] ——, "The fiducial argument in statistical inference," *Annals of eugenics*, vol. 6, pp. 391–398, 1935.

[44] J. Hannig, H. Iyer, R. C. S. Lai, and T. C. M. Lee, "Generalized fiducial inference: A review and new results," *Journal of the American Statistical Association*, vol. 111, pp. 1346–1361, 2016.

[45] D. J. Aldous, "Representations for partially exchangeable arrays of random variables," *Journal of Multivariate Analysis*, vol. 11, pp. 581–598, 1981.

[46] D. N. Hoover, "Relations on probability spaces and arrays of random variables," *Preprint, Institute for Advanced Study, Princeton*, 1979.

[47] P. J. Bickel and A. Chen, "A nonparametric view of network models and newman–girvan and other modularities," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 21 068–21 073, 2009.

[48] S. Janson, "A graphon counter example," *Discrete Mathematics*, vol. 343, p. 111836, 2020.

[49] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, "Clustering network layers with the strata multilayer stochastic block model," *IEEE transactions on network science and engineering*, vol. 3, pp. 95–105, 2016.

[50] S. L. Zabell *et al.*, "RA Fisher and fiducial argument," *Statistical Science*, vol. 7, pp. 369–387, 1992.

[51] R. Martin and C. Liu, "Inferential models: A framework for prior-free posterior probabilistic inference," *Journal of the American Statistical Association*, vol. 108, pp. 301–313, 2013.

[52] R. Martin, J. Zhang, and C. Liu, "Dempster–Shafer theory and statistical inference with weak beliefs," *Statistical Science*, vol. 25, pp. 72–87, 2010.

[53] M. Xie, K. Singh, and W. E. Strawderman, "Confidence distributions and a unifying framework for meta-analysis," *Journal of the American Statistical Association*, vol. 106, pp. 320–333, 2011.

[54] M.-G. Xie and K. Singh, "Confidence distribution, the frequentist distribution estimator of a parameter: A review," *International Statistical Review*, vol. 81, pp. 3–39, 2013.

[55] J. Zhang and C. Liu, "Dempster-Shafer inference with weak beliefs," *Statistica Sinica*, vol. 21, pp. 475–494, 2011.

[56] J. Hannig, "Generalized fiducial inference via discretization," *Statistica Sinica*, vol. 23, pp. 489–514, 2013.
[57] D. L. Sonderegger and J. Hannig, "Fiducial theory for free-knot splines," in *Contemporary Developments in Statistical Theory*. Springer, 2014, pp. 155–189.
[58] J. Hannig and T. C. M. Lee, "Generalized fiducial inference for wavelet regression," *Biometrika*, vol. 96, pp. 847–860, 2009.
[59] R. C. S. Lai, J. Hannig, and T. C. M. Lee, "Generalized fiducial inference for ultrahigh-dimensional regression," *Journal of the American Statistical Association*, vol. 110, pp. 760–772, 2015.
[60] D. Cai, N. Ackerman, and C. Freer, "An iterative step-function estimator for graphons," *arXiv preprint arXiv:1412.2129*, 2014.
[61] D. Choi, P. J. Wolfe *et al.*, "Co-clustering separately exchangeable network data," *The Annals of Statistics*, vol. 42, pp. 29–63, 2014.
[62] S. C. Olhede and P. J. Wolfe, "Network histograms and universality of blockmodel approximation," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 14 722–14 727, 2014.
[63] S. Chatterjee, "Matrix estimation by universal singular value thresholding," *The Annals of Statistics*, vol. 43, pp. 177–214, 2015.
[64] R. Rossi and N. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.
[65] O. Marchal, J. Arbel *et al.*, "On the sub-gaussianity of the beta and dirichlet distributions," *Electronic Communications in Probability*, vol. 22, 2017.
[66] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*. Wiley New York, 1994.

Yi Su received the B.S. degree in mathematics and statistics in 2015 from Nanjing University, Nanjing, China. In 2020, he completed a Ph.D. degree in Statistics at the University of California, Davis. His research interests include network data analysis, graph models and machine learning.

He currently works as a Data Scientist in the Applied Research group at LinkedIn, with a focus on forecasting models.

Jan Hannig received his Mgr (MS equivalent) in mathematics in 1996 from the Charles University, Prague, Czech Republic. He received Ph.D. in statistics and probability in 2000 from Michigan State University under the direction of Professor A.V. Skorokhod.

From 2000 to 2008 he was on the faculty of the Department of Statistics at Colorado State University where he was promoted to an Associate Professor. He has joined the Department of Statistics and Operation Research at the University of North Carolina at Chapel Hill in 2008 and was promoted to Professor in 2013. He is an elected member of International Statistical Institute and a fellow of the American Statistical Association and Institute of Mathematical Statistics. He has been a PI and co-PI on several federally funded projects. To date he has advised and co-advised 26 Ph.D. students and published over 75 peer reviewed publications. He serves or served as an associate editor of Journal of American Statistical Association, Journal of Computational and Graphical Statistics, Sankhya, Statistical Theory and Related Fields, Electronic Journal of Statistics and Stat. His research interests are: theoretical statistics, generalized fiducial inference, and applications to biology, engineering and forensic science.

Thomas C. M. Lee received the B.App.Sc. (Math) degree in 1992, and the B.Sc. (Hons) (Math) degree with University Medal in 1993, all from the University of Technology, Sydney, Australia. In 1997 he completed a Ph.D. degree jointly at Macquarie University and CSIRO Mathematical and Information Sciences, Sydney, Australia.

Currently he is Professor of Statistics and Associate Dean of the Faculty in Mathematical and Physical Sciences at the University of California, Davis. He is an elected Fellow of the American Association for the Advancement of Science (AAAS, Section M, Engineering), the American Statistical Association (ASA) and the Institute of Mathematical Statistics (IMS). From 2013 to 2015 he served as the editor-in-chief for the Journal of Computational and Graphical Statistics, and from 2015 to 2018 he served as the Chair of the Department of Statistics at UC Davis. His research interests include nonparametric and semiparametric modeling, statistical image and signal processing, machine learning, and statistical applications in other scientific disciplines.