Biometrika (2022), **xx**, x, *pp.* 1–19
Printed in Great Britain

Subsampling Sparse Graphons Under Minimal Assumptions

By Robert Lunde

Department of Statistics, University of Michigan,
Ann Arbor, Michigan 48109, U.S.A.
rlunde@umich.edu

Purnamrita Sarkar

Department of Statistics and Data Sciences, University of Texas
Austin, Texas 78712, U.S.A.
purna.sarkar@austin.utexas.edu

Summary

We study the properties of two subsampling procedures for networks (vertex subsampling and p-subsampling) under the sparse graphon model. The consistency of network subsampling is demonstrated under the minimal assumptions of weak convergence of corresponding network statistics and an (expected) subsample size growing to infinity slower than the number of vertices in the network. Furthermore, under appropriate sparsity conditions, we derive limiting distributions for the nonzero eigenvalues of an adjacency matrix under the sparse graphon model. Our weak convergence result implies the consistency of our subsampling procedures for eigenvalues under appropriate conditions.

1. Introduction

1.1. *Statistical Inference for Network Data*

The analysis of network data has quickly become one of the most active research areas in statistics. Many results are now known about canonical network models such as the stochastic block model and its many variants (Holland et al., 1983; Airoldi et al., 2008; Karrer & Newman, 2011), the generalized random dot product model (Young & Scheinerman, 2007; Rubin-Delanchy et al., 2022), and the latent space model (Hoff et al., 2002), among others. For recent developments on minimax rates of nonparametric estimation in the sparse graphon model defined in Section 1.2, see Gao & Ma (2021). However, the problem of statistical inference for common network statistics, particularly in the nonparametric setting, has been less studied. One exception is count statistics; the validity of two subsampling schemes for normalized count functionals of sparse graphons has been established by Bhattacharyya & Bickel (2015).

Subsampling is a general methodology that has been shown to exhibit first-order correctness under minimal assumptions for a wide range of data generating processes; for an overview, see Politis et al. (1999). While count functionals are an important class of statistics in network analysis, we will show that a more general theory is possible for sparse graphons. Before discussing our results in more detail, we will introduce the network model that we consider below.

© 2021 Biometrika Trust

1.2. *Graphons and Sparse Graphons*

Let $\{A^{(n)}\}_{n\in\mathbb{N}}$ denote a sequence of $n \times n$ adjacency matrices generated by the following model:

$$A_{ij}^{(n)} = A_{ji}^{(n)} = \mathbb{1}(\eta_{ij} \leq h_n(\xi_i, \xi_j)) \sim \text{Bernoulli}(h_n(\xi_i, \xi_j)), \tag{1}$$

where $h_n : [0,1]^2 \mapsto [0,1]$ is a symmetric measurable function and $\xi_i \sim \text{Uniform}[0,1]$ for $1 \leq i \leq n$ and $\eta_{ij} \sim \text{Uniform}[0,1]$ for $1 \leq i < j \leq n$. We assume that $A_{ii}^{(n)} = 0$. We drop dependence on n when appropriate. Formally, we require that $\{\xi_i\}_{i=1}^{\infty}$ and $\{\eta_{ij}\}_{(i,j)\in\mathbb{N}^2}$ are defined on the common probability space $(\Omega, \mathcal{F}, P)$.

When $h_n = h$ is fixed for all n, h is often referred to as a graphon, short for graph function, and (1) is known as the graphon model. Graphons are natural models for graphs that exhibit vertex exchangeability. The theorems of Aldous (1981) and Hoover (1979) imply that any binary jointly exchangeable infinite array may be represented as a mixture of processes for which the data generating process is given above, with h_n fixed for all n. For modeling purposes, it is common to fix one component of the mixture, and assume that a size n adjacency matrix is a partial observation from an infinite array. Alternatively, graphons arise as limits of convergent graph sequences, where "convergence" may be defined by one of several equivalent notions; see Lovász (2012).

Graphons are known to imply either empty or dense graphs; in the latter case, the expected number of edges is $\Theta(n^2)$. However, many real-world networks are known to have expected number of edges given by $o(n^2)$; therefore having h_n fixed as $n \to \infty$ is often inappropriate. Instead, following Bickel & Chen (2009), we will consider the following parametrization. Let:

$$\rho_n = P(A_{ij} = 1) = \int_0^1 \int_0^1 h_n(u,v)\ du\ dv.$$

It follows that we may express $h_n(u,v)$ as:

$$h_n(u,v) = \rho_n w_n(u,v),$$

where $w_n(u,v) : [0,1]^2 \mapsto \mathbb{R}$ is the conditional density of (ξ_i, ξ_j) given $A_{ij}^{(n)} = 1$. With this parametrization, it is natural to keep $w_n(u,v)$ fixed and let ρ_n vary with n. Doing so, we arrive at the following model:

$$h_n(u,v) = \rho_n w(u,v) \wedge 1, \tag{2}$$

where w is a symmetric, nonnegative function that satisfies $\int_0^1 \int_0^1 w(u,v)\ du\ dv = 1$. By letting $\rho_n \to 0$ at an appropriate rate, we may generate an appropriately sparse sequence of graphs. With this generalization, note that $w(u,v)$ and ρ_n may lose their original interpretation for any finite n. One may alternatively arrive at this model by considering L^p graphons of Borgs et al. (2019), where w is an element of $L^p([0,1]^2)$ instead of $L^\infty([0,1]^2)$. As noted in the above reference, an unbounded graphon allows power law degree distributions and allows sparse graphs to contain dense spots.

Notable alternative frameworks for sparse network models include the graphon process or graphex (Veitch & Roy, 2015; Borgs et al., 2017) and edge exchangeable random graphs (Crane & Dempsey, 2018). These models are based on different notions of exchangeability. While we focus on sparse graphons only, we will consider a subsampling procedure based on a natural sampling mechanism for graphexes in Section 2.5. See

Orbanz (2017) for further discussion on natural sampling mechanisms associated with various network models.

1.3. *Related Work on Inference for Network Data*

Aside from the work by Bhattacharyya & Bickel (2015), there has been some work involving subsampling/resampling networks in the statistics literature. For example, Ali et al. (2016) develop a subsampling method named *Netdis*, which consists of sampling nodes and forming two-step ego-networks for these nodes. In addition, Levin & Levina (2019) consider the problem of bootstrapping network statistics that are expressible as perturbed U-statistics for random dot product graphs. The procedures they consider involve estimating the latent positions first and bootstrapping the associated U-statistics. Green & Shalizi (2022) also propose two bootstrap procedures for conducting inference for count functionals, one based on an "empirical graphon" and the other based on a sieve procedure. Recently, Zhang & Xia (2022) have derived an Edgeworth expansion for count functionals as well as higher-order correctness of both subsampling and the empirical graphon bootstrap for counts under certain conditions. Around the time of this submission, we have also become aware of Naulet et al. (2021), which studies subsampling for count functionals of graphex processes.

In the computer science literature, various subsampling approaches have also been studied for approximate subgraph counting. In contrast to Bhattacharyya & Bickel (2015), who consider inference for graphon parameters, the aim here is to approximate subgraph counts in the graph up to a multiplicative constant in a computationally efficient manner. Recently, a literature on sublinear algorithms has emerged, where many of the proposed procedures are based on edge sampling; see Feige (2006); Goldreich & Ron (2008); Assadi et al. (2019); Eden et al. (2017); Gonen et al. (2010) and references therein. Approximate subgraph counting has also received significant attention in various streaming settings; see for example, Bar-Yosseff et al. (2002); Tsourakakis et al. (2009); Kane et al. (2012); McGregor et al. (2016); Bera & Chakrabarti (2017); Kallaugher & Price (2017); McGregor et al. (2019).

For the general problem of nonparametric inference for network data, some work has started to emerge involving hypothesis testing. Ghoshdastidar et al. (2017) develop a framework for nonparametric two-sample testing based on network statistics that satisfy a certain concentration property. Among the network statistics considered are eigenvalues of independent edge random graphs. Tang et al. (2017) also consider nonparametric two-sample testing, but in a setting where the networks are generated by a random dot product graph. Their test involves estimating the latent positions of the network and using the kernel MMD (Gretton et al., 2012) to test whether the latent positions are generated by the same distribution.

2. General TheoreâĹńms for Subsampling

2.1. *Problem Setup and Notation*

Consider a parameter θ of inferential interest and let $\hat{\theta}_n : \mathbb{R}^{n\times n} \mapsto \mathbb{R}$ be an estimator of θ. For certain statistics, it is more natural to view $\hat{\theta}_n$ as a function on the corresponding graph G with vertex set $V(G) = \{1, \ldots, n\}$ and edge set $E(G) = \{\{i, j\} \mid i, j \in V(G), A_{ij} = 1\}$; in these cases, we will use the equivalent graph-theoretic notation. We will assume $\hat{\theta}_n$ is invariant to permutations of the vertices. Let τ_n be a normalizing sequence satisfying $\tau_n \to \infty$ as $n \to \infty$; for statistics with standard asymptotics, $\tau_n = n^{1/2}$.

The distribution function of interest is given by:

$$J_n(t) = P\left[\tau_n\{\hat{\theta}_n(A^{(n)}) - \theta\} \leq t\right]. \tag{3}$$

Let $G_{n,b}$ denote the graph induced by the vertices $1, \ldots, b$, which has the corresponding edge set $\{(i,j) \in G_n \mid i,j \in V(G_n)\}$. Since subgraphs induced by b vertices all follow the same distribution under the sparse graphon model (1), ensuing statements related to the distribution subsampled statistics computed on $G_{n,b}$ hold for any induced subgraph. For each $n \geq 2$ and $2 \leq b \leq n$, let $\hat{\theta}_{n,b} : \{0,1\}^{b \times b} \mapsto \mathbb{R}$ be an estimator defined on the induced subgraph; for reasons that will be explained shortly, $\hat{\theta}_{n,b}$ need not be equal to $\hat{\theta}_b$, but will be closely related. Let $A^{(n,b)}$ denote the adjacency matrix corresponding to $G_{n,b}$ and define the following cumulative distribution function:

$$J_{n,b}(t) = P\left[\tau_b\{\hat{\theta}_{n,b}(A^{(n,b)}) - \theta\} \leq t\right]. \tag{4}$$

We impose the following condition:

Assumption. For a given sequence $\{b_n\}_{n\in\mathbb{N}}$ satisfying $b_n \to \infty$ as $n \to \infty$, there exists some non-degenerate limiting distribution $J(t)$, such that, for all continuity points of $J(\cdot)$:

$$|J_n(t) - J(t)| \to 0 \quad \text{and} \quad |J_{n,b}(t) - J(t)| \to 0.$$

The network statistics considered will often be normalized by some power of ρ_n. Assumption 1 requires that $J_{n,b}$ converges to J even when $\hat{\theta}_{n,b}$ is normalized by ρ_n.

2.2. *Examples*

Example 1 (*Maximum eigenvalue*). Let $\lambda_1(w)$ denote the eigenvalue of the integral operator associated with w; see Section 3 for details. The corresponding cumulative distribution functions of interest may be expressed as:

$$J_n(t) = P\left[n^{1/2}\{\lambda_1(A^{(n)}/n\rho_n) - \lambda_1(w)\} \leq t\right]$$
$$J_{n,b}(t) = P\left[b^{1/2}\{\lambda_1(A^{(n,b)}/b\rho_n) - \lambda_1(w)\} \leq t\right].$$

For Assumption 1 to be satisfied, $\rho_n = \omega(b_n^{-1/2})$ is necessary; however, stronger conditions are required if the graphon is unbounded; see Section 3 for details.

Example 2 (*Count functionals*). The count functionals that we consider here were first studied in Bickel et al. (2011). We begin by preparing some notation. Let G_n denote a graph generated by the sparse graphon model, with vertex set $V(G_n) = \{1, \ldots, n\}$ and edge set $E(G_n)$. Let R be a graph with vertex set $V(R) = \{1, \ldots, p\}$ and let $G_n[R]$ denote the subgraph induced by $V(R)$. Let $p = |V(R)|$, $e = |E(R)|$, and consider the following graphon parameter, which provides the probability that a given subgraph is contained in G_n:

$$\widetilde{Q}(R) = \rho_n^{-e} P(R \subseteq G_n[R]) = \int_{[0,1]^p} \prod_{\{i,j\} \in E(R)} w(x_i, x_j)\, dx_1 \ldots dx_p. \tag{5}$$

Now, will define some additional notation needed to define our estimator of $\widetilde{Q}(R)$. We say that two graphs are isomorphic ($R_1 \sim R_2$) if there exists a bijection $\sigma : V(R_1) \mapsto V(R_2)$ such that there is an edge between $\sigma(i)$ and $\sigma(j)$ if and only there is an edge present

between i and j; mathematically $\{i,j\} \in E(R_1) \iff \{\sigma(i),\sigma(j)\} \in E(R_2)$. Let $|\text{Iso}(R)|$ denote number of isomorphisms of R into $\{1,\ldots,p\}$. In addition, for a graph H such that $V(R) \subseteq V(H)$, let $H[R]$ denote the subgraph induced by $V(R)$. For some graph H with $|V(H)| = m$ where $m \geq p$, consider the following estimator of $\widetilde{Q}(R;H)$:

$$\widehat{Q}(R;H) = \frac{1}{\binom{m}{p}\rho_n^e |\text{Iso}(R)|} \sum_{S \sim R} \mathbb{1}(S \subseteq H[S]). \tag{6}$$

The corresponding cumulative distribution functions are given by:

$$J_n(t) = P\left[n^{1/2}\{\widehat{Q}(R;G_n) - \widetilde{Q}(R)\} \leq t\right]$$
$$J_{n,b}(t) = P\left[b^{1/2}\{\widehat{Q}(R;G_{n,b}) - \widetilde{Q}(R)\} \leq t\right].$$

where $G_{n,b}$ is a subgraph of G_n formed from b nodes. For acyclic graphs, Assumption 1 is satisfied if $b_n\rho_n \to \infty$ holds.

Example 3 (*Sample variance of rooted subgraph frequencies*). Now, we will introduce a class of count functionals that provides more local information compared to the count functionals considered in the previous example. Recall that R is a k-star if $|V(R)| = k+1$, $|E(R)| = k$, and the edges in the graph have the form (i,v) for some v $\in V(S)$ and each $i \neq v$. Let $C_{k,v}(H)$ denote the total number of k-stars in the graph H rooted at the vertex v, $|V(H)| = m$, and $\widehat{C}_{k,x}(H)$ denote the following scaled quantity:

$$\widehat{C}_{k,v}(H) = \frac{1}{\binom{m-1}{k}\rho_n^e} C_{k,v}(H).$$

Now let $\widehat{\sigma}_n^2(C_{k,v}(G_n))$ denote the sample variance of the rooted subgraph counts:

$$\widehat{\sigma}_n^2(C_{k,v}(G_n)) = \frac{1}{n}\sum_{v=1}^{n}\left(\widehat{C}_{k,v}(G_n) - \frac{1}{n}\sum_{v=1}^{n}\widehat{C}_{k,v}(G_n)\right)^2.$$

Furthermore, let $\sigma^2 = \lim_{n\to\infty}\widehat{\sigma}_n^2(C_{k,x}(G_n))$. The corresponding cumulative distribution functions are given by:

$$J_n(t) = P\left[n^{1/2}\{\widehat{\sigma}_n^2(C_{k,v}(G_n)) - \sigma^2\} \leq t\right]$$
$$J_{n,b}(t) = P\left[b^{1/2}\{\widehat{\sigma}_n^2(C_{k,v}(G_{n,b})) - \sigma^2\} \leq t\right].$$

In Supplement S6, we show that Assumption 1 holds so long as $\rho_n = \omega(1/b_n^{1/2})$. With the exception of Maugis (2020), who considers Lipschitz functions of local subgraph counts, theoretical results for local subgraphs are scarce in the literature. However, we choose not to elaborate on our result in the main text since our argument for establishing asymptotic normality is related to arguments in Bickel et al. (2011). Nevertheless, this example demonstrates the generality of our approach.

Example 4 (*Empirical Degree Distribution Function*). Suppose that $\rho_n := \rho$; that is the graphon is dense. Let $D_i = \frac{1}{n-1}\sum_{i=1}^{n} A_{ij}$ denote the normalized degree of node i and $\mathfrak{D}(y) = \int_0^1 w(u,v)\,dv$ denote its conditional expectation given ξ_i. Consider the cen-

tered and scaled empirical degree distribution function:

$$\mathbb{G}_n(y) = n^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(D_i \leq \mathfrak{D}(y)) - y\right].$$

Under the conditions $w \in C^3([0,1]^2)$, $w < 1-\epsilon$ and $\mathfrak{D}(y) > \epsilon$ for some $\epsilon \in (0, 1/2)$, Delmas et al. (2021) establish weak convergence of finite dimensional distributions; therefore, under these conditions, the empirical degree function is amenable to subsampling.

Example 5 (*Likelihood Ratio Statistics*). To assess goodness of fit of a stochastic block model or its variants, one often computes likelihood ratio statistics of the form:

$$L(\Theta_0, \Theta_1) = \frac{\sup_{\theta\in\Theta_0} g(A^{(n)}, \theta)}{\sup_{\theta\in\Theta_1} g(A^{(n)}, \theta)},$$

where $g(A^{(n)}, \theta)$ is formed by marginalizing out the unobserved latent variables in the full likelihood. In Wang & Bickel (2017), asymptotic normality of the likelihood ratio is established in the case where the comparison is between the true model and an underfitting model. Therefore, this likelihood ratio statistic is amenable to subsampling; in fact, calculating the likelihood ratio statistic on subsamples may be computationally advantageous for fitting methods such as semi-definite programming, studied by Yan et al. (2018); Amini & Levina (2018); Perry & Wein (2017); Guédon & Vershynin (2016).

To emphasize, Assumption 1 imposes implicit limitations on the sparsity level. Even if subsampling is valid for certain sequences, this condition will also impose an implicit restriction on how slowly b_n can grow. Furthermore, in our definition of $J_n(t)$, note that ρ_n is unknown. For ease of exposition, we will not introduce estimation of ρ_n in our theorems below, but in Section 2.4 we will show that subsampling may still be used to approximate $J_n(t)$ when ρ_n is estimated.

Since $J_n(t)$ is inaccessible, we will approximate it with an empirical quantity defined on subsamples of the data. Our first subsampling scheme computes the statistic of interest on each induced subgraph with b vertices. Bhattacharyya & Bickel (2015) refer to this procedure as uniform subsampling; in the present work, we use the terminology vertex subsampling to avoid confusion with the notion of uniform validity, which is discussed in Supplement S3 due to space limitations. Let $N_n = \binom{n}{b_n}$; we will drop the subscripts when there is no ambiguity. For a given n, let $S_{b,1}, S_{b,2}, \ldots S_{b,N}$ denote subsets of size b constructed from $\{1, \ldots, n\}$, arranged in any order. Furthermore, let $A^{(n,b,i)}$ denote the $b \times b$ adjacency matrix induced by the nodes in $S_{b,i}$. The resulting empirical cumulative distribution function may be expressed as follows:

$$L_{n,b}(t) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\left[\tau_b\{\hat{\theta}_{n,b}(A^{(n,b,i)}) - \hat{\theta}_n(A^{(n)})\} \leq t\right]. \tag{7}$$

Since θ is unobservable, it is customary to replace it with its empirical counterpart estimated on G_n. As long as $b_n = o(n)$, this substitution is asymptotically negligible.

2.3. *Consistency of Vertex Subsampling for Sparse Graphons*

In Theorem 1, we establish the consistency of vertex subsampling for sparse graphons under minimal conditions. Our proof hinges on a technique involving the independence of induced subgraphs when the node sets are disjoint. To make the similarites to the

theory for i.i.d. processes transparent, we have stated the theorem below in a manner analogous to Theorem 2.2.1 of Politis et al. (1999) (see Supplement S1.1 for the proof).

THEOREM 1 (CONSISTENCY OF VERTEX SUBSAMPLING FOR SPARSE GRAPHONS). *Assume* $\tau_b/\tau_n \to 0$, $b_n \to \infty$, *and* $b_n = o(n)$. *Further suppose that for the sequence* $\{b_n\}_{n\in\mathbb{N}}$, *Assumption 1 is satisfied. Then,*

i. If t *is a continuity point of* $J(\cdot)$, *then* $L_{n,b}(t) \to J(t)$ *in probability.*

ii. If $J(\cdot)$ *is continuous, then,*

$$\sup_{t\in\mathbb{R}} |L_{n,b}(t) - J_n(t)| \to 0 \quad \textit{in probability.}$$

iii. Let $c_{n,b}(1-\alpha) = \inf\{\ t \in \mathbb{R} \mid L_{n,b}(t) \geq 1-\alpha\}$.
Correspondingly define:

$$c(1-\alpha) = \inf\{\ t \in \mathbb{R} \mid J(t) \geq 1-\alpha\}.$$

If $J(\cdot)$ *is continuous at* $c(1-\alpha)$ *then,*

$$P\left[\tau_n\{\hat{\theta}_n(G_n) - \theta\} \leq c_{n,b}(1-\alpha)\right] \to 1-\alpha \quad \textit{as} \quad n \to \infty.$$

Remark 1. The above result may be generalized in several ways. The latent positions may instead follow any distribution as in Hoff et al. (2002). In addition, the above theorem holds for weighted/directed graphs as well as graphon models with nodal covariates, so long as the pairs $(\xi_i, X_i) \in \mathbb{R}^p$ are i.i.d.

We would like to mention that several auxiliary results in Politis et al. (1999) may be proved following similar reasoning. In particular, one does not need all $\binom{n}{b}$ subsamples of the data; a stochastic approximation involving $B \to \infty$ subsamples chosen with or without replacement suffices, following analogous reasoning to the proof of Corollary 2.4 of the above reference. However, certain results, such as the validity of subsampling with a data-driven subsample size (Theorem 2.7.1 of Politis et al. (1999)) require slightly stronger assumptions since $J_{n,b}(t)$ need not be equal to $J_b(t)$. We will not pursue this here, but we would like to mention that a similar issue arises in Section 2.5; granting Assumption 2 should be enough to extend the aforementioned result.

2.4. *On Subsampling with an Estimated Sparsity Parameter*

While ρ_n is assumed to be known in the above result, in practice, it will need to be estimated from data. A natural idea is to plug in the following estimator:

$$\hat{\rho}_n = \frac{1}{\binom{n}{2}} \sum_{i<j} A_{ij}.$$

In Theorem 1 of Bickel et al. (2011), it is shown that $\hat{\rho}_n/\rho_n$ converges to 1 at a $\sqrt{n}$-rate; therefore, replacing ρ_n with $\hat{\rho}_n$ turns out to be too large of a perturbation to be considered negligible. One may often use the Delta Method to establish convergence with an estimated sparsity parameter, but this often leads a higher variance. The maximum eigenvalue appears to be an exception; see Supplement S7 for a figure depicting this.

Fortunately, in many situations we may still approximate the sampling distribution through subsampling. In essence, if we plug in $\hat{\rho}_n$ estimated on G_n in each of the subsampled statistics, it will turn out that the estimate is accurate enough relative to the subsampled statistics so that it is asymptotically negligible. For confidence intervals, the

cost is a slight loss of efficiency, as one needs to consider a functional calculated on a subgraph with $m_n = o(n)$ as the centering of interval. However, this choice of centering does have the secondary benefit of reducing bias, which in finite samples substantially affects the coverage properties of subsampling; see Section 4 for further discussion.

Suppose that $\hat{\theta}_n(A^{(n)}; \rho) = \rho^{-s} f(A^{(n)})$ and $\hat{\theta}_{n,b}(A^{(n,b)}; \rho) = \rho^{-s} f_b(A^{(n,b)})$ for some $s \geq 0$ where the function f is independent of ρ, a sparsity parameter or estimate thereof. All of the examples considered in Section 2.2 may be expressed in this form and further have $\tau_n = n^{1/2}$. Define the following subsampling estimate:

$$\hat{L}_{n,b}(t) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left[\tau_b \left\{\hat{\theta}_{n,b}(A^{(n,b,i)}; \hat{\rho}_n) - \hat{\theta}_n(A^{(n)}; \hat{\rho}_n)\right\} \leq t\right]. \tag{8}$$

Furthermore, let $\hat{c}_{n,b}(\cdot)$ denote the quantile function associated with the distribution $\hat{L}_{n,b}(\cdot)$. We have the following result; the proof is given in Supplement S7. Note that in the proposition below, $\tau_n = n^{1/2}$ is imposed for simplicity but is not necessary.

PROPOSITION 1 (SUBSAMPLING WITH ESTIMATED SPARSITY). *Suppose that the assumptions in Theorem 1 are satisfied. Then, i. – iii. of Theorem 1 hold for* $\hat{L}_{n,b}(t)$. *Moreover, if* $m_n = o(n)$, $\tau_n = n^{1/2}$, *and* J *is continuous at* $c(\alpha/2)$ *and* $c(1-\alpha/2)$,

$$P\left(\theta \in \left[\hat{\theta}_{n,m}(A^{(n,m)}; \hat{\rho}_n) - \frac{\hat{c}_{n,b}(1-\alpha/2)}{\tau_m},\ \hat{\theta}_{n,m}(A^{(n,m)}; \hat{\rho}_n) - \frac{\hat{c}_{n,b}(\alpha/2)}{\tau_m}\right]\right) \to 1 - \alpha. \tag{9}$$

2.5. *Consistency of p-subsampling for Sparse Graphons*

We will now consider the validity of a subsampling procedure that involves repeatedly p-sampling a given graph. The notion of p-sampling was introduced by Veitch & Roy (2019) in the context of sampling for graphexes. This procedure is described below.

DEFINITION 1 (P-SAMPLING). *A p-sample of* G, *denoted* $\mathrm{Smpl}_p(G)$, *is a random subgraph obtained by including vertices independently with probability* p, *and taking the induced subgraph, with isolated vertices removed.*

Given a size n realization of a graphex, a p-sampling of the graph generates a smaller graphex; therefore, p-sampling may be viewed as the natural sampling mechanism for this process. Even under certain forms of misspecification for the sampling process, our result demonstrates that subsampling still produces valid inferences for sparse graphons under weak assumptions on the (expected) subsample size.

Since the subsample size is random, we will now denote it with B_i. Let $X_{ij}^{(n)} \sim \mathrm{Bernoulli}(p_n)$; this random variable indicates whether the jth node is included in the subsampled graph before the deletion of isolated vertices. We will assume the Bernoulli trials are generated independently from G_n. Denote the ith p-sample $\mathrm{Smpl}_{p,i}(G_n)$. Furthermore, let M denote the number of p-subsamples. In addition, we will modify Assumption 1 to accommodate the random sample size. Consider the following condition:

Assumption. For sequences $\{l_n\}_{n\in\mathbb{N}}$ and $\{u_n\}_{n\in\mathbb{N}}$ satisfying $l_n < u_n$ and $l_n \to \infty$, there exists some non-degenerate limiting distribution $J(t)$ such that, for all continuity points

of $J(t)$:

$$|J_n(t) - J(t)| \to 0 \text{ and } \sup_{l_n \le j \le u_n} |J_{n,j}(t) - J(t)| \to 0.$$

Now, let

$$l_n = \lfloor np_n - 3(np_n \log n)^{1/2} \rfloor, \quad u_n = \lceil np_n + 3(np_n \log n)^{1/2} \rceil. \tag{10}$$

Recall the following Chernoff bound for binomial random variables, which yields, for $0 < \epsilon < np_n$:

$$P\left(\left| \sum_{j=1}^{n} X_{ij}^{(n)} - E(X_{ij}^{(n)}) \right| > \epsilon \right) \le 2 \exp\left(\frac{-\epsilon^2}{3np_n} \right). \tag{11}$$

In light of Assumption 2 and the Chernoff bound, our choice of l_n and u_n ensures that the sequence with a random sample size converges in distribution with high probability. In all of the examples considered, Assumption 2 holds so long as suitable sparsity conditions for each example are met. If p_n is chosen so that a size np_n graph with sparsity parameter ρ_n satisfies these sparsity conditions, then Assumption 2 holds since the subsample sizes are all $\Theta(np_n)$. However, as a technicality, Assumption 2 is stronger than Assumption 1.

Since the number of vertices of the input graph is random, we will have to be slightly careful in formally defining $\hat{\theta}_{n,B_i}(\cdot)$. Let $\hat{\theta}_{n,B_i}(\cdot) : \cup_{b=0}^{n} \{0,1\}^{b \times b} \mapsto \mathbb{R}$ be equal to $\hat{\theta}_{n,j}(\cdot)$ when $|V(\text{Smpl}_{p,i}(G_n)| = j$ for $j \ge 2$ and 0 otherwise. Define the following empirical distribution function of the statistic of interest following a p-subsampling procedure:

$$L'_{n,B}(t) = \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}\left[\tau_{B_i} \left\{ \hat{\theta}_{n,B_i}(\text{Smpl}_{p,i}(G_n)) - \hat{\theta}_n(G_n) \right\} \le t \right]. \tag{12}$$

We are now ready to state our result. Our proof strategy involves approximating the empirical cumulative distribution function of the p-subsampled statistic with a convex combination of U-statistics, corresponding to its conditional expectation given G_n. This approximation demonstrates that vertex sampling and p-sampling are closely linked, which is an interesting concept in its own right. See Supplement S1.3 for details.

Theorem 2 (Consistency of $p-$subsampling for Sparse Graphons). *Assume that $\tau_n = n^\alpha$ for some $\alpha > 0$, $M \to \infty$, $np_n \to \infty$, and $p_n = o(1)$. Further suppose for l_n, u_n given in (10), Assumption 2 is satisfied and $(\rho_n, w(u,v))$ satisfy $\rho_n = \omega(\log l_n / l_n)$, and $w(\xi_i, \xi_j) \ge c$ almost surely, where $c > 0$. Then, $i. - iii.$ of Theorem 1 hold for $L'_{n,B}(t)$.*

Remark 2. The conditions on $(\rho_n, w(u,v))$ above ensure that the deletion of isolated vertices is negligible. However, certain statistics such as eigenvalues are invariant to deletion of isolated vertices; Lovász (2012) refers to these graph functionals as isolate-indifferent. For such statistics, these explicit conditions are not necessary.

3. Weak Convergence of Nonzero Eigenvalues of Sparse Graphons

In this section, we state a weak convergence result for eigenvalues of adjacency matrices generated by sparse graphons of finite rank. Our motivation is to derive conditions under which subsampling may be used to conduct inference for eigenvalues, which capture salient features of the underlying graph. Eigenvalues of the adjacency matrix have been

used widely in applied research. For example, the maximum eigenvalue of brain functional networks have been used to estimate mental fatigue (Li et al., 2020). The inverse of the largest eigenvalue equals the epidemic threshold in a network (Van Mieghem et al., 2009) and is proportional to its synchronization threshold (Restrepo et al., 2005). The eigenspectrum of adjacency matrices are often used to compare two networks (Gera et al., 2018; Mukherjee et al., 2017). The ratio of eigenvalues has been used to understand how individual vs. collective motivations drive complex socio-economic networks (Iranzo et al., 2020). The spectral gap between the two largest eigenvalues of an adjacency matrix is used to measure expansion properties of networks (Hoory et al., 2006) and have been used to conduct exploratory analysis of large social networks (Malliaros & Megalooikonomou, 2011). The spectral gap is also used to understand how fast a dynamic process in a network will reach the steady state (Mieghem, 2010).

To our knowledge, we are the first to consider inference for eigenvalues under the sparse graphon model. Limit theorems are one of the most important topics in random matrix theory, so our result here is of independent interest. The main tools we use are results on random matrix approximations of integral operators due to Koltchinskii & Giné (2001) and refined eigenvalue perturbation bounds developed by Eldridge et al. (2018). To use the latter, an upper bound on the operator norm of the centered adjacency matrix is needed; however, sharp results (e.g. Vu (2007)) require independent entries. To handle the graphon dependence structure, we develop a technique that involves conditioning on the latent positions and using Egorov's Theorem to control the conditional probability uniformly on a high probability set.

For Erdős-Rènyi graphs, limiting distributions of the eigenvalues are well-known. After appropriate centering and centering, the leading eigenvalue of the adjacency matrix generated by Erdős-Rènyi graphs has been shown to converge to the normal distribution by Füredi & Komlós (1981). The recent work of Tang (2018) considers limiting distributions for (dense) random graphs centered at a data-dependent quantity, which rules out subsampling. Beyond this, weak convergence results were not previously known for more general classes of models.

We would like to note that Borgs et al. (2012) show that the scaled eigenvalues of an adjacency matrix generated by a graphon converge to a limiting quantity. Their result is a law of large numbers for spectra; what we show is along the lines of a central limit theorem. We will now introduce some concepts needed to state our result.

Let $w : [0,1]^2 \mapsto \mathbb{R}$ be a symmetric element of $L^2([0,1]^2)$. Consider the following integral operator associated with w, which we will denote $T_w : L^2([0,1]) \mapsto L^2([0,1])$:

$$T_w f = \int_0^1 w(u,v) f(v) dv. \tag{13}$$

By the spectral theorem for compact, self-adjoint operators (see for example, Corollary 4.10.2 of Debnath et al. (2005)), there exists an orthonormal collection of eigenfunctions $\{\phi_r,\ r \in J\}$, where J is either finite or countably infinite, and a sequence of real numbers $\{\lambda_r,\ r \in J\}$ satisfying $\sum_{r \in J} \lambda_r^2 < \infty$ such that:

$$w(u,v) = \sum_{r \in J} \lambda_r \phi_r(u) \phi_r(v). \tag{14}$$

in the $L^2([0,1]^2)$ sense. We will consider the ordering $\lambda_1 \geq \lambda_2 \geq \ldots > 0 > \ldots \geq \lambda_{-2} \geq \lambda_{-1}$, where negative indices correspond to negative eigenvalues. We will not consider zero eigenvalues. We will denote the eigenvalues associated with w as $\lambda_r(w)$.

Let $\widetilde{A}^{(n)} = A^{(n)}/n\rho_n$. In the theorem below, we will establish conditions under which:

$$Z_{n,r} = n^{1/2}\{\lambda_r(\widetilde{A}^{(n)}) - \lambda_r(w)\} \tag{15}$$

weakly converges to a limiting distribution $Z_{\infty,r}$. In fact, we may show stronger statements involving convergence of joint distributions. While we require that the kernel is finite rank, this assumption is, strictly speaking, not necessary for establishing marginal convergence; see Remark 3 for details. We will now state our result below under various assumptions on the pair (ρ, w). A proof of this theorem is given in Supplement S2.1.

THEOREM 3 (WEAK CONVERGENCE OF NONZERO EIGENVALUES). *Let $\{G_n\}_{n\in\mathbb{N}}$ be a sequence of graphs generated by the model (2). Suppose $w(u,v)$ is an element of $L^2([0,1]^2)$ with an eigendecomposition of the form:*

$$w(u,v) = \sum_{r=1}^{k} \lambda_r \phi_r(u)\phi_r(v)$$

for some $k < \infty$ where $\lambda_r \neq 0$ for all $1 \leq r \leq k$, and one of the following conditions are satisfied:

A. (Boundedness) $\|w(u,v)\|_\infty < \infty$ almost surely, $\rho_n = o(1)$ and $\rho_n = \omega(n^{-1/2})$.

B. (sub-Weibull) There exists some universal constant $K < \infty$ and $\gamma > 0$ such that:

$$E\left[\exp\left\{\frac{w(\xi_i,\xi_j) - E(w(\xi_i,\xi_j))}{K}\right\}^\gamma\right] \leq 2. \tag{16}$$

Furthermore, for $1 \leq r \leq k$, the eigenfunctions satisfy:

$$\int_0^1 \phi_r^4(u)\, du < \infty. \tag{17}$$

Further suppose that $\rho_n = O(n^{-\delta})$ and $\rho_n = \omega(n^{-1/2+\delta})$ for some $\delta > 0$.

C. (Moment) Suppose that $E[w^s(\xi_i,\xi_j)] < \infty$ for some $s > 3 + \sqrt{5}$. Further suppose the eigenfunctions satisfy (17) and ρ_n satisfies $n^{1+\delta}\rho_n^{s-2} = o(1)$ and $\rho_n = \omega\left(n^{(2-s+\delta)/2s}\right)$ for some $\delta > 0$.

Let $Z_n = (Z_{n,1}, \ldots, Z_{n,k})$, where $Z_{n,r}$ are given by (15). Then, there exists some limiting random variable Z_∞ such that:

$$Z_n \to Z_\infty \quad \text{in distribution.} \tag{18}$$

Furthermore, if $\lambda_1, \ldots, \lambda_k$ are distinct, then Z_∞ is multivariate Gaussian.

Remark 3. Following Theorem 5.1 of Koltchinskii & Giné (2001), one may impose the following condition for graphons that are not finite rank. For some $R_n \to \infty$, suppose that w satisfies $\sum_{|r|>R_n} \lambda_r^2 = o(n^{-1})$. Further suppose that:

$$\left(\sum_{|r|\leq R_n,\ |s|\leq R_n} \int \phi_r^2\phi_s^2 dP\right) \times \left(\sum_{|r|\leq R_n,\ |s|\leq R_n} (\lambda_r^2 + \lambda_s^2) \int \phi_r^2\phi_s^2 dP\right) = o(n).$$

Moreover, suppose that $\sum_{r\in\mathbb{Z}} |\lambda_r|\phi_r^2 \in L^2(P)$. Then, one may show convergence of finite-dimensional distributions of the nonzero eigenvalues of interest. However, it seems that verifying these conditions is non-trivial outside of the finite-rank case. For instance, it appears that smoothness properties by themselves do not imply the conditions above.

Remark 4. Cases B and C impose various tail conditions on unbounded graphons, which are more expressive than bounded graphons (Borgs et al., 2019), allowing a wider range of degree distributions. However, they complicate arguments that involve conditioning on the latent positions. To overcome this difficulty, we use Egorov's Theorem to establish uniform convergence of conditional probability statements over a high probability set, which implies convergence in probability of a nuisance term.

Remark 5. For bounded graphons, we show in Supplement S5.2 that the conditions in the previous remark may be simplified. For this case, we require $\rho_n = \omega(n^{-1/2})$, T_w is trace class, and for some $R_n \to \infty$, $\frac{R_n}{n}(\sum_{1\leq r\leq R_n} \frac{1}{\lambda_r^2})^{3/2} \to 0$. The trace class condition implies $\sum_{r\in\mathbb{Z}} |\lambda_r|\phi_r^2 \in L^2(P)$; see for example, the proof of Proposition 3.2 of Lei (2021).

We will now discuss some of the conditions in our theorem. Finite rank graphons are equivalent to the rich family of generalized random dot product models (Rubin-Delanchy et al., 2022), which are particularly expressive when the dimension of the latent space is allowed to be any fixed natural number. In addition, while we require that the graphon is rank k, the rank need not be known a priori. For any $k' \leq k$, the theorem above provides joint convergence of the k' nonzero eigenvalues. Furthermore, since we derive joint convergence, one may use the Delta Method to derive weak convergence for certain differentiable functions of the nonzero k' eigenvalues. We provide examples of eigenvalue statistics that may be of interest below.

Example 6 (*Statistics Based on Disparities between Eigenvalues*). For the purpose of comparing two networks, it may be of interest to consider certain statistics of two eigenvalues. Two natural choices are spectral gaps and eigenvalue ratios, as defined below:

$$\hat{\theta}_{\text{gap}}(A^{(n)}) = \frac{\lambda_1(A^{(n)}) - \lambda_2(A^{(n)})}{n\rho_n}, \quad \hat{\theta}_{\text{ratio}}(A^{(n)}) = \frac{\lambda_1(A^{(n)})}{\lambda_{k'}(A^{(n)})}. \tag{19}$$

Example 7 (*Approximate Trace*). Another important statistic in network analysis is $\text{tr}(A^p)$. This statistic is related to subgraph counts; more precisely, it provides the number of closed walks of length p from any vertex back to itself. It is well known that

$$\text{tr}(A^p) = \sum_{r=1}^{n} \lambda_r^p(A).$$

Consider the statistic $\hat{\theta}_{\text{trace},p,k'}(A^{(n)}) = \sum_{r=1}^{k'} \lambda_r^p(A^{(n)}/n\rho_n)$.

This statistic is a suitable approximation to its population counterpart $\sum_{r=1}^{k'} \lambda_r^p(w)$.

While our result here is sufficiently general, it should be mentioned that, even for bounded graphons, our theorem requires $\rho_n = \omega(n^{-1/2})$. In general, it seems difficult to improve the eigenvalue perturbation bounds to weaken conditions for concentration; O'Rourke et al. (2018) derive similar bounds under a low rank hypothesis for the mean matrix. In addition, for sub-Weibull and L^p graphons, we impose upper bounds on the rate of decay of ρ_n, which may seem unusual. These upper bounds allow us to control the difference between eigenvalues of the mean matrix, with entries given by $\rho_n w(\xi_i, \xi_j) \wedge 1$, and a matrix with entries given by $\rho_n w(\xi_i, \xi_j)$. In essence, if we allow w to be unbounded, we need a sparser graph sequence to observe most of the uncensored values of $w(\xi_i, \xi_j)$.

4. Simulation Study

We investigate the finite-sample properties of confidence intervals for various eigenvalue statistics formed by subsampling. We consider two sparse graphon models; for these models, it will be more natural to consider the following parameterization:

$$h_n(u,v) = P(A_{ij} = 1 \mid \xi_i = u, \xi_j = v) = \nu_n h(u,v), \tag{20}$$

where $h(u,v)$ is a dense graphon and ν_n is a sparsity parameter. It then follows that $\rho_n = \nu_n \int h(\xi_i, \xi_j)\, dP$. We study the performance of our method for sample sizes ranging from $n = 1000$ to $n = 7000$ and sparsity parameters ranging from $\nu_n = n^{-0.1}$ to $\nu_n = n^{-0.45}$. For vertex subsampling, we consider subsample sizes of the form $b_n = cn/\log n$ for $c \in \{1, 1.5, 2, 2.5, 3\}$; for p-subsampling, we consider p_n corresponding to an equivalent expected subsample size.

The data generating processes considered in our simulation study are described below. For each of these processes, we approximate the true parameter $\lambda_i(w)$ by simulating a graph of size $n = 20000$ and $\nu_n = 1$. We then approximate the population parameter with $\lambda_i(A^{(n)})/n\hat{\rho}_n$. We assess coverage by counting the number of times the parameter falls within our confidence interval; to this end, we simulate the model 500 times and construct a confidence interval from $N = 500$ subsamples for each iteration. We construct confidence intervals of the form (9), with $m = 5n/\log n$.

4.1. *Stochastic Block Model*

We consider a three class stochastic block model studied in Lei (2021), with parameters:

$$B = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/6 \end{pmatrix}, \quad \pi = (0.3, 0.3, 0.4). \tag{21}$$

The corresponding graphon is rank 2 and has one positive and one negative eigenvalue, with $\lambda_1 = 1.035$ and $\lambda_2 = -0.267$.

4.2. *Gaussian Latent Space Model*

We also investigate the properties of our procedure for a graphon model that is not low rank. The following model is a special case of the Gaussian latent space model studied in Rastelli et al. (2016). Let $\xi_i \sim N(0,1)$ and define:

$$h_n(u,v) = \nu_n \exp\{-25(u-v)^2\}. \tag{22}$$

We study the behavior of our procedure for the top 3 positive eigenvalues; the associated parameters are: $\lambda_1 = 1.311, \lambda_2 = 1.147$, and $\lambda_3 = 1.011$.

4.3. *Simulation Results*

Our simulation results suggest that p-subsampling and vertex subsampling have very similar coverage properties for eigenvalues. It also appears that subsampling offers strong finite sample performance for an appropriate choice of subsample size. Even for the most difficult sparsity setting $\nu_n = n^{-0.45}$, we see that coverage of our subsampling procedure with an appropriately chosen subsample size approaches the nominal level as n increases for the eigenvalues considered.

Another takeaway is that the sparsity level increases, the performance of subsampling becomes more sensitive to the choice of subsample size. Nevertheless, it is reassuring that certain subsample regimes perform well in nearly all settings considered. Subsampling appears to perform best with larger subsampling sizes, corresponding to $b_n = 2.5n/\log n$

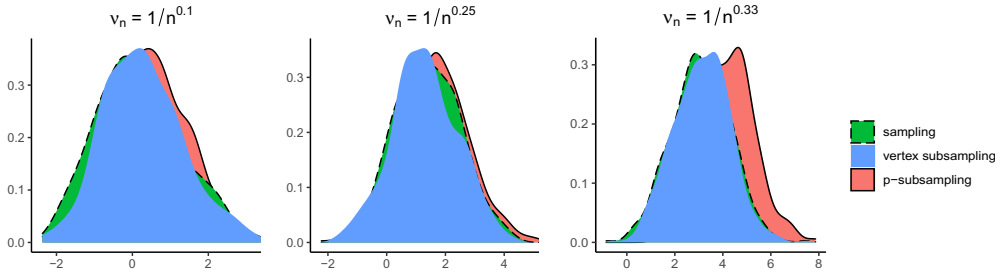

Fig. 1: Sampling and subsampling distributions for inference on $\lambda_2(w)$ of the Gaussian latent space model. The sampling distribution is formed from 500 graphs, where for each graph of 7000 vertices, we compute the eigenvalues of interest on a size $5n/\log n$ subsample. The subsampling distributions are formed from 500 subsamples of a given graph with $n = 7000$ vertices, and an (expected) subsample size of $b_n = 3n/\log n$.

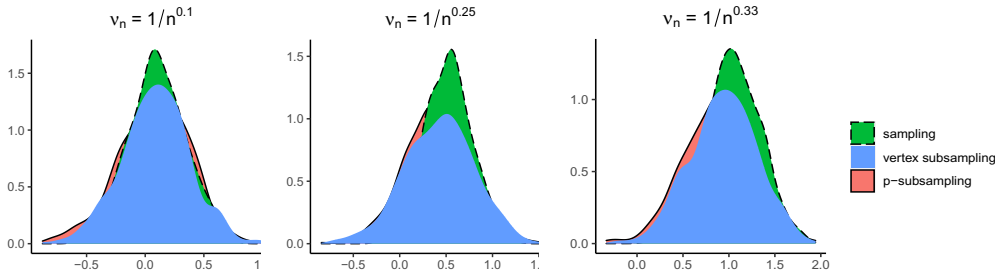

Fig. 2: Sampling and subsampling distributions for inference on $\lambda_1(w)$ of the stochastic block model. Simulation settings are identical to those described in Figure 1.

and $b_n = 3n/\log n$. This is most likely due to the bias of our estimator; for larger subsample sizes, the bias of the subsampled eigenvalues and that of the size m functional are more likely to be close.

We present the distributions of the subsampled λ_2 for a network with 7000 nodes generated from a Gaussian latent space model (Section 4.2), with average degree decaying from the left to the rightmost panel in Figure 1. The behavior of λ_1 (see the table in Section S8 of the Supplement) shows less degradation as one increases sparsity. We include similar plots of λ_1 and λ_2 for the stochastic block model (Section 4.1) in Figures 2 and 3. As Figures 1 and 2 indicate, the bias appears to be less problematic for eigenvalues that are well-separated from the bulk.

In Supplement S9, we also provide a comparison of subsampling with the empirical graphon bootstrap of Green & Shalizi (2022) with resample size n and a semiparametric bootstrap for random dot product graph models proposed by Levin & Levina (2019). Neither procedure has been proven to work for eigenvalues and our simulation study suggest that they may not be suitable for doing inference for eigenvalues.

5. Real Data Example: Facebook Networks

Using Facebook networks from 100 universities in 2005 (Traud et al., 2012), we perform two sample tests to see whether two social networks are generated by the same sparse

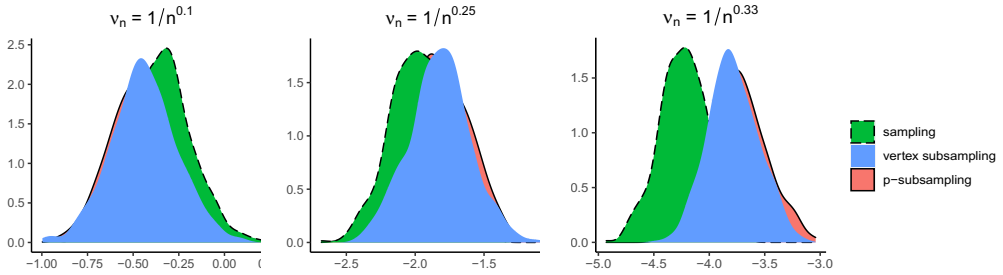

Fig. 3: Sampling and subsampling distributions for inference on $\lambda_2(w)$ of the stochastic block model. Simulation settings are identical to those described in Figure 1.

graphon. Facebook networks are well-suited for our method, as they are typically large and not too sparse. They are also undirected, ensuring real-valued spectra.

We will consider a two-sample test based on subsampled eigenvalues for comparing social networks of different universities, similar to the two sample test with subsampled count functionals used by Bhattacharyya & Bickel (2015). Here, we examine the relative merits of using subsampled eigenvalue statistics for comparing two networks.

Before proceeding, we would like to mention some of the strengths and weaknesses of the eigenvalue approach compared to the subgraph approach that are not directly related to the power of the test. First, note that the entire spectrum of a symmetric matrix can be computed in $O(n^3)$ time and calculating the top few eigenvalues can be performed even faster. For count functionals, a brute-force subgraph search with $p = |\mathcal{V}(R)|$ has complexity $O(n^p)$. However, as noted in Bickel et al. (2011), for sparser graphs, certain functionals such as kl-wheels can be counted faster. From a computational perspective, eigenvalue statistics generally may be preferable for network comparisons. On the other hand, eigenvalue statistics require stronger conditions on the sparsity level; consequently, count functionals may be safer to use for sparser graphs.

With eigenvalue statistics, it is also not clear a priori which statistics will be most effective at distinguishing two networks. To limit false discoveries, we consider a sample splitting procedure in which one subnetwork is used for formulating hypotheses and the other is used for testing the hypotheses. The most natural procedure under the model (Eq 1) is node-splitting, in which nodes are randomly split into two disjoint sets and the corresponding induced subgraphs are used. Under our model, the induced subgraphs are independent and thus, inferences based on node-splitting are valid.

5.1. *Data Analysis Results*

In this section, we compare the Facebook networks of University of Pennsylvania (Penn) with Columbia and Yale with Princeton. For each comparison, we randomly split the vertex set of each network into two parts. The "training" halves are used to find k ($k \leq 5$) such that λ_k has the largest difference between the training subnetworks of the two exploratory networks. Then, on the held-out subnetworks, we use subsampling to perform a two-sample test. We mainly consider principal eigenvalues since inference becomes harder for eigenvalues closer to the bulk. For each network, a 97.5% confidence interval is constructed for the test statistic. If the confidence intervals are disjoint, we may reject the null hypothesis that the graphs were generated by the same graphon

Schools	n	Test statistic	Value	97.5% CI	Decision
Columbia	5885	$\lambda_1/m\hat{\rho}_n$	2.757	$(2.202, 3.080)$	Reject H_0
Penn	7458		2.067	$(1.830, 2.166)$	
Princeton	3298	$\lambda_4/m\hat{\rho}_n$	0.943	$(0.771, 0.957)$	Fail to reject H_0
Yale	4289		1.019	$(0.846, 1.011)$	

Table 1: Two-sample test results

with a significance level of $\alpha = 0.05$. The confidence intervals are constructed via vertex subsampling, with a subsample size given by $b_n = 2.5n/\log n$ and $m_n = 5n/\log n$.

As shown in Table 1, we reject the null hypothesis for Columbia vs Penn, but fail to reject for Princeton vs Yale. Eigenvalue plots (see Supplement S10) suggested that it would be difficult to distinguish between the spectra of the Yale and Princeton networks, so it is not surprising that we were not able to reject the null hypothesis here.

We would like to emphasize that the validity of subsampling hinges on a weak convergence result, which requires a non-trivial sparsity condition. Since our notion of sparsity pertains to a sequence of graphs and we only have one observation for each school, it is difficult to say whether this condition is satisfied. However, we would like to note that one would expect the confidence intervals to be very wide if the graphs were too sparse, which does not appear to be the case.

6. Discussion

In this paper, we establish the validity of two subsampling schemes for network data. Our work leaves open the possibility of more computationally tractable approaches to inference for count functionals of sparse graphons. If a weak convergence result can be shown for approximate subgraph counting methods, then subsampling may be established using the results in this paper. It would be interesting to investigate the relationship between the maximum eigenvalue and the average degree for graphon models in greater detail. For general graphs, it is well-known that the maximum eigenvalue is bounded between the average degree and maximum degree; see, for example, Spielman (2012). For sparse graphons, our simulations in Section S7 of the Supplement suggest that the maximum eigenvalue and the average degree are also highly correlated under general conditions. It would be of interest to derive conditions under which an estimator of the maximum eigenvalue of the graphon operator with an estimated sparsity parameter outperforms an oracle estimator that uses the true sparsity parameter.

Acknowledgements

The authors would like to thank the editor and reviewers for helpful suggestions. The authors also gratefully acknowledge support from NSF 2019844 and NSF HDR-1934932.

Supplementary material

The supplementary material contains proofs of the main theorems, results and proofs on uniform validity, weak convergence results for the sample variance of rooted stars, a comment on a phenomenon involving the maximum eigenvalue, additional simulation results including comparison to other resampling methods, and additional details on real data analysis.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E. & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* **9**, 1981–2014.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis* **11**, 581–598.

Ali, W., Wegner, A. E., Gaunt, R. E., Deane, C. M. & Reinert, G. (2016). Comparison of large networks with sub-sampling strategies. *Scientific Reports* **6**.

Amini, A. A. & Levina, E. (2018). On semidefinite relaxations for the block model. *Ann. Statist.* **46**, 149–179.

Assadi, S., Kapralov, M. & Khanna, S. (2019). A simple sublinear-time algorithm for counting arbitrary subgraphs via edge sampling. In *Proceedings of the Tenth Innovations in Theoretical Computer Science Conference*, ITCS 2019.

Bar-Yosseff, Z., Kumar, R. & Sivakumar, D. (2002). Reductions in streaming algorithms, with an application to counting triangles in graphs. In *Proceedings of SODA*.

Bera, S. K. & Chakrabarti, A. (2017). Towards tighter space bounds for counting triangles and other substructures in graph streams. In *34th Symposium on Theoretical Aspects of ComputerScience*, STACS 2017.

Bhattacharyya, S. & Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *Annals of Statistics* **43**, 2384–2411.

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 (USA)* **106**, 21068–21073.

Bickel, P. J., Chen, A. & Levina, E. (2011). The method of moments and degree distributions for network models. *Annals of Statistics* **39**, 38–59.

Borgs, C., Chayes, J. T., Cohn, H. & Holden, N. (2017). Sparse exchangeable graphs and their limits via graphon processes. *Journal of Machine Learning Research* **18**, 7740–7810.

Borgs, C., Chayes, J. T., Cohn, H. & Zhao, Y. (2019). An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society* **372**, 3019–3062.

Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T. & Vesztergombi, K. (2012). Convergent sequences of dense graphs II: Multiway cuts and statistical physics. *Annals of Mathematics* **176**, 151–219.

Crane, H. & Dempsey, W. (2018). Edge exchangeable models for interaction networks. *JASA* **113**, 1311–1326.

Debnath, L., & Mikusinski, P. (2005). *Hilbert Spaces with Applications*. Elsevier Academy Press.

Delmas, J.-F., Dhersin, J.-S. & Sciauveau, M. (2021). Asymptotic for the cumulative distribution function of the degrees and homomorphism densities for random graphs sampled from a graphon. *Random Structures and Algorithms* **58**, 94–149.

Eden, T., Levi, A., Ron, D. & Seshadhri, C. (2017). Approximately counting triangles in sublinear time. *SIAM Journal of Computing* **46**, 1603–1646.

Eldridge, J., Belkin, M. & Wang, Y. (2018). Unperturbed: spectral analysis beyond Davis-Kahan. In *Proceedings of Algorithmic Learning Theory (ALT)*, vol. 83.

Feige, U. (2006). On sums of independent random variables with unbounded variance, and estimating the average degree in a graph. *SIAM Journal of Computing* **35**, 964–984.

Füredi, Z. & Komlós, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1**, 233–241.

Gao, C. & Ma, Z. (2021). Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statistical Science* **36**, 16–33.

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**, 2.

Ghoshdastidar, D., Gutzeit, M., Carpentier, A. & von Luxburg, U. (2017). Two-sample tests for large random graphs using network statistics. In *Conference on Learning Theory (COLT)*.

GOLDREICH, O. & RON, D. (2008). Approximating average parameters of graphs. *Random Structures and Algorithms* **32**, 473–493.

GONEN, M., RON, D. & SHAVITT, Y. (2010). Counting stars and other small subgraphs in sublinear time. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*.

GREEN, A. & SHALIZI, C. (2022). Bootstrapping exchangeable random graphs. *Electronic Journal of Statistics* **16**, 1058–1095.

GRETTON, A., BORGWARDT, K. M., RASCHAND, M. J., SCHÖLKOPF, B. & SMOLA, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research* **13**, 723–773.

GUÉDON, O. & VERSHYNIN, R. (2016). Community detection in sparse networks via GrothendieckâĂŹs inequality. *Probability Theory and Related Fields* **165**, 1025–1049.

HOFF, P. D., RAFTERY, A. E. & HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97**, 1090–1098.

HOLLAND, P. W., LASKEY, K. & LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5**, 109 – 137.

HOORY, S., LINIAL, N. & WIGDERSON, A. (2006). Expander graphs and their applications. *BULL. AMER. MATH. SOC.* **43**, 439–561.

HOOVER, D. N. (1979). *Relations on probability spaces arrays of random variables*. Institute for Advanced Study,RI.

IRANZO, J., PABLO-MARTÍ, F. & AGUIRRE, J. (2020). Emergence of complex socioeconomic networks driven by individual and collective interests. *Phys. Rev. Research* **2**, 043352.

KALLAUGHER, J. & PRICE, E. (2017). A hybrid sampling scheme for triangle counting. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA 2017.

KANE, D. M., MEHLHORN, K., SAUERWALD, T. & SUN, H. (2012). Counting arbitrary subgraphs in data streams. In *International Colloquium on Automata, Languages, and Programming*, ICALP 2012.

KARRER, B. & NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E* **83**, 1–24.

KOLTCHINSKII, V. & GINÉ, E. (2001). Random matrix approximation of spectra of integral operators. *Bernoulli* **6**, 1–24.

LEI, J. (2021). Network representation using graph root distributions. *Annals of Statistics* **49**, 745–768.

LEVIN, K. & LEVINA, E. (2019). Bootstrapping networks with latent space structure. *ArXiv e-prints* .

LI, G., JIANG, Y., JIAO, W., XU, W., HUANG, S., GAO, Z., ZHANG, J. & WANG, C. (2020). The maximum eigenvalue of the brain functional network adjacency matrix: Meaning and application in mental fatigue evaluation. *Brain Sciences* **10**.

LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. Providence,Rhode Island: American Mathematical Society Colloquium Publications.

MALLIAROS, F. D. & MEGALOOIKONOMOU, V. (2011). Expansion properties of large social graphs. In *DASFAA Workshops*.

MAUGIS, P. (2020). Central limit theorems for local network statistics. *ArXiv e-prints* .

MCGREGOR, A., VOROTNIKOVA, S. & VU, H. T. (2016). Better algorithms for counting triangles in data streams. In *Proceedings of PODS*.

MCGREGOR, J. K. A., PRICE, E. & VOROTNIKOVA, S. (2019). The complexity of counting cycles in the adjacency list streaming model. In *Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, PODS '19. New York, NY, USA: ACM.

MIEGHEM, P. V. (2010). *Graph Spectra for Complex Networks*. Cambridge University Press.

MUKHERJEE, S. S., SARKAR, P. & LIN, L. (2017). On clustering network-valued data. In *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett, eds., vol. 30. Curran Associates, Inc.

NAULET, Z., ROY, D. M., SHARMA, E. & VEITCH, V. (2021). Bootstrap estimators for the tail-index and for the count statistics of graphex processes. *Electron. J. Statist.* **15**, 282–325.

ORBANZ, P. (2017). Subsampling large graphs and invariance in networks. *ArXiv e-prints* .

O'ROURKE, S., VU, V. & WANG, K. (2018). Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra and its Applications* **540**, 26–59.

PERRY, A. & WEIN, A. S. (2017). A semidefinite program for unbalanced multisection in the stochastic block model. In *2017 International Conference on Sampling Theory and Applications (SampTA)*.

POLITIS, D. N., ROMANO, J. P. & WOLF, M. (1999). *Subsampling*. New York: Springer.

RASTELLI, R., FRIEL, N. & RAFTERY, A. E. (2016). Properties of latent variable network models. *Network Science* **4**, 407–432.

RESTREPO, J. G., OTT, E. & HUNT, B. R. (2005). Onset of synchronization in large networks of coupled oscillators. *Phys. Rev. E* **71**, 036151.

RUBIN-DELANCHY, P., PRIEBE, C. E., TANG, M. & CAPE, J. (2022). A statistical interpretation of spectral embedding: the generalised random dot product graph. *Journal of the Royal Statistical Society Series B (to appear)* .

Spielman, D. A. (2012). Lecture 3, Yale Computer Science: CPSC 662: Spectral Graph Theory. URL: https://www.cs.yale.edu/homes/spielman/561/2012/lect03-12.pdf.

Tang, M. (2018). The eigenvalues of stochastic blockmodel graphs.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V. & Priebe, C. E. (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli* **23**, 1599–1630.

Traud, A. L., Kelsic, E. D., Mucha, P. J. & Porter, M. A. (2012). Social structure of Facebook networks. *Physica A: Statistical Mechanics and its Applications* **391**, 4165–4180.

Tsourakakis, C. E., Kang, U., Miller, G. L. & Faloutsos, C. (2009). Doulion: Counting triangles in massive graphs with a coin. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Van Mieghem, P., Omic, J. & Kooij, R. (2009). Virus spread in networks. *IEEE/ACM Transactions on Networking* **17**, 1–14.

Veitch, V. & Roy, D. P. (2015). The class of random graphs arising from exchangeable random measures. *ArXiv e-prints* .

Veitch, V. & Roy, D. P. (2019). Sampling and estimation for (sparse) exchangeable random graphs. *Annals of Statistics* **47**, 3274–3299.

Vu, V. (2007). Spectral norm of random matrices. *Combinatorica* **27**, 721–736.

Wang, Y. R. & Bickel, P. (2017). Likelihood-based model selection for stochastic block models. *Annals of Statistics* **45**, 500–528.

Yan, B., Sarkar, P. & Cheng, X. (2018). Provable estimation of the number of blocks in block models. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, A. Storkey & F. Perez-Cruz, eds., vol. 84 of *Proceedings of Machine Learning Research*. PMLR.

Young, S. J. & Scheinerman, E. R. (2007). Random dot product graph models for social networks. In *Proceedings of the 5th international conference on algorithms and models for the web-graph*.

Zhang, Y. & Xia, D. (2022). Edgeworth expansions for network moments. *Annals of Statistics* **2**, 726–753.