On the Theoretical Properties of the Network Jackknife

Qiaohui Lin [1]   **Robert Lunde** [1]   **Purnamrita Sarkar** [1]

Abstract

We study the properties of a leave-node-out jack-knife procedure for network data. Under the sparse graphon model, we prove an Efron-Stein-type inequality, showing that the network jack-knife leads to conservative estimates of the variance (in expectation) for any network functional that is invariant to node permutation. For a general class of count functionals, we also establish consistency of the network jackknife. We complement our theoretical analysis with a range of simulated and real-data examples and show that the network jackknife offers competitive performance in cases where other resampling methods are known to be valid. In fact, for several network statistics, we see that the jackknife provides more accurate inferences compared to related methods such as subsampling.

1. Introduction

Network-structured data are now everywhere. The Internet is a giant, directed network of webpages pointing to other webpages. Facebook is an undirected network built via friendships between users. The ecological web is a directed network of different species with edges specified by 'who-eats-whom' relationships. Protein-protein interactions are undirected networks consisting of pairs of bait-prey proteins that bind to each other during coaffinity purification experiments arising in mass spectrometry analysis.

In these application areas, it is often of interest to characterize a network using statistics such as the clustering coefficient, triangle density, or principal eigenvalues. There has been a substantial amount of work on approximating these quantities with small error on massive networks (Feige, 2006; Goldreich & Ron, 2008; Assadi et al., 2018; Eden

et al., 2017; Gonen et al., 2010; Kallaugher et al., 2019). However, comparatively little attention has been paid to assessing the variability of these statistics with a few exceptions that we will discuss shortly. Quantifying the uncertainty of these estimators is of utmost importance, as it gives us information about the underlying variability of the data generating process. Take for example the problem of comparing two networks, which is a key question in many biological applications and in social network analysis. A natural approach would be to first obtain resamples of networks to construct distributions of different summary statistics and then compare these distributions. While there has been some recent interest in two-sample tests for networks (Kim et al., 2014; Durante & Dunson, 2018; Ghoshdastidar et al., 2017; Tang et al., 2017), very few works use resampling to compare networks. Uncertainty estimates from resampling also allow the construction of confidence intervals, which is a fundamental task in statistics and machine learning.

Resampling methods have a long and celebrated history in statistics, with the bootstrap, jackknife, and subsampling being the three main forms. There is a now vast literature developing these methods for iid data; see for example, (Quenouille, 1949; Efron & Tibshirani, 1986; Bickel et al., 1997; Politis et al., 1999; Shao & Wu, 1989). Even when the data are not independent, resampling methods have been shown to yield asymptotically valid inferences for a wide range of functionals under various dependence structures. For weakly dependent time series, for example, the key innovation is to resample contiguous blocks of data instead of individual observations. Under mild conditions on the block length and nature of dependence, blocked variants of resampling methods, including the block bootstrap (Künsch, 1989), block subsampling (Politis & Romano, 1994), and the blockwise jackknife (Künsch, 1989) have been shown to asymptotically capture the dependence structure of the data, leading to theories that closely resemble the corresponding theories for iid data.

Recently, some work has started to emerge involving resampling procedures for networks. Levin & Levina (2019) propose two bootstrap procedures for random dot product graphs that involve estimating the latent positions and resampling the estimated positions to conduct inference for the functional of interest. The authors establish bootstrap consistency for functionals that are expressible as U-statistics

---

[1]Department of Statistics and Data Sciences, University of Texas at Austin, TX, USA. Correspondence to: Qiaohui Lin <qiaohui.lin@utexas.edu>, Robert Lunde <rlunde@utexas.edu>, Purnamrita Sarkar <purna.sarkar@austin.utexas.edu>.

of the latent positions, which encompasses many important classes of functionals including subgraph counts.

Lunde & Sarkar (2019) consider a procedure that involves subsampling nodes and computing functionals on the induced subgraphs. This procedure is shown to be asymptotically valid under conditions analogous to the iid setting; that is, the subsample size must be $o(n)$ and the functional of interest must converge to a non-degenerate limit distribution. Previously, Bhattacharyya & Bickel (2015) had shown the validity of subsampling for count functionals. By proving a central limit theorem for eigenvalues, Lunde & Sarkar (2019) also establish subsampling validity for these functionals under certain conditions. Finally, in Green & Shalizi (2017), the authors propose sieve and nonparametric bootstrap procedures for networks.

We would like to note that that both the sieve approach of Green & Shalizi (2017) and the latent position estimation approaches of Levin & Levina (2019) depend on accurately estimating the underlying graphon. The nonparametric bootstrap procedure described in Green & Shalizi (2017) requires resampling much larger networks from a size n network, leading to computational inefficiency. Even subsampling requires weak convergence and a known rate of convergence; it turns out that the latter may be estimated (Bertail et al., 1999), but doing so entails a substantial increase in computation and is likely to adversely affect the finite-sample performance of the procedure. While asymptotically valid under general conditions, the finite-sample performance of subsampling methods is known to be sensitive to the choice of tuning parameters; see for example, Kleiner et al. (2014).

1.1. Our Contribution

In the present work, we study the properties of a network jackknife introduced by Frank & Snijders (1994) under the sparse graphon model. On the theoretical side, we make two primary contributions. First, analogous to the iid setting, we show that the network jackknife produces variance estimates that are conservative in expectation under general conditions. Our result here justifies the network jackknife as a rough-and-ready tool that produces reasonable answers (erring on the side of caution) even when the asymptotic properties of the functional of interest are poorly understood.

While the upward bias of the network jackknife is a favorable property, it does not provide information as to how the jackknife compares to other resampling procedures for more well-understood functionals. As another theoretical contribution, we establish consistency of the jackknife for a general class of count statistics studied in Bickel et al. (2011). We also extend our result to smooth functions of counts, which encompasses widely used measures such as the transitivity coefficient.

We complement our theoretical results with an empirical investigation of the network jackknife on both simulated and real datasets. In our simulation study, we study the rate of convergence of the jackknife variance estimate for two sparse graphon models under a range of choices for the network functional. Our results suggest that by and large, the jackknife has better finite-sample properties than subsampling. For real data, we conduct network comparisons of Facebook networks constructed from a number of different colleges such as Caltech, Berkeley, Stanford, Wellesley, etc.

The paper is organized as follows. In Section 2, we do problem setup and introduce notation. In Section 3, we present our theoretical results and some proof sketches. Finally in Section 4 we present experimental results on simulated and real networks.

2. Background

We first briefly recall the original Jackknife for iid data, followed by a description of the sparse graphon model. We conclude this section with the network jackknife procedure.

2.1. The Jackknife for IID Data

The jackknife, attributed to Quenouille (1949) and Tukey (1958), is a resampling procedure that involves aggregating leave-one-out estimates. More precisely, let $X_1, \ldots, X_n \sim P$ and let S_n be a permutation-invariant function of n variables. Furthermore, let $S_{n,i}$ denote the functional computed on the dataset with X_i removed and let $\bar{S}_n = \frac{1}{n} \sum_{i=1}^{n} S_{n,i}$. The jackknife estimate of the variance of $S_{n-1} = S(X_1, \ldots, X_{n-1})$ is given by:

$$\widehat{\mathrm{Var}}_{\mathrm{JACK}} \, S_{n-1} := \sum_{i=1}^{n} (S_{n,i} - \bar{S}_n)^2 \qquad (1)$$

For appropriately smooth functionals, it is well-known that the jackknife consistently estimates the variance; see for example, Shao & Tu (1995). The bootstrap, introduced by Efron (1979), typically requires weaker regularity conditions than the jackknife for consistency. In fact, it is well-known that the jackknife is inconsistent for the median (Miller, 1974) while the bootstrap variance remains consistent under reasonable conditions (Ghosh et al., 1984)[1].

However, for more complicated functionals, it may often be the case that both the bootstrap and the jackknife are inconsistent[2]. Even in these cases, the jackknife still provides reasonable answers. The remarkable inequality of Efron &

---

[1]It should be noted that the delete-d jackknife is valid under more general conditions; see Shao & Wu (1989).

[2]Recent work by Fang & Santos (2019) suggests that Hadamard differentiability of g is both necessary and sufficient for bootstrap consistency of $g(\hat{\theta}_n)$ whenever $\hat{\theta}_n$ is asymptotically Gaussian.

Stein (1981) asserts that the jackknife is always upwardly biased, ensuring a conservative estimate of the variance.

Since networks are inherently high-dimensional objects, asymptotic results are often harder to come by compared to the iid setting. The theory of the jackknife for iid processes suggests that, even in this challenging regime, a network analogue of the jackknife may have some advantageous properties. Before delving into our method, we introduce the network model under consideration below.

2.2. The Sparse Graphon Model

The model below is due to Bickel & Chen (2009). Let $\{A^{(n)}\}_{n \in \mathbb{N}}$ denote a sequence of $n \times n$ adjacency matrices and $\xi_1, \xi_2, \dots \xi_n \sim \text{Uniform}[0,1]$. Furthermore, for $i \neq j$, let $\eta_{ij} \sim \text{Uniform}[0,1]$ and define $A_{ij}^{(n)}$ as follows:

$$
\begin{aligned}
A_{ij}^{(n)} = A_{ji}^{(n)} &= 1(\eta_{ij} \leq \rho_n w(\xi_i, \xi_j) \wedge 1) \\
&\stackrel{d}{=} \text{Bernoulli}(\rho_n w(\xi_i, \xi_j) \wedge 1)
\end{aligned}
\tag{2}
$$

where $w : [0,1]^2 \mapsto \mathbb{R}_{\geq 0}$ is a symmetric measurable function that satisfies $\int_0^1 \int_0^1 w(u,v) \, du \, dv = 1$ and $\rho = \{\rho_n\}_{n \in \mathbb{N}} \in [0,1]^{\mathbb{N}}$. When $\|\rho_n w(u,v)\|_{\infty} \leq 1$, ρ_n may be interpreted as the probability of an edge and w as the conditional density of the latent positions given an edge. For $i = j$, we will assume that $A_{ii}^{(n)} = 0$. We refer to w as a graphon, the pair (ρ, w) as a sparse graphon, and (2) as graphs generated by the sparse graphon model.

Bounded graphons arise as a limiting object in the theory of graph limits; see Lovász (2012). Alternatively, graphons are a natural representation for (infinite-dimensional) jointly exchangeable arrays, where this notion of exchangeability corresponds to invariance under vertex permutation; see for example, Diaconis & Janson (2008). Bounded graphons subsume many other commonly studied network models, including stochastic block models (Holland et al., 1983) and random dot product graphs (Young & Scheinerman, 2007).

For applications, one major issue with graphons is that they generate dense graphs, where the expected number of edges is $O(n^2)$. However, most real-world graphs are known to be sparse in the sense that the number of edges is $o(n^2)$. Here, the sequence ρ determines the sparsity level; by letting $\rho_n \to 0$ at an appropriate rate, we may generate a graph sequence with the desired sparsity properties. Furthermore, notice that in our formulation above, w is allowed to be unbounded. As noted in Borgs et al. (2019), unbounded graphons are more expressive than bounded graphons, allowing graphs that exhibit power law degree distributions.

2.3. The Network Jackknife Procedure

Let $f : \{0,1\}^{n-1 \times n-1} \mapsto \mathbb{R}$ denote a function that takes as input a $n-1 \times n-1$ adjacency matrix and let $Z_{n,i}$ denote the random variable formed by applying f to an induced subgraph with node i removed. Under the model (2), observe that each induced subgraph formed by leaving a node out is identically distributed as a consequence of vertex exchangeability. Therefore, functionals calculated on these induced subgraphs are similar in spirit to the the leave-one-out estimates for the jackknife in the iid setting. Following Frank & Snijders (1994), a natural generalization of the jackknife to the sparse graphon setting is given by:

$$
\widehat{\text{Var}}_{\text{JACK}} \, Z_{n-1} := \sum_{i=1}^{n} (Z_{n,i} - \bar{Z}_n)^2
\tag{3}
$$

where $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^{n} Z_{n,i}$ and $\widehat{\text{Var}}_{\text{JACK}} \, Z_{n-1}$ is an estimate of $\text{Var} \, Z_{n-1}$, the variance with respect to an induced subgraph with node set $\{1, \dots, n-1\}$. We would like to note that letting $Z_{n-1} := Z_{n,n}$ constitutes a slight abuse of notation since ρ_{n-1} need not equal ρ_n, but doing so substantially improves the readability of our proofs.

3. Theoretical Results

3.1. The Network Efron-Stein Inequality

The first result we state here is our generalization of the Efron-Stein inequality to the network setting. Intuitively, the Efron-Stein inequality may be thought of as a general property of functions of independent random variables. While edges in the adjacency matrix are dependent through the latent positions, the fact that they are functions of independent random variables allow us to prove the following:

Theorem 1 (Network Efron-Stein Inequality)**.**

$$
\text{Var} \, Z_{n-1} \leq E(\widehat{\text{Var}}_{\text{JACK}} \, Z_{n-1})
\tag{4}
$$

The main ingredients in our proof are an adaptation of a martingale argument due to Rhee & Talagrand (1986) and an appropriate filtration for graphon models inspired by Borgs et al. (2008). We provide a proof sketch below; for details, see Supplement Section A.

Proof Sketch. As discussed in the Supplementary Material, for $1 \leq i \leq n$, we may express $Z_{n,i}$ as a measurable function of latent positions $\xi_i \sim \text{Unif}[0,1]$ for $1 \leq i \leq n$ and $\eta_{ij} \sim \text{Unif}[0,1]$ for $1 \leq i < j \leq n$. More precisely, $Z_{n,i}$ is a function of the variables that are not shaded below:

$$
Z_{n,i} = g \begin{pmatrix} \xi_1 & \eta_{12} & \eta_{13} & \dots & \eta_{1i} & \dots & \eta_{1n} \\ & \xi_2 & \eta_{23} & \dots & \eta_{2i} & \dots & \eta_{2n} \\ & & \xi_3 & \dots & \eta_{3i} & \dots & \eta_{3n} \\ & & & \dots & \dots & \dots & \dots \\ & & & & \xi_i & \eta_{i,i+1} & .. & \eta_{in} \\ & & & & & \xi_{n-1} & \eta_{n-1,n} \\ & & & & & & \xi_n \end{pmatrix}.
\tag{5}
$$

We design a martingale difference sequence d_i,

$$d_i = E(Z_{n-1}|\Sigma_i) - E(Z_{n-1}|\Sigma_{i-1}), \qquad (6)$$

based on filtration Σ_i:

$$\Sigma_i = \sigma\{\xi_1,,\xi_i,\eta_{12},,\eta_{1i},\eta_{23},,\eta_{2i},,,\eta_{i-1,i}\}$$

$$= \sigma \left\{ \begin{array}{ccccc} \xi_1 & \eta_{12} & ... & \eta_{1,i-1} & \eta_{1i} \\ & \xi_2 & ... & \eta_{2,i-1} & \eta_{2i} \\ & & ... & .. & \\ & & \xi_{i-2} & \eta_{i-2,i-1}, & \eta_{i-2,i} \\ & & & \xi_{i-1} & \eta_{i-1,i} \\ & & & & \xi_i \end{array} \right\}. \qquad (7)$$

Then we can show that,

$$\text{Var } Z_{n-1} = \sum_{i=1}^{n-1} E d_i^2.$$

On the other hand, the expectation of Jackknife estimate is:

$$E \sum_{i=1}^{n} (Z_{n,i} - \bar{Z}_n)^2 = (n-1) \frac{E(Z_{n,1} - Z_{n,2})^2}{2}. \qquad (8)$$

Now, we construct another filtration $\mathcal{A}$ such that $E(Z_{n,1}|\mathcal{A}) = E(Z_{n,2}|\mathcal{A})$. In particular, we use:

$$\mathcal{A} = \sigma\{\xi_3, \ldots, \xi_{i+1}, \eta_{34}, \ldots, \eta_{3,i+1}, \ldots, \eta_{i,i+1}\}. \qquad (9)$$

This is essentially Σ_{i+1}, with the first and second row and columns removed. Define

$$U = E(Z_{n,1}|\Sigma_{i+1}) - E(Z_{n,1}|\mathcal{A})$$
$$V = E(Z_{n,2}|\Sigma_{i+1}) - E(Z_{n,2}|\mathcal{A})$$

Using the fact that $E(X^2) = E(E(X^2|\Sigma)) \geq E(E[X|\Sigma]^2)$ for some random variable X which is measurable w.r.t to some Sigma field Σ, we get:

$$E(Z_{n,1} - Z_{n,2})^2 \geq E(U - V)^2 = 2E(d_i^2) \qquad (10)$$

The result follows from plugging in Eq 10 to Eq 8. $\qquad \square$

Remark 1. *Using the aforementioned filtration for graphon models, is also possible to prove another network variant of the Efron-Stein inequality following arguments in Boucheron et al. (2004). This alternative procedure does not require the functional to be invariant to node permutation and allows flexibility with the leave-one-out estimates. However, the resulting estimate is often not sharp. See the Supplement Section F for more details.*

3.2. Beyond the Efron-Stein inequality

While the Efron Stein inequality in Theorem 1 is surprising and useful for estimating uncertainty for network statistics, it would be much more satisfying if indeed the jackknife estimate of variance in fact coincided with the true underlying variance, at least asymptotically. We want to draw the attention of the reader to leftmost panel in Figure 1. The solid black line shows the mean and standard error of the ratio between the jackknife estimate of the variance and the true variance for edge density for a blockmodel and a smooth graphon (details in Section 4), as graph size grows. This figure shows the surprising trend that, in fact, the jackknife estimate is not only an upper bound on the true variance; it is in fact asymptotically unbiased. Our next proposition establishes exactly that. In what follows, let Z_n denote the edge density (see Section 4).

Proposition 1. *Suppose that $\int_0^1 \int_0^1 w^2(u,v) \, du \, dv < \infty$ and $n\rho_n \to \infty$. Let $\sigma^2 = \lim_{n\to\infty} n \cdot \text{Var}(Z_{n-1})$. Then,*

$$n \cdot E(\widehat{\text{Var}}_{\text{JACK}} Z_{n-1}) \to \sigma^2 \qquad (11)$$

The proof of the above result involves tedious combinatorial arguments and is deferred to the Supplement Section D. The above proposition says that, the jackknife estimate of variance of the edge density of a sparse graphon model (see Eq 2), in expectation, converges to the true variance. This is a somewhat weak result, since it does not say anything about the jackknife estimate obtained from one network. However, it begs the question, whether a stronger result is true. In fact, in the next section, we prove that for a broad class of count functionals, the jackknife estimate is in fact consistent. This paves the way to the next section, which we start by introducing count functionals.

3.3. Jackknife Consistency for Count Functionals

In this section, we study the properties of the jackknife for subgraph counts, which are an important class of functionals in network analysis. In graph limit theory, convergence of a sequence of graphs can be defined as the convergence of appropriate subgraph frequencies (Lovász, 2012). More practically, subgraph counts have been used to successfully conduct two-sample tests in various settings. In social networks, for example, the frequency of triangles provides information about the likelihood of mutual friendships/connections and is therefore a useful summary statistic.

The count functionals that we consider were first studied in Bickel et al. (2011). We will now introduce some notation needed to define these functionals. Let G_n denote a graph with vertex set $V(G_n) = \{1, 2, \ldots, n\}$ and edge set $E(G_n) \subset V(G_n) \times V(G_n)$. Let $R \subseteq E(G_n)$ denote the subgraph of interest parameterized by its edge set. For convenience, we assume $V(R) = \{1, 2, \ldots, p\}$. Furthermore,

let $G_n[R]$ denote the subgraph induces by the vertices of R.

We consider two different types of count functionals. The first notion we consider counts exact matches and has the following probability under the sparse graphon model: (Eq 2):

$$P(R) = P(G_n[R] = R)$$

$$= E \left[\prod_{(i,j) \in R} \rho_n w(\xi_i, \xi_j) \wedge 1 \prod_{(i,j) \in \overline{R}} (1 - \rho_n w(\xi_i, \xi_j) \wedge 1) \right]$$
(12)

We also consider the functional $Q(R)$, which provides the probability of an induced subgraph containing the subgraph R:

$$Q(R) = P(R \subseteq G_n(R)) = E \left[\prod_{(i,j) \in R} \rho_n w(\xi_i, \xi_j) \right]$$
(13)

Note that $Q(R)$ is agnostic to the presence of additional edges. When the graph sequence is sparse, $P(R)$ and $Q(R)$ are uninformative, as $P(R)$, $Q(R) \to 0$. Let $e = |E(R)|$ and $p = |V(R)|$. Instead, define the following normalized subgraph frequency:

$$\tilde{P}(R) = \rho_n^{-e} P(R) \qquad \tilde{Q}(R) = \rho_n^{-e} Q(R) \qquad (14)$$

Furthermore, let $Iso(R)$ denote the class of graphs isomorphic to R, and $|Iso(R)|$ its cardinality. Our estimator of $\tilde{P}(R)$ is given by:

$$\hat{P}(R) = \rho_n^{-e} \frac{1}{\binom{n}{p} |\text{Iso}(R)|} \sum_{S \sim R} 1(S = G_n[S]) \qquad (15)$$

Similarly, define $\hat{Q}(R)$ as:

$$\hat{Q}(R) = \rho_n^{-e} \frac{1}{\binom{n}{p} |\text{Iso}(R)|} \sum_{S \sim R} 1(S \subseteq G_n[S]) \qquad (16)$$

Due to magnification by ρ_n^{-e}, (14), (15), and (16) are not necessarily upper bounded by 1; nevertheless, they are still meaningful quantities related to subgraph frequencies.

Bickel et al. (2011) establish a central limit theorem for these functionals under general conditions on the sparsity level and structure of the subgraph. Under analogous conditions, we establish consistency of the network jackknife for counts and smooth functions of counts.

Theorem 2 (Jackknife Consistency for Counts). *Suppose that R is a p-cycle or acyclic graph. Furthermore, suppose that $\int_0^1 \int_0^1 w^{2e}(u,v) \, du \, dv < \infty$ and $n\rho_n \to \infty$. Let $\sigma^2 = \lim_{n \to \infty} n \cdot \text{Var} \, \hat{P}(R)$. Then,*

$$n \cdot \widehat{\text{Var}}_{\text{JACK}} \, \hat{P}(R) \xrightarrow{P} \sigma^2 \qquad (17)$$

Our proof relies on a signal-noise decomposition of the jackknife variance. Bickel et al. (2011) establish that the variance of a count functional is largely driven by the variance of a U-statistic related to the edge structure of the subgraph. For this U-statistic component of the decomposition, results for jackknifing U-statistics due to Arvesen (1969) may be used to show convergence in probability towards the variance of the corresponding U-statistic. Since the jackknife is a sum of squares, we are able to decouple the effects of a remainder term and show that it is negligible. We provide a sketch below, and defer the details to Supplement Section B.

Proof Sketch. Define density (normalized counts) of R when leaving ith node out is $Z_{n,i} = \binom{n-1}{p}^{-1} \rho_n^{-e}(T - T_i)$, where T is total counts of R in G_n, p and e are number of vertices and edges in R. Define $Z_n = \binom{n}{p}^{-1} \rho_n^{-e} T$. Then $\widehat{\text{Var}}_{\text{JACK}} = \sum_i (Z_{n,i} - \overline{Z_n})^2$, $\text{Var} \, \hat{P}(R) = \text{Var} \, Z_n$.

Theorem 1 of Bickel et al. (2011) establishes that $n\text{Var}(Z_n)$ converges to a positive constant. Thus we scale $\widehat{\text{Var}}_{\text{JACK}}$ by n, and decompose $n\widehat{\text{Var}}_{\text{JACK}}$ into

$$n \left[\sum_i (\alpha_i - \bar{\alpha}_n)^2 - 2 \sum_i (\alpha_i - \bar{\alpha}_n)(\beta_i - \bar{\beta}_n) + \sum_i (\beta_i - \bar{\beta}_n)^2 \right],$$
(18)

where $\alpha_i = Z_{n,i} - E(Z_{n,i}|\boldsymbol{\xi}_n)$, $\beta_i = E(Z_{n,i}|\boldsymbol{\xi}_n)$, $\bar{\alpha}_n = \frac{1}{n}\sum_{i=1}^n \alpha_i$, $\bar{\beta}_n = \frac{1}{n}\sum_{i=1}^n \beta_i$ and $\boldsymbol{\xi}_n = (\xi_1, \dots \xi_n)$. The term $n \sum_i (\beta_i - \bar{\beta}_n)^2$ corresponds to the signal component discussed before the theorem statement.

We show in the Supplement that $E \sum_i (\alpha_i - \bar{\alpha}_n)^2$ can be further written as $(1 - o(1)) \cdot E \left[\text{Var}(\hat{P}(R) \mid \boldsymbol{\xi}_n) \right] = o(1/n)$, by Theorem 1 of Bickel et al. (2011). $n \sum_i (\alpha_i - \bar{\alpha}_n)^2$ is thus negligible by Markov Inequality. The cross term $n \sum_i (\alpha_i - \bar{\alpha}_n)(\beta_i - \bar{\beta}_n)$ is also negligible by the Cauchy-Schwartz Inequality.

$\square$

Remark 2. *Our theoretical results hold for both notions of subgraph frequencies. However, note that $\tilde{Q}(R)$ is independent of n, but $\tilde{P}(R)$ depends on n and approaches $\tilde{Q}(R)$. While $\sqrt{n}[\hat{P}(R) - \tilde{P}(R)]$ and $\sqrt{n}[\hat{Q}(R) - \tilde{Q}(R)]$ have the same limiting variance, inference for a fixed target using $\hat{P}(R)$ requires stronger sparsity conditions; namely $\rho_n = \omega(1/\sqrt{n})$. See Section 3.4 for a related discussion.*

Remark 3. *Central limit theorems and jackknife consistency can also be shown for more general (cyclic) graphs. However, in these cases, more stringent sparsity conditions are needed.*

Now, let $f(G_n)$ denote a function of the vector $(\hat{P}(R_1), \dots, \hat{P}(R_d))$. Furthermore, let ∇f denote the gradient of f and $\mu \in \mathbb{R}^d$ the limit of $(\tilde{P}(R_1), \dots, \tilde{P}(R_d))$

as $n \to \infty$; it turns out that μ corresponds to an integral parameter of the graphon related to the edge structure of the subgraph. We have the following result.

Theorem 3 (Jackknife Consistency for Smooth Functions of Counts). *Suppose that $(R_1, \ldots, R_d)$ are p-cycles or acyclic graphs and $n\rho_n \to \infty$. Let $e^* = \max\{|E(R_1), \ldots E(R_d)\}$ and suppose that $\int_0^1 \int_0^1 w^{2e^*}(u, v)\, du\, dv < \infty$. Furthermore, suppose that ∇f exists in a neighborhood of μ, $\nabla f(\mu) \neq 0$, and that ∇f is continuous at μ. Let σ_f^2 denote the asymptotic variance of $\sqrt{n}[f(G_n) - f(E(G_n))]$. Then,*

$$n \cdot \widehat{\mathrm{Var}}_{\mathrm{JACK}}\, f(G_n) \xrightarrow{P} \sigma_f^2$$

Proof Sketch. Let $Z_{n,i} = (Z_{n,i}(1), \ldots Z_{n,i}(d))$, where d is a constant w.r.t n and each entry corresponds to a count functional with node i removed. Let $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_{n,i}$. We use a Taylor expansion of $f(Z_{n,i})$ around $\bar{Z}_n$.

$$f(Z_{n,i}) = f(\bar{Z}_n) + \nabla f(\mu)^T (Z_{n,i} - \bar{Z}_n)$$
$$+ \underbrace{(\nabla f(\zeta_i) - \nabla f(\mu))^T (Z_{n,i} - \bar{Z}_n)}_{E_i},$$

where $\zeta_i = (\zeta_{i1}, \ldots, \zeta_{id}) = c_i Z_{n,i} + (1 - c_i)\bar{Z}_n$ for some $c \in [0, 1]$. Thus, we also have:

$$f(Z_{n,i}) - \overline{f(Z_{n,i})} = \underbrace{\nabla f(\mu)^T (Z_{n,i} - \bar{Z}_n)}_{I_i} + \underbrace{E_i - \frac{1}{n}\sum_i E_i}_{II_i}$$

(19)

We bound $n\sum_i (I_i)^2$ and $n\sum_i (II_i)^2$ separately. Let Σ denotes the covariance matrix of a multivariate U-statistic with kernels $(h_1, \ldots, h_d)$, where each h_j is the kernel corresponding to the count functional in the jth coordinate of the vector Z_n (see Eq A.7 for detailed definition). We can show that

$$n\sum_i (I_i)^2 - \nabla f(\mu)^T \Sigma \nabla f(\mu) = o_p(1).$$

We can also show that $n\sum_i (II_i)^2$ is also $o_p(1)$. Then, let $\mu_n = E[Z_n]$. Note that if one counts subgraphs by an exact match as in (Bickel et al., 2011) $\mu_n \to \mu$. If one counts subgraphs via edge matching, $\mu_n = \mu$. Thus, both these types of subgraph densities, which asymptotically have the same limit, can be handled by our theoretical results. By Theorem 3.8 in Van der Vaart (2000),

$$\sqrt{n}(f(Z_n) - f(\mu_n)) \quad N(0, \nabla f(\mu)^T \Sigma \nabla f(\mu))$$

This shows that the jackknife estimate of variance converges to the asymptotic variance of $f(Z_n)$. The proof is deferred to the Supplement Section C. $\qquad\square$

3.4. A Remark on the use of the Network Jackknife for Two-Sample Testing

In principle, the jackknife variance provides a quantification of uncertainty that may be used for many inference tasks. When the limiting distribution is Normal, one may use a Normal approximation; otherwise, one may use Chebychev's inequality. However, in these cases, the natural centering is $\theta_{n-1} = E(Z_{n-1})$, which depends on n. Inferences about θ_{n-1} are often useful for a single graph, but for two-sample testing, issues may arise when comparing networks of different sizes. Probability statements involving a fixed population parameter θ are needed. To ensure that the jackknife yields valid inferences for an appropriate population parameter, we will need to impose some additional assumptions. In what follows, let $\{\tau_n\}_{n\in\mathbb{N}}$ denote a sequence of normalizing constants and let $U_{n-1} = \hat{\theta}_{n-1} - \theta$ for some $\theta \in \mathbb{R}$. We have the following result; see Supplement Section E for details.

Proposition 2. *Suppose that $\tau_n \to \infty$ and $\tau_n U_n \quad U$ for some non-degenerate U with mean 0 and variance σ^2 and $\{(\tau_n U_n)^2\}_{n\in\mathbb{N}}$ is uniformly integrable. Then,*

$$\tau_n(\hat{\theta}_n - E(\hat{\theta}_n)) \quad U,\ \mathrm{Var}\, U_n \to \mathrm{Var}\, U \qquad (20)$$

As a consequence of Proposition 2, if a central limit theorem is known for $\sqrt{n}\, U_n$ and a uniform integrability condition is satisfied, then one may use the jackknife variance in conjunction with a Normal approximation to conduct (possibly conservative) inference for θ. Asymptotically, these inferences are the same as those centered at θ_{n-1}. For count functionals, we have mentioned when this condition holds, so in this case it does not need to be checked.

4. Experiments

In this section, we present simulation experiments and experiments on real data. For simulations, we compare our variance estimate with that estimated using subsampling. We present our results on two graphons. For real data, we compare networks based on C.I.'s constructed using jackknife estimates of variance of network statistics like edge or triangle density and normalized transitivity.

Count functionals used: In this section we consider the edge, triangle, two star density, and normalized transitivity, which we explicitly define below.

$$\text{Edge density} := \frac{\sum_{i<j} A_{ij}}{\binom{n}{2}\rho_n}$$

$$\text{Triangle density} := \frac{\sum_{i<j<k} A_{ij}A_{jk}A_{ki}}{\binom{n}{3}\rho_n^3}$$
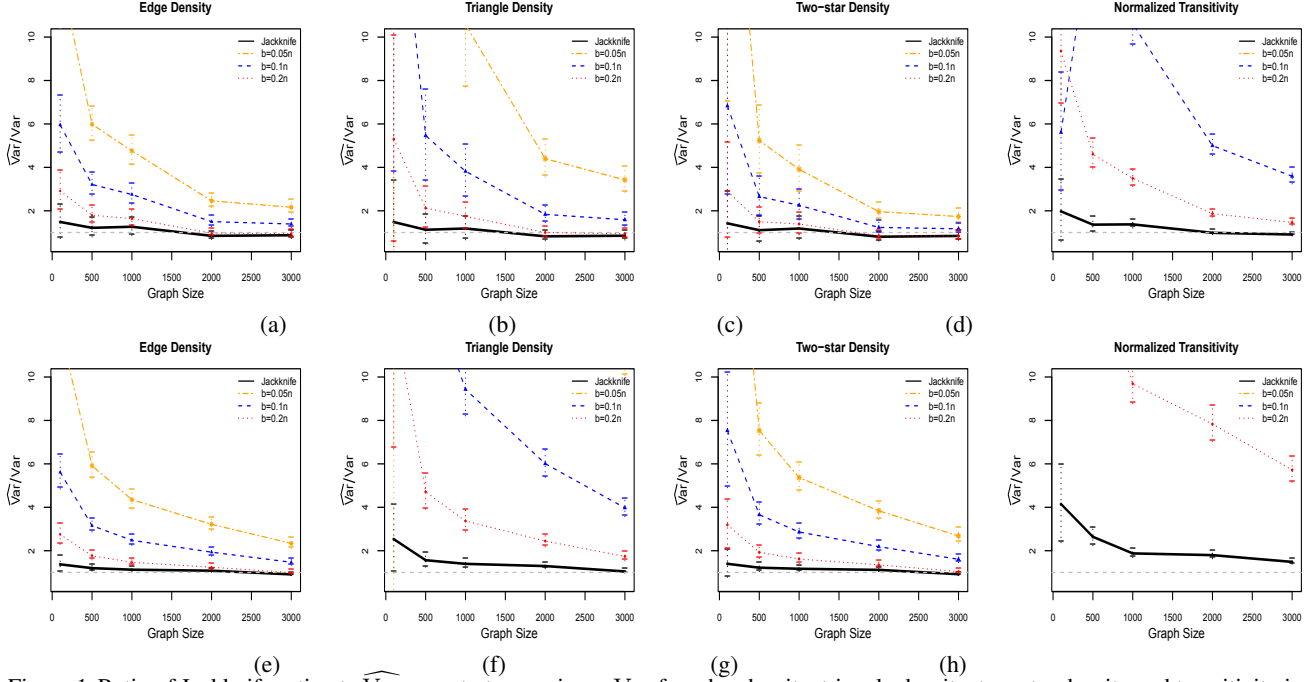
Figure 1. Ratio of Jackknife estimate $\widehat{\text{Var}}_{\text{JACK}}$ to true variance Var for edge density, triangle density, two-star density and transitivity in size $n = 100, 500, 1000, 2000, 3000$ graphs simulated from the SBM (top) and GR2 (bottom), compared to subsampling with $b = 0.05n$, $b = 0.1n$, $b = 0.2n$ variance estimation on the same graphs.

$$\text{Two star density} := \frac{\sum_{i,j<k,j,k\neq i} A_{ij} A_{ik}}{\binom{n}{3} \rho_n^2}$$

As a smooth function of count statistics, we use:

$$\text{Normalized transitivity} := \frac{\text{Triangle density}}{\text{Two star density}}$$

4.1. Simulated Data

We simulate graphs from two different graphons. The first is a Stochastic Block Model (SBM) (Holland et al., 1983), which is a widely used model for networks with communities. A SBM is characterized by a binary cluster membership matrix $Z \in \{0,1\}^{n \times r}$, where r is the number of communities, and a community-community interaction matrix B. Conditioned on $Z_{ia} = 1$ and $Z_{jb} = 1$, nodes i and j form a link with probability B_{ab}. We use $B = ((0.4, 0.1, 0.1), (0.1, 0.5, 0.1), (0.1, 0.1, 0.7))$ and generate a Z from a Multinomial$(0.3, 0.3, 0.4)$.

For the other graphon, we consider the following parameterization:

$$h_n(u,v) = P(A_{ij} = 1 \mid \xi_i = u, \xi_j = v) = \nu_n |u - v|$$
$$\text{(GR2)}$$

where ν_n is a sparsity parameter. We use $\nu_n = n^{-1/3}$. We denote this graphon by GR2.

From these two graphons, we consider graph size n of $n = 100, 500, 1000, 2000, 3000$. For each n, we simulated 100

graphs to calculate the approximate true variance of edge density, triangle density, two-star density and normalized transitivity among these graphs.

Computation: For each simulated network, we remove one node at a time, recalculate a statistic $Z_{n,i}$ on the graph with $(n-1)$ nodes left. Next we compute the jackknife estimate of the variance $\widehat{\text{Var}}_{\text{JACK}} := \sum_i (Z_{n,i} - \bar{Z}_n)^2$, where $\bar{Z}_n$ is the average of the $Z_{n,i}$'s. It should be noted that for some statistics, jackknife, owing to its leave-one-node-out characteristic, can be implemented to reduce computation. For example, in calculating triangles, we calculate the number of triangle on the whole graph once and the number of triangles each node is involved in. This can be done by keeping track of the number of common neighbors between a node and its neighbors.

For each statistic mentioned, we report the mean of the ratio $\widehat{\text{Var}}_{\text{JACK}}/\text{Var } Z_n$ among 100 graphs of each n and a 95% confidence interval from the standard deviation from a normal approximation of these 100 ratios. We also plot a dotted line to denote 1. Closer to this line a resampling procedure is, the better. In Figure 1, we plot this on the Y axis with n on the X axis. We also plot the same for subsampling with $b = 0.05n$, $b = 0.1n$, $b = 0.2n$ performed on the same graphs. Figure 1 a,b,c, and d contain results for the SBM, whereas the rest are for the smooth graphon.

We see that for both graphons, $\widehat{\text{Var}}_{\text{JACK}}/\text{Var}$ converges to 1 much more quickly in comparison to subsampling and

has much smaller variance. These figures also show how susceptible the performance of subsampling is to the choice of b. For $b = 0.05n$, subsampling overestimates the variance, and exceeds the upper bound on Y axis of some of the figures. In Figure 1 (h) we see that $\widehat{\text{Var}}_{\text{JACK}}$ for the normalized transitivity converges slowly for GR2, and subsampling with all choices of b are worse as well.

Eigenvalues: Here we examine the performance of jackknife on assessing the variance of eigenvalues, to which we have not yet extended our theoretical guarantees. In Figure 2 we show the $\widehat{\text{Var}}_{\text{JACK}}/\text{Var}$ for the two principal eigenvalues of the SBM ((a) and (b)) and two graphons described before. Here we only compared with subsampling with $b = 0.3n$ and $n = 1000, 2000, 3000$ as subsampling for eigenvalues in sparse graphs only works asymptotically for very large n (Lunde & Sarkar, 2019). For smaller n and b in our sparsity setting, we saw that subsamples of adjacency matrices often were too sparse leading to incorrect estimate of the variance.

Let us first look at Figure 2 (a) and (b) for the SBM setting. For both the eigenvalues in this case, $\widehat{\text{Var}}_{\text{JACK}}/\text{Var}$ converges to 1, whereas subsampling consistently underestimates the true variance. For graphon GR2, we see from Figure 2 (c) that both jackknife and subsampling estimate the true variance well, whereas for the second eigenvalue (see (d)) they both perform extremely poorly. These preliminary results of jackknife estimates show tentative evidence that our theory can be applied to statistics beyond count statistics, like eigenvalues, which we aim to investigate in future work.

4.2. Real-world Data

Below, we present two experiments using Facebook network data (Rossi & Ahmed, 2015); additional data results are discussed in Supplement Section G. In the first experiment, we compared three colleges: Caltech, Williams and Wellesley. While Caltech is known for its strength in natural sciences and engineering, Williams and Wellesley are strong liberal arts colleges. They all have relatively small number of students (800-3000), but with different demographics. For example, Wellesley is a women's liberal arts college, whereas the other have a mixed population. We present the 95% confidence intervals (CI) obtained using a normal approximation with the estimated variances for two-star and triangle densities for these networks.

We see that while all three have similar two-star density, Wellesley has significantly higher triangle density. We also see that CI's from jackknife and subsampling with $b = 0.1n$ and $b = 0.2n$ are comparable. Subsampling with $b = 0.05n$ tends have a wider CI, as it overestimates the variance. It is interesting to note that, for triangles, subsampling with $b = 0.2n$ took nearly 10 times as much time as jackknife, since

Table 1. Triangle, two-star density and normalized transitivity and their variances estimated in college training sets

College	Triangle		Two-star		Norm. Trans.	
	Est	$\widehat{Var}$	Est	$\widehat{Var}$	Est	$\widehat{Var}$
Berkeley	77.95	18.10	6.31	0.27	37.05	5.57
Stanford	36.62	5.12	5.90	0.11	18.61	0.16
Yale	24.20	2.40	5.22	0.09	13.90	0.06
Princeton	20.87	2.34	5.25	0.11	11.91	0.06
Harvard	38.56	5.11	6.28	0.10	18.43	0.10
MIT	30.20	7.89	6.11	0.24	14.82	0.15

we used the leave-one-node-out structure. In comparison, for both methods, two-star counting is overall much faster than counting triangles.

In the second experiment, we look at three college pairs: Berkeley and Stanford, Yale and Princeton, Harvard and MIT. First we decide which statistic differentiates between a given pair. For this, we split each college data set in half, into a training set and test set. On each of training set, we estimated their triangle density, two-star density, normalized transitivity and their variances estimated by jackknife, demonstrated in Table 1. Interestingly, in Table 1, the triangle density is large for all colleges, owing to the sparsity of the networks. From Table 1 we can see normalized transitivity estimates have relatively smaller variance and well separates each of the pairs in training sets. Thus we choose normalized transitivity as the test statistic. We now obtain jackknife estimate of variance of normalized transitivity using the the test sets.

Figure 4 presents 97.5% CI's for normalized transitivity for each college. Thus, two disjoint CI's are equivalent to rejecting a level 0.05 test. Figure 4 basically shows that transitivity can in fact separate Berkeley and Stanford Facebook networks, as well as Harvard and MIT Facebook networks, giving us interesting information about the inherent differences between the network structures of these colleges.

5. Discussion

In the present work, we have shown that the network jackknife is a versatile tool that may be used in a wide variety of situations. For poorly understood functionals, the Network Efron-Stein inequality ensures that the jackknife produces conservative estimates of the variance in expectation. For a general class of functionals related to counts, we establish consistency of the jackknife. Our empirical investigation is encouraging regarding the finite sample properties of the procedure, as the network jackknife outperforms subsampling in many simulation settings. Our simulations suggest that, for variance estimation of eigenvalues, on certain
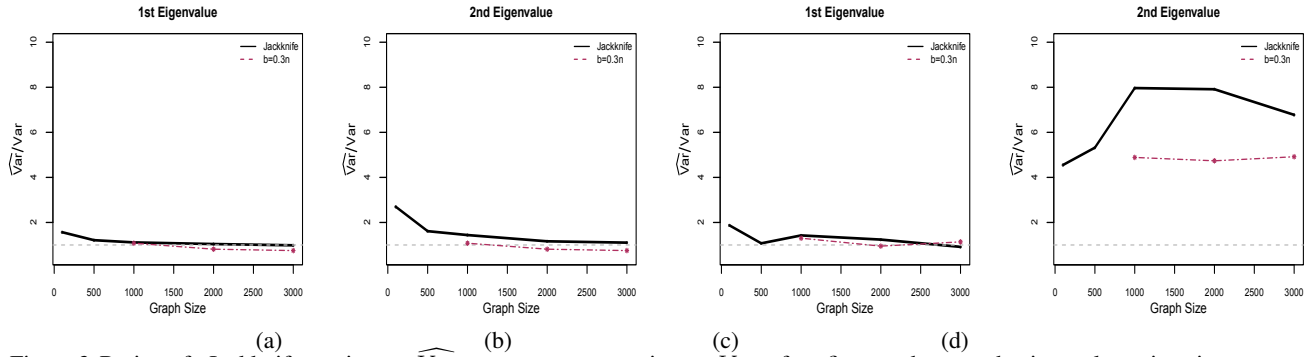
(a)   (b)   (c)   (d)

Figure 2. Ratio of Jackknife estimate $\widehat{\mathrm{Var}}_{\mathrm{JACK}}$ to true variance Var for first and second eigenvalues in size n = $100, 500, 1000, 2000, 3000$ graphs simulated from stochastic block model in (a) and (b) and the graphon GR2 in (c) and (d), compared to subsampling with $b = 0.3n$ variance estimation on the same graphs.
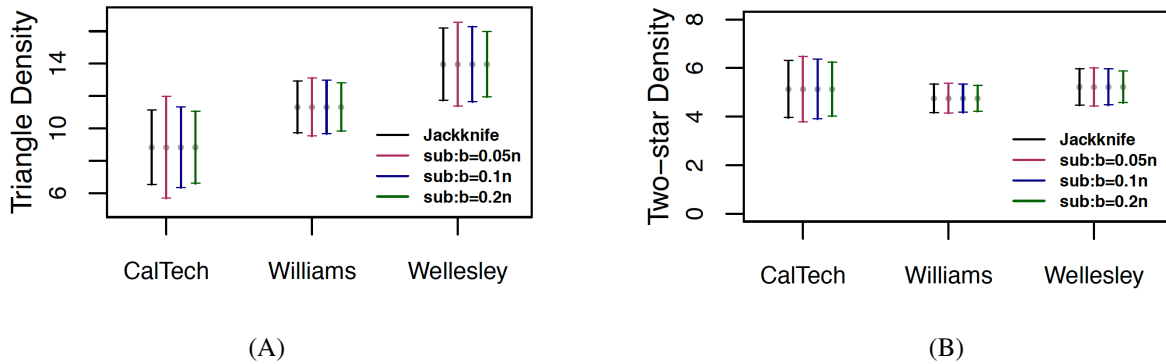

(A)

(B)

Figure 3. (A) Triangle density , and (B) two-star density (bottom) and their CI's based on jackknife and subsampling variance estimates.

Figure 4. For 3 pairs of colleges, 97.5% CI constructed using $\widehat{\mathrm{Var}}_{\mathrm{JACK}}$ on normalized transitivity

graphons, the jackknife's performance is similar to subsampling. However we also see that, for some graphons, it is in fact worse than subsampling. We intend to investigate this further both theoretically and empirically.

6. Acknowledgements

References

Arvesen, J. N. Jackknifing U-statistics. *The Annals of Mathematical Statistics*, 40(6):2076–2100, 1969.

Assadi, S., Kapralov, M., and Khanna, S. A simple sublinear-time algorithm for counting arbitrary subgraphs via edge sampling. *ArXiv e-prints*, 2018.

Bertail, P., Politis, D. N., and Romano, J. P. On subsampling estimators with unknown rate of convergence. *Journal of the American Statistical Association*, 94(446):569–579, 1999.

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

Bickel, P. J. and Chen, A. A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences (USA)*, 106:21068–21073, 2009.

Bickel, P. J., Götze, F., and van Zwet, W. Resampling fewer than n obersvations: Gains, losses, and remedies for losses. *Statistica Sinica*, pp. 1–31, 1997.

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

Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., and Veszter-gombi, K. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801 – 1851, 2008.

Borgs, C., Chayes, J. T., Cohn, H., and Zhao, Y. 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(5): 3019–3062, 2019.

Boucheron, S., Lugosi, G., and Bousquet, O. Concentration inequalities. In Bousquet, O., von Luxburg, U., and Rätsch, G. (eds.), *Advanced Lectures in Machine Learning*, pp. 208âĂŞ40. Springer, New York, 2004.

Diaconis, P. and Janson, S. Graph limits and exchangeable random graphs. *Rend. Mat. Appl.*, 7(28):33–61, 2008.

Durante, D. and Dunson, D. B. Bayesian inference and testing of group differences in brain networks. *Bayesian Anal.*, 13(1):29–58, 03 2018. doi: 10.1214/16-BA1030.

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

Efron, B. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(3):1–26, 1979.

Efron, B. and Stein, C. The jackknife estimate of variance. *Annals of Statistics*, 9(3):586–596, 1981.

Efron, B. and Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.*, 1(1):54–75, 02 1986. doi: 10.1214/ss/1177013815.

Fang, Z. and Santos, A. Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1): 377–412, 2019.

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

Frank, O. and Snijders, T. Estimating the size of hidden populations using snowball sampling. *Journal of Official Statistics*, 10(1):53 – 67, 1994.

Ghosh, M., Parr, W. C., Singh, K., and Babu, G. J. A note on bootstrapping the sample median. *The Annals of Statistics*, 12(3):1130–1135, 9 1984.

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

Goldreich, O. and Ron, D. Approximating average parameters of graphs. *Random Structures and Algorithms*, 32: 473–493, 2008.

Gonen, M., Ron, D., and Shavitt, Y. Counting stars and other small subgraphs in sublinear time. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, 2010.

Green, A. and Shalizi, C. Bootstrapping Exchangeable Random Graphs. *arXiv e-prints*, art. arXiv:1711.00813, Nov 2017.

Holland, P. W., Laskey, K., and Leinhardt, S. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109 – 137, 1983.

Kallaugher, J., McGregor, A., Price, E., and Vorotnikova, S. 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, pp. 119–133, New York, NY, USA, 2019. ACM.

Kim, J., Wozniak, J. R., Mueller, B. A., Shen, X., and Pan, W. Comparison of statistical tests for group differences in brain functional networks. *NeuroImage*, 101:681–694, 2014.

Kleiner, A., Talwalkar, A., Sarkar, P., and Jordan, M. I. A scalable bootstrap for massive data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76 (4):795–816, 2014.

Künsch, H. R. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17:1217–1241, 1989.

Levin, K. and Levina, E. Bootstrapping networks with latent space structure. *ArXiv e-prints*, 2019.

Lovász, L. *Large Networks and Graph Limits*. American Mathematical Society Colloquium Publications, Providence,Rhode Island, 2012.

Lunde, R. and Sarkar, P. Subsampling sparse graphons under minimal assumptions. *ArXiv e-prints*, 2019.

Miller, R. G. The jackknife - a review. *Biometrika*, 61(1): 1–15, 1974.

Politis, D. N. and Romano, J. P. Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, 22(4):2031–2050, 1994.

Politis, D. N., Romano, J. P., and Wolf, M. *Subsampling*. Springer, New York, 1999.

Quenouille, M. H. Problems in plane sampling. *The Annals of Mathematical Statistics*, 20(3):355–375, 1949.

Rhee, W. T. and Talagrand, M. Martingale inequalities and the jackknife estimate of variance. *Statistics & Probability Letters*, 4(1):5– 6, 1986.

Rossi, R. A. and Ahmed, N. K. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015. URL http://networkrepository.com.

Shao, J. and Tu, D. *The Jackknife and the Bootstrap*. Springer, 1995.

Shao, J. and Wu, C. J. A general theory for jackknife variance estimation. *Annals of Statistics*, 17(3):1176–1197, 1989.

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

Tukey, J. W. Bias and confidence in not quite large samples (abstract). *The Annals of Mathematical Statistics*, 29:614, 1958.

Van der Vaart, A. W. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Young, S. J. and Scheinerman, E. R. Random dot product graph models for social networks. In *Proceedings of the 5th international conference on algorithms and models for the web-graph*, pp. 138–149, 2007.