SUBSAMPLING BOOTSTRAP OF COUNT FEATURES OF NETWORKS

By Sharmodeep Bhattacharyya*† and Peter J. Bickel*†

University of California, Berkeley†

Abstract Analysis of stochastic models of networks is quite important in light of the huge influx of network data in social, information and bio sciences. But a proper statistical analysis of features of different stochastic models of networks is still underway. We propose bootstrap subsampling methods for finding empirical distribution of count features or ‘moments’ (Bickel, Chen & Levina, AoS, 2011) and smooth functions of these features for the networks. Using these methods, we can not only estimate the variance of count features but also get good estimates of such feature counts, which are usually expensive to compute numerically in large networks. In our paper, we prove theoretical properties of the bootstrap estimates of variance of the count features as well as show their efficacy through simulation. We also use the method on some real network data for estimation of variance and expectation of some count features.

1. Introduction. The study of networks have received recent increased attention not only in the social sciences, mathematics and statistics but also in physics and computer science. With the information boom, a huge number of network data sets have appeared. In biology - gene regulation networks, protein-protein interaction networks, neural networks, ecological and epidemiological networks have become increasingly important. In social media - Facebook, Twitter, Linkedin networks have come into prominence. Information networks have arisen in connection with text mining. Technological networks such as the Internet and many other networks related to internet have also become objects of study.

In this paper, we consider a nonparametric formulation for network models where node labels carry no information. The model was proposed in Bickel and Chen (2009) [5] and has its origins in the works of Aldous [1] and Hoover [16]. Exchangeable probability models on infinite networks have a general representation based on the results of Aldous [1], Hoover [16], Kallenberg [17], Diaconis and Janson [10]. The result is analogous to deFinetti’s.

*Supported in part by NSF grant DMS-1160319
AMS 2000 subject classifications: Primary 62F40, 62G09; secondary 62D05
Keywords and phrases: Networks, Subsampling, Bootstrap, Count features, Model-based sampling
theorem. Note that, numerical representation of networks come in form of adjacency matrix $A$, where $A_{ij} = 1$ if there is an edge from node $i$ to $j$ and 0 otherwise. We assume $A_{ii} = 0$, that is, there are no self-loops. It is natural to assume exchangeable property for probability distribution on unlabeled random networks, which means that the probability distribution on the set of all adjacency matrices $\mathcal{L}\{[A_{ij}], i, j \geq 1\}$ satisfy $\mathcal{L}\{[A_{ij}]\} = \mathcal{L}\{[A_{\sigma(i),\sigma(j)}]\}$, where $\sigma$ is an arbitrary permutation function on node indices. Such exchangeable probability distributions on random infinite binary arrays can be characterized as follows - for $i, j \geq 1$,

$$\alpha, \xi_i, \eta_{ij} \overset{iid}{\sim} U(0, 1)$$

$$A_{ij} = f(\alpha, \xi_i, \xi_j, \eta_{ij}),$$

where, $f : [0, 1]^4 \to [0, 1]$ is a measurable function symmetric in its second and third arguments and $\eta_{ij} = \eta_{ji}$. $\alpha$ as in de Finetti’s theorem corresponds to the mixing distribution and is not identifiable. This representation is not unique, and $f$ is not identifiable. These distributions can be parametrized through the function

$$h(u, v) = P[A_{ij} = 1|\xi_i = u, \xi_j = v].$$

The function $h$ is still not unique, but it can be shown that if two functions $h_1$ and $h_2$ define the same distribution $\mathcal{L}$, they can be related through a measure-preserving transformation. This leads to the Bickel and Chen (2009) [5] characterization of “nonparametric” unlabeled graph models which is closely related to Lovász’s notion of “graphons” [20]. The model will be described in more detail in Section 2. Other researchers have also studied similar general class of models such as latent space models of Hoff, Raftery and Handcock (2002) [14] and inhomogeneous random graph models of Bollobás, Janson and Riordan (2007) [8]. Many previously studied probability models for networks fall into this class. The class includes the stochastic block models (Holland, Laskey and Leinhardt (1983) [15], Nowicki and Snijders (2001) [24]) and the configuration model (Chung and Lu (2002) [9]). Dynamically defined models such as the “preferential attachment” models (which seem to have been first mentioned by Yule in the 1920s and given its modern name by Barabási and Albert (1999) [2]) can also be thought of in this way if the dynamical construction process continues forever producing an infinite graph. More details are given in Section 5.

Motifs or count statistics are the main statistics that we consider in this paper. Count statistics can be defined as smooth function of counts of subgraphs in the network. Counts of special subgraphs have been extensively
used in the network literature for analyzing network behavior [3] [23] [26].

The count statistics have appeared earlier under the names of “motif” counts
in biology [22] and “subgraph” counts in probability [20]. It also follows from
the work of Lovász [20], Diaconis and Janson [10] and in part from Bickel
and Chen [5] that there is a unique set of statistics whose joint distribu-
tion characterize the probability distribution on unlabeled networks. These
statistics, called “empirical moments” by Bickel, Chen and Levina [6] are
the counts of subgraphs in the network. The subgraphs most used are small
cycles like triad, tetrad and small acyclic graphs.

The expectation and variances of count statistics can, in principle be com-
puted (Picard et.al. [25]) and more usefully be asymptotically approximated
[6]. Under appropriate conditions normalized count statistics has limiting
Gaussian distribution. They have many uses [31] [30] [3], particularly in dis-
tinguishing between the mechanisms generating different graphs as well as
providing characterization of network distributions. The general asymptotic
Gaussian distribution of count statistics was provided in Theorem 1 of [6]
with an expression for the asymptotic mean and variance. However, no way
of calculating the quantities was provided in the paper [6].

Motifs or count statistics have been used in testing equality of features of
networks and finding confidence intervals of the count features [27] [21]. But,
a major stumbling block in their use has been the calculation of motifs that
have even moderately large number of vertices (like more than five) and even
more challenging problem of finding estimates of their variances. Finding
the correct count statistics or motifs is a computationally hard problem for large
networks as the complexity of finding the count of a subgraph is polynomial
in terms of number of vertices and when the number of vertices in the
network is even in thousands the computation becomes difficult and if it is in
millions, the computation becomes infeasible. Using subsampling methods
to calculate the count statistics we can greatly reduce the computational
burden of computing the statistics and inference using them.

In the statistical literature on networks, some work has been done on de-
vising sampling designs to select network samples. Various sampling designs
has been proposed in the statistical and computer science literature to derive
representative samples of a given network. [18], [19] and [28]. Many of these
sampling designs have been analyzed from the design-based sampling point
of view [29] [12]. There has also been work done on analyzing some of these
methods from the model-based sampling point of view, where, mostly the
exponential random graph model (ERGM) was considered as the model gen-
erating the network and a likelihood-based approach was taken for inference
[13]. As a result only parametric inference was possible. On the other hand,
our approach is not restricted to parametric models as we try to estimate the certain functionals of the underlying nonparametric generating model, using the samples obtained from the network data.

1.1. Contribution and Structure of Our Work. We use subsampling based bootstrap approaches to estimate the count statistics as well as find the approximate distributions for such count statistics under the general model of Bickel and Chen [5].

Along with the bootstrap methods and their theoretical analysis, we give two examples where use of count statistics provide some useful insights into the behavior of the networks. One of the two examples is the Jefferson high-school network given in Bearman et.al. (2004) [3] and the other example uses Facebook collegiate networks provided in Traud et.al. (2010) [18]. The high school network is a nice example where counts of specific types of subgraphs in the network, and their confidence intervals based on different generating models, give us useful insight into the behavior of nodes in the network [3]. The Facebook collegiate networks are larger and denser networks and calculation of count statistics for these networks would not be computationally feasible without the use of subsampling methods.

In section 2 we outline our main results. In section 3 we describe the bootstrap subsampling methods and the theoretical properties of each bootstrap estimators. We also indicate a method for estimating asymptotic variances of these estimators using bootstrap. Additionally, we give a theoretical comparison of the methods. In section 4, we give the general theorem on asymptotic Gaussianity of bootstrap subsampling estimates count statistics and their variance. In section 5 we perform a simulation study under two special cases of the general “nonparametric” model: stochastic block model and preferential attachment model respectively. In 6 we apply our method to test hypotheses about the count statistics of real networks.

2. Model and Statistics. We consider a random unlabeled graph $G_n$ as the data. Let $V(G_n) = \{v_1, \ldots, v_n\}$ denote the vertices of $G_n$ and $E(G_n)$ denote the set of edges of $G_n$. So, the number of vertices in $G_n$ is $|V(G_n)| = n$. We shall consider undirected, unweighted graphs only in this paper. For the sake of notational simplicity, we may denote $G_n$ by $G$.

As usual we suppose the network is represented by an adjacency matrix $A_{n \times n}$ whose elements are $A_{ij} \in \{0, 1\}$,

$$A_{ij} = \begin{cases} 1 & \text{if node } i \text{ links to node } j, \\ 0 & \text{otherwise}. \end{cases}$$
A finite sample version of the Aldous-Hoover representation for exchangeable adjacency matrices $A_{n \times n}$ becomes - for $i, j \in \{1, \ldots, n\}$,

$$\alpha, \xi_i, \eta_{ij} \overset{iid}{\sim} U(0, 1)$$

$$A_{ij} = f_n(\alpha, \xi_i, \xi_j, \eta_{ij})$$

where, $f_n : [0, 1]^4 \rightarrow [0, 1]$ is a measurable function symmetric in its second and third arguments and $\eta_{ij} = \eta_{ji}$. Note that this is not a representation of all exchangeable probability distributions on finite networks.

Bickel and Chen [5] considered a special form of the general Aldous-Hoover representation

$$h_n(u, v) \equiv P(A_{ij} = 1 | \xi_i = u, \xi_j = v)$$

The above mentioned form can be simplified by decoupling $n$ from the contribution of $(\xi_i, \xi_j)$. So, $h_n$ is modeled as product of a scale function in terms of $n$, $\rho_n$, defined as

$$\rho_n = \int_0^1 \int_0^1 h_n(u, v) du dv$$

and a bivariate function independent of $n$, the latent variable density, $w(\xi_i, \xi_j)$. We call the resulting model nonparametric latent variable model and the model equation as described in Bickel, Chen and Levina (2011) [6] becomes - for $i, j \in \{1, \ldots, n\}$, $\xi_i \overset{iid}{\sim} U(0, 1)$ and

$$(2.1) \quad P(A_{ij} = 1 | \xi_i = u, \xi_j = v) = h_n(u, v) = \rho_n w(u, v) 1(w \leq \rho_n^{-1}),$$

where, $w(u, v) \geq 0$, symmetric, $0 \leq u, v \leq 1$, $\int \int w(u, v) du dv = 1$, $0 < \rho_n < 1$ and we define expected degree $\lambda_n = n \rho_n$.

The graph statistics that we are concerned with, are count statistics of subgraphs. Let $R$ be a subgraph of $G$, with $V(R) \subseteq V(G)$ and $E(R) \subseteq E(G)$. We have $|V(R)| = p$ and $|E(R)| = e$. For notation, if two graphs $R$ and $S$ are equivalent, we denote them by $R \cong S$ and if $R$ is a subgraph of $S$, we denote them by $R \subseteq S$. The integral parameter corresponding to a subgraph $R$ is defined as $P(R)$,

$$(2.2) \quad P(R) = \mathbb{E}\left\{ \prod_{(i,j) \in E(R)} h(\xi_i, \xi_j) \prod_{(i,j) \in E(\bar{R})} (1 - h(\xi_i, \xi_j)) \right\}$$

where, $\bar{R}$ is a subgraph of $K_p$ ($K_p$ is complete graph on $p$ vertices) with $V(\bar{R}) = \{i, j : (i, j) \notin E(R), i, j \in V(R)\}$ and $E(\bar{R}) = \{(i, j) : (i, j) \notin E(R), i \in V(R), j \in V(R)\}$. 
Now, the empirical statistic corresponding to \( P(R) \), which is the count statistics for subgraph \( R \) is

\[
\hat{P}(R) = \frac{1}{\binom{n}{p}|\text{Iso}(R)|} \sum_{S \subseteq K_n, S \cong R} 1(S \subseteq G)
\]

where, \( \text{Iso}(R) \) is the group of Isomorphisms of \( R \) and \( K_n \) is the complete graph on \( n \) vertices.

We also have from [6],

\[
\mathbb{E}(\hat{P}(R)) = P(R)
\]

Examples of subgraphs and corresponding count statistics include -

**Example 1.** \( R = \text{edge} \) is a subgraph with two vertices and one edge connecting them, so, \( \hat{P}(R) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} D_i \), where, \( D_i = \text{degree of } v_i, v_i \in V(G) \).

**Example 2.** \( R = \text{triangle} \) is a 3-clique subgraph, so,

\[
P(R) = \int_{0}^{1} \int_{0}^{1} \int_{0}^{1} h_n(u,v)h_n(v,s)h_n(s,u)\,du\,dv\,ds.
\]

**Example 3.** We define a smooth function of counts of triangle and ‘\( V \)’s, known as transitivity, \( T_{Tr} \), defined as

\[
\hat{P}_{Tr} = \frac{\hat{\rho}_n^{-3} \hat{P}(R_1)}{\hat{\rho}_n^{-3} P(R_1) + \hat{\rho}_n^{-2} \hat{P}(R_2)}
\]

where, \( R_1 \) is a triangle or a 3-cycle and \( R_2 \) is a ‘\( V \)’ or a path with three vertices and \( \hat{\rho}_n = \hat{P}(\text{edge}) \).

**Example 4.** \( R = p\)-cycle is a cyclic subgraph with \(|V(R)| = p\), \(|E(R)| = p\) and \( R \) is a ring containing all \( p \) vertices. Triangle is a 3-cycle.

\[
P(R) = \int_{0}^{1} \cdots \int_{0}^{1} h_n(u_1,u_2) \cdots h_n(u_{p-1},u_p)h_n(u_p,u_1)du_1 \cdots du_p.
\]

**Definition 5 (Wheels).** A \((k,l)\)-wheel is an acyclic graph with \( kl + 1 \) vertices and \( kl \) edges and “hub” vertex, (say, \{1\}), isomorphic to the graph with edges \\{((1, 2), (2, 3), \ldots, (k, k+1)), ((1, k+2), (k+2, k+3), \ldots, (2k, 2k+1)) \ldots, ((1, (l-1)k+2), ((l-1)k+2, (l-1)k+3), \ldots, (lk, lk+1))\}.

Edges, ‘\( V \)’, ‘\( W \)’ are all examples of \((k, l)\)-wheels. An edge is a \((1,1)\)-wheel, a ‘\( V \)’ is a \((1,2)\)-wheel and a ‘\( W \)’ is a \((2,2)\)-wheel.
Now, as \( \rho_n \to 0 \), which is the case for graphs which are not fully dense that is \(|E(G_n)| = O(n^2)\), \( P(R) \to 0 \) as well as its estimator \( \hat{P}(R) \xrightarrow{P} 0 \) and the asymptotics on \((\hat{P}(R) - P(R))\) become uninformative. So, in order to get a proper analysis of the behavior of \( \hat{P}(R) \) in relation to \( P(R) \), we have to appropriately normalize both \( P(R) \) and \( \hat{P}(R) \). The normalized versions of parameter \( P(R) \) is defined as

\[
\tilde{P}(R) = \rho_n^{-e} P(R) \tag{2.4}
\]

where, \( e \equiv |E(R)| \), then, we define the corresponding normalized statistic to be

\[
\hat{T}(R) = \hat{\rho}^{-e} \hat{P}(R) \tag{2.5}
\]

where,

\[
\hat{\rho} = \frac{\bar{D}}{n - 1}. \tag{2.6}
\]

where, \( D_i = \text{degree of } v_i, v_i \in V(G_n) \) for \( i = 1, \ldots, n \) and \( \bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i \).

Now the investigation on asymptotic behavior of \( \sqrt{n}(\hat{T}(R) - \tilde{P}(R)) \) is possible, as both terms are asymptotically non-zero quantities. This investigation was done in the paper [6].

We wish to approximate the statistic \( \hat{P}(R) \) and functional \( \text{Var}(\hat{P}(R)) \) by nonparametric bootstrap. We consider two bootstrap procedures.

(I) uniform subsampling bootstrap procedure
(II) subgraph subsampling bootstrap procedure.

Let us consider the bootstrap estimate of \( \hat{P}(R) \) to be \( \hat{T}_B(R) \) and bootstrap estimate of \( \text{Var}(\hat{P}(R)) \) to be \( \hat{\sigma}^2_B(R) \). We consider \( B \) as the bootstrap repetition or resampling parameter (number of bootstrap subsamples). How, we get the bootstrap estimates will be discussed in next section and we will state theorems justifying the use of these bootstrap estimation in next two sections.

2.1. Bootstrap and Model-based Sampling. Our work can be viewed from two different perspectives. The first perspective is that of the bootstrap. In non-parametric bootstrap, we use resamples or subsamples of the data, where the data comes from an unknown distribution, to find the functionals of the unknown distribution. In our situation also, we have a network that has been generated from an underlying probability model. We want to subsample networks from our given network and use those subsampled networks
to approximate estimates of functionals of the underlying population model generating the given network. Note that here we are interested in subsampling not resampling of a network. Our use of the bootstrap corresponds to Efron’s [11] use of bootstrap for approximations made by Monte Carlo quantities which in principle could be calculated using the data such as the empirical variance of complicated estimates.

The second perspective is that of sampling. In sampling, we consider that the population, from which the sample is selected according to some sampling design, is a realization of a probabilistic event. So, in our case, we consider the given network as the population and it is generated from an underlying probability model. We used subsampling bootstrap or sampling of network data to get estimates for population quantity (count statistics) and underlying probability model (integral parameter).

3. Bootstrap Methods. We consider two different bootstrap methods. Both of the methods of bootstrap consider finding subsamples from the whole network given as the data. In the following subsections, we shall define each of these subsampling bootstrap methods. We shall also compare the theoretical performance between the two bootstrap schemes.

Let the adjacency matrix of $G_n$ be denoted by $A_{n \times n}$. Let $R$ be a subgraph of $G$, with $V(R) \subseteq V(G)$ and $E(R) \subseteq E(G)$. We have $|V(R)| = p$ and $|E(R)| = e$.

3.1. Uniform Subsampling Bootstrap. In the uniform subsampling bootstrap scheme, at each bootstrap iteration a subset of vertices of the full network $G$ is selected without replacement and the graph induced by the selected subset of vertices is the subnetwork we consider. This is a vertex subsampling or induced network sampling scheme. Given subnetwork size $m$ and number of bootstrap iterates $B$, the uniform subsampling bootstrap scheme has the following steps.

1. For $b^{th}$ iterate of the bootstrap, $b = 1, \ldots, B$.
2. Choose $m$ vertices without replacement from $V(G)$ and form the induced subgraph of $G$ based on the selected vertices. Denote the graph formed by $H$.
3. Calculate $\hat{P}_{b1}(R)$, given by formula

$$
(3.1) \quad \hat{P}_{b1}(R) = \frac{1}{{m \choose p}|Iso(R)|} \sum_{S \subseteq K_m, S \equiv R} 1(S \subseteq H)
$$
The uniform subsampling bootstrap estimate of \( \hat{P}(R) \) is given by

\[ \tilde{P}_{B1}(R) = \frac{1}{B} \sum_{b=1}^{B} \hat{P}_{b1}(R) \]

The uniform subsampling bootstrap scheme is the network version of the common subsampling bootstrap scheme seen in Bickel et. al. [7]. Note that, there are other ways of forming uniformly subsampled bootstrap estimates as mentioned in [7], however, we just mention one of them in this discourse.

For the bootstrap method, we prove a theorem of following type -

**Theorem 3.1.** Suppose \( R \) is fixed acyclic or p-cycle with \(|V(R)| = p \) and \(|E(R)| = e\),

(i) Given \( G \), \( \hat{P}_{b1}(R) \) is an unbiased estimate of \( \hat{P}(R) \).

(ii) Given \( G \), \( \text{Var} \left( \rho_{n}^{-e} \hat{P}_{b1}(R) | G \right) = O \left( \frac{1}{m^{p\rho_{n}} n} \vee \frac{1}{m} \right). \)

(iii) Also if \( B \to \infty, n \to \infty, m \to \infty, m/n \to 0 \) and \( B(m^{p\rho_{n}} m) > O(n) \), under \( G \) generated from (2.1),

\[ \sqrt{n} \left( \rho_{n}^{-e} \tilde{P}_{B1}(R) - \rho_{n} P(R) \right) \to 0 \]

**Proof.** The proof is given in Appendix A1 in [4]. □

### 3.2. Subgraph Subsampling bootstrap.

In the subgraph subsampling bootstrap scheme we use an enumeration scheme to find all possible subgraphs \( R \) of size \(|V(R)| = p \) in the graph \( G \). Then, we convert the enumeration scheme into a sampling scheme by selecting each subgraph \( R \) of size \( p \) of \( G \) with a fixed probability and counting the number of sampled subgraphs. The enumeration scheme was proposed by Wernicke et al. (2006) [32]. A random version of the enumeration scheme was also proposed in the paper [32]. We use the random version of the enumeration scheme to form our sampling scheme.

Let us first discuss the enumeration scheme of Wernicke et al [32], which we shall henceforth call ESU. The enumeration algorithm is a breadth-first search algorithm. The enumeration scheme creates a forest of tree structures such that each tree corresponds to one vertex of the network \( G \) and each leaf of each tree is a size-\( p \) subgraph (we have, \(|V(R)| = p \) of \( G \). Since, the counting scheme follows a breadth-first search route, before performing the ESU algorithm, we need an ordering of the vertices based on breadth-first search of the graph starting from any particular vertex (say \( v_{1} \)). We get such a particular fixed ordering of the vertices of the network with \( v_{1} \)
getting lowest order value and subsequently searched vertices getting higher order values. The ordering is described in the algorithm AssignOrder or \( \text{AO} \), where, given any set of vertices \( \mathcal{V} \), we denote the set of vertices connected to \( \mathcal{V} \), that is the neighbors of \( \mathcal{V} \), by \( N(\mathcal{V}) \). Also, the based on the ordering defined by \( \text{AO} \), we denote \( v_i \succ v_j \), if \( v_i \) has higher order than \( v_j \).

The enumeration algorithm starts with an available vertex of lowest possible order (where, order is specified by \( \text{AO} \)), say \( v_1 \). We construct a tree with the vertex \( v_1 \) as the root node. We consider \( v_1 \) as the “parent” node and neighbors of \( v_1 \), which have a higher order than \( v_1 \) as its “children”. In the next step, the “children” node become the “parent” node in the tree and has its own neighbors which have higher order than the nodes that have already come into the tree as their “children”. We define \( N_{\text{excl}}(v, \mathcal{V}) \) ( \( v \) is a vertex and \( \mathcal{V} \) is a set of vertices) for \( N(\mathcal{V}) \setminus \mathcal{V} \). The tree is allowed to grow up to a height \( p \), if we are counting size-\( p \) subgraphs. So, we can see that each leaf of the tree represents a collection of \( p \) nodes coming from the path connecting the leaf to the root \( v_1 \). For each vertex, we have such a tree and over counting is averted as we maintain the order of vertices assigned by \( \text{AO} \) while forming the trees. So, with the help of the particular ordering of vertices, each of the size-\( p \) subgraphs (\( |R| = p \)) is counted only once.

**Algorithm 3.1 AssignOrder(\( G, p \))**

**Input:** A graph \( G = (V, E) \), where, \( |V(G)| = n \).

**Output:** A vector \( \sigma = (\sigma(1), \ldots, \sigma(n)) \), where, \( \sigma \) is some permutation of \( \{1, \ldots, n\} \) and \( \sigma(i) \) is associated with vertex \( v_{\sigma(i)} \in V(G) \) for all \( i = 1, \ldots, n \).

1: \( \sigma_1 \leftarrow 1 \)
2: \( V \leftarrow \{v_1\} \)
3: \( i \leftarrow 1 \)
4: while \( |V| < n \) do
5: \( \text{Denote } k \leftarrow |N(V) \setminus V| \text{ and } \{v_{h_1}, \ldots, v_{h_k}\} = N(V) \setminus V \)
6: \( \text{Define } \sigma(i + j) \leftarrow h_j \text{ for } j = 1, \ldots, k \).
7: \( i \leftarrow i + k \)
8: \( V \leftarrow V \cup N(V) \)
9: end while

The randomized enumeration algorithm RAND-ESU 3.2 also creates a forest of tree structures such that each tree corresponds to one vertex of the network \( G \) and each leaf of each tree is a size-\( p \) subgraph (we have, \( |V(R)| = p \)) of \( G \). However only a random selection of leaves of each tree is present in RAND-ESU with uniform probability of selection of each leaf. The random enumeration algorithm starts with an available vertex of lowest possible order (where, order is specified by \( \text{AO} \)), say \( v_1 \) chosen with probability \( q_1 \). We construct a tree with the vertex \( v_1 \) as the root node. We consider \( v_1 \) as the “parent” node and neighbors of \( v_1 \), which have a higher order than \( v_1 \) as
its “children” and each “child” is selected with probability $q_2$ independently. In the next step, the “children” node become the “parent” node in the tree and has its own neighbors, which have higher order than the nodes that have already come into the tree, as their “children” and each “child” is selected with probability $q_3$. The tree is allowed to grow up to a height $p$, if we are counting size-$p$ subgraphs and at step $d$ the probability of selection is $q_d$. So, we can see that each leaf of the tree represents a collection of $p$ nodes coming from the path connecting the leaf to the root. For each vertex, we have such a tree. So, with the help of the particular ordering of vertices, a subsample of the size-$p$ subgraphs ($|R| = p$) is obtained. The pseudo-code is given in Algorithm 3.2.

The ordering is needed for success of the ESU algorithm and its randomized counterpart 3.2. We formally state the subsampling algorithm, $\text{RAND}-\text{ESU} \ 3.2$ in this paper with an extra set of parameters $(q_1, \ldots, q_p)$. The enumeration version can be found in [32].

Algorithm 3.2 $\text{RANDOMIZEDENUMERATESUBGRAPH}(G, p)$

**Input:** A graph $G = (V, E)$, an integer $p$ and an vector $(q_1, \ldots, q_p)$, where, $1 \leq p \leq |V|$ and $q_d \leq 1$ for all $d = 1, \ldots, p$.

**Output:** $S_{p}^{R} = A$ sample of subgraphs, $R$ of $G$, such that $|R| = p$.

1: for each vertex $v \in V$ do
2: $V_{\text{Extension}} \leftarrow \{ u \in N(\{v\}) : u > v \}$
3: $d \leftarrow 1$
4: With probability $q_d$ Call $\text{RandExtendSubgraph}(\{v\}, V_{\text{Extension}}, v, d)$
5: end for
6: function $\text{RandExtendSubgraph}(V_{\text{Subgraph}}, V_{\text{Extension}}, v, d)$
7: **Input:** Graphs $V_{\text{Subgraph}}, V_{\text{Extension}}$ and vertex $v$.
8: **Output:** A sample of subgraphs, $R$ of $G$, such that $|R| = p$ and $v$ is a vertex of $R$.
9: if $|V_{\text{Subgraph}}| = p$ then
10: return Subgraph of $G$ induced by $V_{\text{Subgraph}}$
11: else
12: while $V_{\text{Extension}} \neq \phi$ do
13: Remove an arbitrarily chosen vertex $w$ from $V_{\text{Extension}}$
14: $V_{\text{Extension}}' \leftarrow V_{\text{Extension}} \cup \{ u \in N_{\text{excl}}(w, V_{\text{Subgraph}}) : u > v \}$
15: $d \leftarrow |V_{\text{Subgraph}}| + 1$
16: With probability $q_d$ Call $\text{RandExtendSubgraph}(V_{\text{Subgraph}} \cup \{w\}, V_{\text{Extension}}', v, d)$
17: end while
18: end if
19: return
20: end function

From the sampling scheme $\text{RAND}-\text{ESU}$ we have a sample $S_{p}^{R}$ of size-$p$ subgraphs of $G$. Now, if we consider each item to be one size-$p$ subgraph of $G$, that is, an element of $S_p$, then, we can try to calculate the inclusion
probability of each item in the sample $\mathcal{S}_p^R$.

An item, $S \in \mathcal{S}_p$ is a subgraph of $G$ induced by the set of vertices \{w_1, \ldots, w_p\}, where, we take that $w_{i+1} \succ w_i$, $i = 1, \ldots, p - 1$. So,

$$\pi \equiv \text{Inclusion Probability of } S = P[(w_1, \ldots, w_p) \text{ is selected}] = P[w_p|(w_1, \ldots, w_{p-1}) \text{ is selected}] = q_p \cdot P[(w_1, \ldots, w_{p-1}) \text{ is selected}] = q_p \cdot q_{p-1} \cdot P[(w_1, \ldots, w_{p-2}) \text{ is selected}] = \cdots = q_p \cdot q_{p-1} \cdots q_1 = \prod_{d=1}^{p} q_d$$

So, each item $S \in \mathcal{S}_p$ has an inclusion probability $\pi$ to be in the sample $\mathcal{S}_p^R$.

In Theorem 2 of [32] it was proved that the output of ESU algorithm, $\mathcal{S}_p$ contains all subgraphs $R$ of $G$, such that $|R| = p$, exactly once. So, we can write the statistic (2.3) for a specific subgraph $R$ with $|R| = p$ in the following way

\begin{equation}
\hat{P}(R) = \frac{1}{(\begin{smallmatrix} n \\ p \end{smallmatrix})} \sum_{S \in \mathcal{S}_p} 1(S \cong R)
\end{equation}

Essentially we have a normalized population total in terms of sampling theory. Our goal is to form a sampling design and devise a corresponding sampling estimator of $\hat{P}(R)$ given $G$. To meet that goal we use a sampling version of the enumeration scheme ESU.

So, now we have a sampling scheme, by which we select a sample $\mathcal{S}_p^R$ from the population $\mathcal{S}_p$, where, each element of $\mathcal{S}_p$ has probability of inclusion of $\pi$. So, we can define a Horvitz-Thompson estimator (for reference see Chapter 6.2 pf [28]) of $\hat{P}(R)$ based on $\mathcal{S}_p^R$ as

\begin{equation}
\hat{P}_{b2}(R) = \frac{1}{(\prod_{d=1}^{p} q_d)(\begin{smallmatrix} n \\ p \end{smallmatrix})} \sum_{S \in \mathcal{S}_p^R} 1(S \cong R)
\end{equation}

Now, if we repeat the same procedure $B$ number of times, each time getting independent copies of $\mathcal{S}_p^R$ with replacement from $\mathcal{S}_p$, we can get the subgraph subsampling bootstrap estimate -

\begin{equation}
\hat{P}_{B2}(R) = \frac{1}{B} \sum_{b=1}^{B} \hat{P}_{b2}(R)
\end{equation}

For the bootstrap method, we prove a theorem of following type -
**Theorem 3.2.** Suppose $R$ is fixed acyclic or $p$-cycle with $|V(R)| = p$ and $|E(R)| = e$,

(i) Given $G$, $\hat{P}_b(R)$ is an unbiased estimate of $\hat{P}(R)$.

(ii) Given $G$, $\text{Var}(\rho_n^{-e}\hat{P}_b(R)|G) = O\left((\frac{1}{q_1} - 1)\frac{1}{n} + \frac{1}{n\rho_n^{p+1}}\cdot\prod_{d=2}^{p}\frac{1}{\chi_n q_d}\right)$.

(iii) For $B \to \infty$ and $q_d \to 0$ for all $d = 1, \ldots, p$ such that $\frac{1}{B}\left(\frac{1}{q_1} - 1\right) \to 0$ and $B\prod_{d=2}^{p} q_d \geq \frac{1}{n} \rho_n$ and $n \to \infty$, $\lambda_n \to \infty$ and under $G$ generated from (2.1),

\[(3.7) \quad \sqrt{n}(\rho_n^{-e}\hat{P}_b(R) - \rho_n^{-e}P(R)) \xrightarrow{P} 0 \]

**Proof.** The proof is given in Appendix A2 in [4].

Note that, the main reason of taking repeated independent samples, $S^R_p$ for this case is to reduce the variance of the bootstrap estimates and making the estimates more stable.

**3.3. Estimation of Variance and Covariance.** We first start with the situation when the source of variation is only the randomness coming from sampling from the underlying model (2.1). We try to give an estimate of asymptotic variance of $\hat{T}(R)$ or $\sigma^2(R)$ and covariance between $\hat{T}(R_1)$ and $\hat{T}(R_2)$ or $\sigma(R_1, R_2)$, where, $e_1 = |E(R_1)|$, $e_2 = |E(R_2)|$, $p_1 = |V(R_1)|$ and $p_2 = |V(R_2)|$.

**Proposition 6.** For connected subgraphs $R$, $R_1$ and $R_2$ of $G$, we have that,

\[
\text{Var}\left[\rho^{-e}\hat{P}(R)\right] = \frac{1}{(\rho^e\binom{n}{p}|\text{Iso}(R)|)^2} \sum_{W:W=\hat{S}\cap\hat{T}, S\cap T\neq\emptyset} \mathbb{E}\left[\sum_{W\subseteq K_n} 1(W \subseteq G)\right] - \left(1 - \frac{((n-p)!)^2}{n!(n-2p)!}\right)^2 \left(\hat{P}(R)\right)^2
\]

\[
\text{Cov}\left(\rho^{-e_1}\hat{P}(R_1), \rho^{-e_2}\hat{P}(R_2)\right) = \frac{1}{(\rho^{e_1+e_2}\binom{n}{p_1}\binom{n}{p_2}|\text{Iso}(R_1)||\text{Iso}(R_2)|)} \times \sum_{W:W=\hat{S}\cup\hat{T}, S\subseteq R_1, T\subseteq R_2, S\cap T\neq\emptyset} \mathbb{E}\left[\sum_{W\subseteq K_n} 1(W \subseteq G)\right] - \left(1 - \frac{(n-p_1)(n-p_2)!}{n!(n-p_1-p_2)!}\right)^2 \hat{P}(R_1)\hat{P}(R_2)
\]

**Proof.** The proof is given in Appendix B1 in [4].
Note that, if we take $k = |V(W)|$ and $e_W \equiv |E(W)|$, then, $k = p, \ldots, 2p-1$ and each term of sum in RHS of previous equation is

$$
\frac{1}{(\rho^e(p)|Iso(R)|)^2} \mathbb{E} \left[ \sum_{W \subseteq K_n} 1(W \subseteq H) \right] = \frac{\rho^e_w(n)|Iso(W)|}{(\rho^e(p)|Iso(R)|)^2} \hat{\rho}(W)
$$

$$
= O(n^{k-2p} \rho e_w^{-2e}).
$$

So, we can analyze each such term separately -

1. If $k = |V(W)| = 2p - 1$, then, $W$ is a connected graph, with $e_W = 2e$. So, we have the main leading term equals $O(\frac{1}{n})$.

2. In the case, $k = |V(W)| < (2p - 1),

   • If $R$ is acyclic, then, $e_W - 2e \leq k - 2p - 1$, since, $e = p - 1$, so $O(n^{k-2p} \rho e_w^{-2e}) = o(n^{-1})$ if $\lambda_n = n \rho_n \to \infty$.

   • If $R$ is a $p$-cycle, $e_W - 2e = k - 2p < 0$ if $k = |V(W)| = p$ and $e_W - 2e \leq k - 2p - 1$ if $p < k < (2p - 1)$, so $O(n^{k-2p} \rho e_w^{-2e}) = O(\lambda_n^{-p}) + o(n^{-1})$ if $\lambda_n \to \infty$.

   • If $R$ is any other cyclic graph, $O(n^{k-2p} \rho e_w^{-2e}) = O \left( \frac{n^{-c} \rho^{-d}}{c} \right)$, where, $0 < c \leq p$ and $0 < d \leq c(c - 1)/2$ for each $c$. So, in order to have $n^{-c} \rho^{-d} \leq Mn^{-1}$, the worst rate that $\lambda_n$ can have is $\lambda_n = O \left( n^{1-2/p} \right)$.

So, for connected and acyclic or $p$-cycle $R, R_1$ and $R_2$, we get that

$$
\text{Var} \left[ \rho^{-e} \hat{\rho}(R) \right] = O \left( \frac{1}{n} \vee \frac{1}{\lambda_n^p} \right)
$$

$$
\text{Cov} \left( \rho^{-e_1} \hat{\rho}(R_1), \rho^{-e_2} \hat{\rho}(R_2) \right) = O \left( \frac{1}{n} \right)
$$

So, for calculation of variance, if $R$ is acyclic or $p$-cycle, we only estimate the count of the features which are $W = S \cup T$ and $|V(W)| = 2p - 1$ and $|V(W)| = p$. So, the estimator of variance becomes

$$
\hat{\sigma}^2(R) = \frac{1/(1 - x)}{\left( \hat{\rho}^e(n)|Iso(R)| \right)^2} \sum_{W \subseteq K_n, W = S \cup T, S,T \subseteq R, |S \cap T|=1,p} 1(W \subseteq G) - \frac{x \hat{\rho}^{-2e} \hat{\rho}(R)^2}{(1 - x)}
$$
where, $x = \left(1 - \frac{(n-p)!}{n!(n-2p)!}\right)$ and the estimator of covariance becomes

\begin{equation}
\hat{\sigma}(R_1, R_2) = \frac{1/(1-y)}{\left(\hat{p}_n^{(n) + (n)_{p_1}^{(n)_{p_2}}}|\text{Iso}(R)|\right)} \sum_{W \subseteq K_n, W = S \cup T, S \supseteq R_1, T \supseteq R_2, |S \cap T| = 1} 1(W \subseteq G) - \frac{y\hat{p}_n^{-(e_1 + e_2)} \hat{P}(R_1) \hat{P}(R_2)}{(1-y)}.
\end{equation}

where, $y = \left(1 - \frac{(n-p)!}{n!(n-2p)!}\right)$.

\[\hat{\sigma}^2(R) \text{ and } \hat{\sigma}(R_1, R_2) \text{ become consistent estimates of } \text{Var} \left[\hat{T}(R)\right] \text{ and } \text{Cov} \left[\hat{T}(R_1), \hat{T}(R_2)\right] \text{ respectively.} \]

**Lemma 7.** As $\lambda_n \to \infty$ and $n \to \infty$, if $R, R_1, R_2$ is connected acyclic or if p-cycle then additionally $\lambda_n^p \geq O(n)$,

\begin{equation}
\frac{\hat{\sigma}^2(R)}{\hat{\sigma}^2(R)} \xrightarrow{p} 1.
\end{equation}

\begin{equation}
\frac{\hat{\sigma}(R_1, R_2)}{\sigma(R_1, R_2)} \xrightarrow{p} 1.
\end{equation}

**Proof.** The proof is given in Appendix B2 in [4].

Now, we can see that $\hat{\sigma}^2(R)$ and $\hat{\sigma}(R_1, R_2)$ are nothing but count statistics on the statistic $W = S \cup T$. So, using bootstrap methods, we can get an estimate of $\hat{\sigma}^2(R)$, for $i = 1, 2$,

\begin{equation}
\hat{\sigma}^2_{Bi}(R) = \sum_{W = S \cup T, S \supseteq R_i, T \supseteq R_i, |S \cap T| = 1, p} \left(\hat{p}_n^{(n)_{pw}}|\text{Iso}(R)|\right)^2 \hat{P}_{Bi}(W) - \frac{x\hat{p}_n^{-(e_2)} \hat{P}_{Bi}(R)^2}{(1-x)}
\end{equation}

where $x = \left(1 - \frac{(n-p)!}{n!(n-2p)!}\right)$. An estimate of $\hat{\sigma}(R_1, R_2)$ is

\begin{equation}
\hat{\sigma}_{Bi}(R_1, R_2) = \sum_{W = S \cup T, S \supseteq R_1, R_2, T \supseteq R_2, |S \cap T| = 1} \left(\hat{p}_n^{(n)_{pw}}|\text{Iso}(W)|\right) \hat{P}_{Bi}(W) - \frac{y\hat{p}_n^{-(e_1 + e_2)} \hat{P}_{Bi}(R_1) \hat{P}_{Bi}(R_2)}{(1-y)}.
\end{equation}
where \( y = \left(1 - \frac{(n-p_1)!}{n!(n-p_1-p_2)!}\right) \) and \( \bar{P}_{B_i}(W) \) \((i = 1, 2)\) are bootstrap count statistics estimates, defined in Eq (3.2) and (3.6).

Lemma 8. As \( \lambda_n \to \infty, n \to \infty, B \to \infty \) and under conditions of Theorem 3.1 and Theorem 3.2, if \( R, R_1 \) and \( R_2 \) is acyclic or if \( p \)-cycle then additionally \( \lambda_n^2 \geq O(n) \),

\[
\frac{\hat{\sigma}^2_{B_i}(R)}{\sigma^2(R)} \xrightarrow{p} 1 \text{ for } i = 1, 2
\]

\[
\frac{\hat{\sigma}_{B_i}(R_1, R_2)}{\sigma(R_1, R_2)} \xrightarrow{p} 1 \text{ for } i = 1, 2
\]

Proof. The proof is given in Appendix B3 in [4].

3.4. Comparison of the Bootstrap Methods. The variance of each of the subsampling bootstrap methods just on the basis of the randomness generated from the bootstrap sampling is given in Theorem 3.1 and 3.2. Also, the worst case computational complexity of finding count statistics for subgraphs \( R \) of size \( p \) for uniform subsampling bootstrap becomes \( O(B m^p) \), while as for subgraph subsampling bootstrap scheme the worst case complexity is \( O(B \prod_{d=1}^{p} (nq_d)) \). Now the question of balancing computational complexity and statistical stability become important.

For dense networks, say when \( \rho_n = n^{-2} \) with \( \varepsilon > 0 \) small (say between \( 0 < \varepsilon < 1/2 \)), we also have \( \lambda_n = n^{1-\varepsilon} \).

- For uniform subsampling from Theom 3.1, we get that \( \text{Var} (\rho^{-\varepsilon} \bar{P}_{B_1}(R)) = O\left(\frac{1}{n^{1+2\varepsilon}}\right) \) with \( m = n^\varepsilon \) and \( B = n^{1+\varepsilon} \). The worst case computational cost becomes \( O\left(n^{1+(p+1)\varepsilon}\right) \).
- For subgraph subsampling from Theom 3.2, we get that \( \text{Var} (\rho^{-\varepsilon} \bar{P}_{B_2}(R)) = O\left(\frac{1}{n^{1+\varepsilon}}\right) \) for \( p \)-cycle \( R \) and \( O\left(\frac{1}{n^{1+\varepsilon}}\right) \) for acyclic \( R \) with \( q_d = O\left(\frac{1}{n^{1-\varepsilon}}\right) \) for \( d = 2, \ldots, p \) and \( B = n^{2\varepsilon} \). The worst case computational complexity becomes \( O\left(n^{1+(p+1)\varepsilon}\right) \).

So, in dense networks, both the subsampling bootstrap methods can achieve low enough bootstrap variance for low computational cost. In fact the gain in computational complexity is quite astonishing as polynomial complexity gets reduced to near-linear complexity. The uniform subsampling bootstrap is a better choice for its ease of use and marginally smaller variance for \( p \)-cycle \( R \). However, since \( m \) has to be greater than \( p \), for large \( R \) the benefit of using uniform subsampling bootstrap starts to reduce and in those cases subgraph subsampling bootstrap might be a better choice.
For sparse case, say when $\rho_n = n^{\varepsilon-1}$ with $\varepsilon > 0$ small (say between $0 < \varepsilon < 1/2$), we also have $\lambda_n = n^{-\varepsilon}$.

- For uniform subsampling from Theorem 3.1, we get that $
abla (\rho - \rho \overline{P}_{B_1}(R)) = O\left(\frac{1}{n^{\varepsilon}}\right)$ for acyclic $R$ and $O\left(\frac{1}{n^{1+\varepsilon}}\right)$ for $p$-cycle $R$ with $m = n^{1-\varepsilon}$ and $B = n^{1+\varepsilon}$. The worst case computational cost becomes $O\left(n^{p-(p-1)\varepsilon-1}\right)$.

- For subgraph subsampling from Theorem 3.2, we get that $
abla (\rho - \rho \overline{P}_{B_2}(R)) = O\left(\frac{1}{n^{\varepsilon}}\right)$ for acyclic $R$ and $O\left(\frac{1}{n^{1+\varepsilon}}\right)$ for $p$-cycle $R$ with $q_d = O\left(\frac{1}{n^{\varepsilon}}\right)$ for $d = 1,\ldots,p$ and $B = n$. The worst case computational complexity becomes $O\left(n^{p-(p-1)\varepsilon-1}\right)$.

So, in sparse networks the computational advantage of using subsampling bootstrap starts to reduce, especially for small subgraphs $R$. However, for large subgraph $R$ there is still computational advantage in using subsampling bootstrap methods. The subgraph subsampling bootstrap scheme is a better choice in this case as it has smaller variance for similar computational complexity.

But for sparse graphs, the methods still remain polynomial in worst case complexity and for large $p$ and $n$ the methods become numerically infeasible. In those cases, it becomes more of a detection problem than a counting problem and a fundamentally different approach will be required for feasible inference.

4. Theoretical Results. In this section, we shall try to provide asymptotic distribution for normalized bootstrap estimates of count statistics. We define normalized bootstrap estimates of count statistic for subgraph $R$ from (3.2) and (3.6) by -

$$\hat{T}_{B_i}(R) = \hat{\rho}_n^{-e} \hat{P}_{B_i}(R)$$

where, $i = 1, 2$ for the two different bootstrap schemes. By obtaining an estimate of the asymptotic variance of $\rho^{-e} \hat{P}(R)$, we can estimate its asymptotic distribution and thus construct hypothesis tests based on the asymptotic distribution. We combine the results obtained in Section 3 to prove Theorem 4.1.

**Theorem 4.1.** Suppose $R$ is fixed, acyclic or $p$-cycle with $|V(R)| = p$ and $|E(R)| = e$ and

$$\int_0^\infty \int_0^\infty w^2(u,v)du dv < \infty.$$
3.2. for \( i = 1, 2 \), if \( \lambda_n(\equiv np_n) \to \infty \) and \( B \to \infty \)

\[
\begin{align*}
(4.2) \quad & \sqrt{n} \left( \hat{T}_{Bi}(R) - \bar{P}(R) \right) \xrightarrow{P} 0 \\
(4.3) \quad & \sqrt{n} \left( \frac{\hat{T}_{Bi}(R) - \bar{P}(R)}{\hat{\sigma}_{Bi}(R)} \right) \xrightarrow{w} N(0, 1)
\end{align*}
\]

If for fixed, acyclic or p-cycle subgraphs \((R_1, \ldots, R_k)\), we define,
\[
T_{Bi}(R) = (\hat{T}_{Bi}(R_1), \ldots, \hat{T}_{Bi}(R_k)) \quad \text{and} \quad P(R) = (\bar{P}(R_1), \ldots, \bar{P}(R_k))
\]

\[
(4.4) \quad \sqrt{n} \left( (T_{Bi}(R) - P(R))^T \hat{\Sigma}_{Bi}^{-1/2}(R) (T_{Bi}(R) - P(R)) \right) \xrightarrow{w} N(0, I)
\]

where, \( \hat{\Sigma}_{Bi} \) is the estimator of \( \Sigma_{Bi} \), and if \( R_s = R_t = R, \hat{\sigma}_{Bi}(R_s, R_t) = \hat{\sigma}_{Bi}^2(R) \). These results also hold for subgraphs \( R \), which are \( r \)-cycles.

4.1. Proof of Theorem 4.1. The proof follows from the lemma and theorems of the previous section. Since, we have \( \sqrt{n} \)-consistent bootstrap estimators of \( \rho - \epsilon \bar{P}_{Bi}(R) \) for \( i = 1, 2 \). Now, from the Theorem 1(a) in [6], we know that as \( \lambda_n \to \infty \), if \( \hat{\rho}_n = \frac{D_n}{n-1} \) as defined in (2.6),

\[
\frac{\hat{\rho}_n}{\rho_n} \xrightarrow{P} 1 \\
\sqrt{n} \left( \frac{\hat{\rho}_n}{\rho_n} - 1 \right) \xrightarrow{w} N(0, \sigma^2)
\]

Now, we define the bootstrap estimates in (4.1). So, we get by applying Slutsky’s Theorem that

\[
\sqrt{n} \left( \hat{T}_{Bi}(R) - \bar{P}(R) \right) \xrightarrow{P} 0 \quad \text{for} \quad i = 1, 2.
\]

The statement about bootstrap estimate of variance follows from Lemma 8 and the definitions of bootstrap variance in the form of equation (3.13).

So, we have \( \sqrt{n} \)-consistent bootstrap estimators, \( \hat{T}_{Bi}(R) \) (for \( i = 1, 2 \)) of \( \hat{T}(R) \) and consistent estimators, \( \hat{\sigma}^2_{Bi}(R) \) (for \( i = 1, 2 \)) of \( \sigma^2(R) \). Also from Theorem 1 of [6], we have, for subgraphs \( R_1, \ldots, R_k \) of \( G_n \),

\[
\sqrt{n} \left( (\hat{T}(R_1), \ldots, \hat{T}(R_k)) - (\bar{P}(R_1), \ldots, \bar{P}(R_k)) \right) \xrightarrow{w} N(0, \Sigma(R))
\]

So, we can combine the result from Theorem 3.1 and 3.2 with the above theorem, using Slutsky and convergence of types theorem, to get the symptomatic normality behavior of \( \hat{T}_{Bi}(R) \). As \( n \to \infty, \lambda_n \to \infty \) and under conditions of Theorem 3.1 and 3.2, if we define, \( T_{Bi}(R) = (\hat{T}_{Bi}(R_1), \ldots, \hat{T}_{Bi}(R_k)) \) and \( P(R) = (\bar{P}(R_1), \ldots, \bar{P}(R_k)) \)

\[
\sqrt{n} \left( (T_{Bi}(R) - P(R))^T \hat{\Sigma}_{Bi}^{-1/2}(R) (T_{Bi}(R) - P(R)) \right) \xrightarrow{w} N(0, I) \quad \text{for} \quad i = 1, 2
\]
where, $[\hat{S}_{Bi}]_{st} = \hat{\sigma}_{Bi}(R_s, R_t)$, $s, t = 1, \ldots, k$ and if $R_s = R_t = R$, $\hat{\sigma}_{Bi}(R_s, R_t) = \hat{\sigma}_{Bi}^2(R)$ for $i = 1, 2$.

5. Simulation Results. We apply the two representative bootstrap subsampling schemes for simulated datasets to find out their performances. We generate data from two different simulation models. Both models are special cases of the nonparametric model described in [5]. The two models that we consider are:

- Stochastic block model
- Preferential attachment model.

For each of the models, we try to find the estimate of the count statistics features and their confidence intervals through bootstrap subsampling. The features that we consider are generalized $(k, l)$-wheels, $p$-cycles and a smooth function of count statistics, transitivity.

5.1. Count Statistics. In these simulations, the main class of acyclic features we consider are $(k, l)$-wheels. We also consider the count of the cyclic patterns such as triads or triangles or 3-cycles and tetrads or quadrilaterals or 4-cycles. We also consider a smooth function of counts of triangle and $(1, 2)$-wheel, known as transitivity, $\hat{P}_{Tr}$, defined in Example 3.

5.2. Stochastic Block Model. Let $w$ correspond to a $K$-block model defined by parameters $\theta = (\pi, \rho_n, S)$, where $\pi_a$ is the probability of a node being assigned to block $a$ as before, and

$$F_{ab} = \mathbb{P}(A_{ij} = 1| i \in a, j \in b) = s_n S_{ab}, \quad 1 \leq a, b \leq K.$$ 

and the probability of node $i$ to be assigned to block $a$ to be $\pi_a$ ($a = 1, \ldots, K$).

We consider a stochastic block model with $K = 2$, $S = \begin{pmatrix} 0.4 & 0.5 \\ 0.4 & 0.7 \end{pmatrix}$, $s_n = \frac{5\nu\sqrt{n}}{n}$ and $\pi = (0.5, 0.5)$. So, we get $\rho_n = \pi^T F \pi$. First, we keep $n = 1000$ fixed and vary $\nu$ such that $\rho_n$ varies from 10 to 100. Second, we vary $\nu$ fixed at 0.5 and vary $n = 500$ to 3000.

In the following figures, we try to see the behavior of mean and variances of the count statistics. In Figure 1(a)-(d), we compare the bootstrap mean and standard deviation of the normalized count statistics, $\hat{T}(R)$, where, $R$ is a 4-cycle and $\hat{T}_{Tr}$. We find the bootstrap mean estimator for both bootstrap schemes and we also find the corresponding estimates of variance by bootstrap. We then plot the plot the parameter, $\hat{P}(R)$, estimator $\hat{T}_{Bi}(R)$.
along with the asymptotic 95% confidence interval using the asymptotic normality result of Theorem 4.1 and the bootstrap estimates of variance \( \hat{\sigma}_{Bi}^2(R) \). The bootstrap estimate of asymptotic variance of \( \hat{T}_{Bi}^R \) is obtained from the bootstrap estimates of \( \hat{\sigma}_{Bi}^2(R_1), \hat{\sigma}_{Bi}^2(R_2) \) and \( \hat{\sigma}_{Bi}(R_1, R_2) \) by using Delta method and using the Theorem 4.1. In Figure 1(a)-(b), we keep \( n \) fixed but vary \( \lambda_n \) from 10 to 100 by varying \( \nu \) and In Figure 1(c)-(d), we keep \( \nu \) fixed and vary \( n \) from 500 to 3000.

![Figure 1](image)

**Figure 1.** Stochastic block model: For \( n = 1000 \), we vary average degree (\( \lambda_n \)) and (a) Plot estimated normalized tetrad count (b) Plot estimated Transitivity and their 95% Confidence Interval (CI), where, CI is estimated using bootstrap estimates of variance of the estimators. For \( \nu = 0.5 \), we vary \( n \) and (c) Plot estimated normalized tetrad count (d) Plot estimated Transitivity and their 95% Confidence Interval (CI). We use different colors to indicate different bootstrap subsampling schemes and graph parameter.

In Figure 2, we compare the mean of the bootstrap estimates with the
parameter $\tilde{P}(R)$. In Figure 2(a), we keep $n$ fixed but vary $\lambda_n$ from 10 to 100 by varying $\nu$ and in Figure 2(b), we keep $\nu$ fixed and vary $n$ from 500 to 3000. So, we get reasonable estimates of integral parameters of graph as we vary average degree and number of vertices of the graph.

![Figure 2](image)

**Figure 2.** Stochastic block model: For $n = 1000$, we vary average degree ($\lambda_n$) and (a) plot estimated normalized $(1,6)$-wheel count. For $\nu = 0.5$, we vary $n$ and (b) plot estimated normalized $(1,6)$-wheel count. We use different colors to indicate different bootstrap subsampling schemes and graph parameter.

In Figure 3, we compare the variance of the bootstrap estimates, based on bootstrap iterations for both the bootstrap schemes. We see that bootstrap variance is usually lower for the subgraph subsampling scheme as we increase number of vertices of the graph for different count statistics.

### 5.3. Preferential Attachment Model

In Preferential Attachment Model, given $k$ initial vertices, $k + 1$th vertex attach to one of the preceding $k$ vertices with probability proportional to degree. Now, we have degree of vertex $v$ defined as $D_v$ and $\bar{D} = \frac{1}{n} \sum_{v=1}^{n} D_v$. Also, we have,

$$\tau(v) \simeq \frac{D_v}{\bar{D}}$$

So, following Eq. (2.1), we have the probability of edge formation as

$$w(u, v) = \frac{\tau(u)}{T(u)} 1(u \leq v) + \frac{\tau(v)}{T'(v)} 1(v \leq u)$$

where, $T(u) = \int_{u}^{1} \tau(s)ds$ and $T'(v) = 1 - T(v)$ and

$$\tau(u) = \int_{0}^{u} w(u, v)dv$$
So, the preferential attachment model can be defined by the following formula on \( w \)

\[
w(u, v) = \frac{\tau(u)}{\int_u^1 \tau(s)ds} 1(u \leq v) + \frac{\tau(v)}{\int_v^1 \tau(s)ds} 1(v \leq u)
\]

Thus, for \( w(u, v) = (1-u)^{-1/2}(1-v)^{-1/2} \)

we have,

\[
\tau(v) = c(1-v)^{-1/2}
\]

which is equivalent to power law of degree distribution \( F \equiv \tau^{-1} \).

In the following figures, we try to see the behavior of mean and variances of the count statistics. In Figure 4(a)-(d), we compare the bootstrap mean and standard deviation of the normalized count statistics, \( \hat{T}(R) \), where, \( R \) is a 4-cycle and \( \hat{T}_{TR} \) just as we did in Figure 1. In Figure 4(a)-(b), we keep \( n \) fixed but vary \( \lambda_n \) from \( \lambda_n \) 10 to 100 and In Figure 4(c)-(d), we vary \( n \) from 500 to 3000 and \( \lambda_n \) varies as constant. \( \sqrt{n} \).

5.4. Comparison Between Stochastic Block Model and Preferential Attachment Model. We simulate networks from both stochastic block models
Figure 4. Preferential attachment model: For $n = 1000$, we vary average degree ($\lambda_n$) and (a) Plot estimated normalized tetrad count (b) Plot estimated Transitivity and their 95% Confidence Interval (CI), where, CI is estimated using bootstrap estimates of variance of the estimators. For $\nu = 0.5$, we vary $n$ and (c) Plot estimated normalized tetrad count (d) Plot estimated Transitivity and their 95% Confidence Interval (CI). We use different colors to indicate different bootstrap subsampling schemes and graph parameter.
Figure 5. (a) For $n = 1000$ we vary $\lambda_n$ and we plot estimated Transitivity $\hat{\mathcal{T}}_{B^2}^{\mathcal{R}}$ and their 95% Confidence Interval (CI), where, CI is estimated using bootstrap estimates of variance of the estimators. (b) We vary $n$ and we plot estimated normalized tetrad count, $\hat{\mathcal{T}}_{B^2}(R)$, $R = \text{tetrad}$ and their 95% Confidence Interval (CI).

and preferential attachment models and then, we try to compare the distribution of count statistics of the graph for two different networks. In Figure 5(a) we vary the parameters of SBM as $F = \mu F^{(1)} + (1 - \mu) F^{(2)}$, where, $F^{(1)} = \text{Diag}(0.035, 0.065$) and $F^{(2)} = 0.001 \mathbf{1}_2$. We increase $\mu$ to increase $\hat{\lambda}_n$ and SBM have more pronounced cluster structure. We keep the average degree, $\lambda_n$, of the two simulated networks same and then, we try to get the asymptotic distribution of the transitivity statistic for the two cases for each $\lambda_n$. We see here that, for low $\lambda_n$, we can not statistically distinguish between the transitivity of networks generated from two different models, but, they become statistically distinguishable as average degree, $\lambda_n$ and $\mu$, increases. In Figure 5(b), take SBM as in 5.2 and PFA as in 5.3 keeping the average degree, $\lambda_n$, of the two simulated networks same and vary $n$ and we can statistically distinguish the normalized tetrad count of networks between the two different models for large $n$ based on subgraph subsampling scheme.

6. Real Data Examples. Social networks recently has become quite large after the introduction of social networking sites. We consider two different social networks as a platform for our experiments. The first one, High School Romantic Relations data is a small social network, whereas the second one, Facebook College Social Network has greater number of nodes and links. For both cases we use subgraph subsampling bootstrap scheme.

6.1. High School Network. In this application, we try to quantitatively verify some of the hypothesis mentioned by the authors in the paper [3] when
presenting the data. The network here is formed by students of Jefferson High school as nodes and if two students have romantic relations then there exists a link between those two nodes. In the paper, [3], where the data was presented, an observation was made about the dearth of short cycles in the network. Our application here is trying to answer the question whether the absence of short cycles in this graph is significant or not. We consider a very simple model for the data.

We consider the data has been generated from two different models -

(a) Stochastic block model with two blocks (Male and Female) and the connection probability matrix is given by

\[ P = \begin{pmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{pmatrix} \]

where, \( P_{ab} = \text{Average number of edges between blocks } a \text{ and } b \text{ in the network, where, } a, b = 1, 2 \text{ are the two blocks with Male } = 1 \text{ and Female } = 2 \). In this network, we have \( P_{11} = 0, P_{12} = 0.0058 \text{ and } P_{22} = 0.000025 \). The probability of belonging to the two blocks are \((0.497, 0.503)\).

(b) Preferential attachment model with \( \rho = \frac{\lambda_n}{n} \), where, \( \lambda_n = \text{Average degree of the network } = 1.66 \text{ and } n \text{ is the number of nodes.} \)

Now, for these two simple models, we can theoretically find the normalized count of small cycles. Then, we can perform a hypothesis test to find out whether the number of small cycles we see in this network is significantly small or not. For both the models, we can find \( \tilde{P}(R) \), where, \( R = \text{cycles of size 3 and 4 based on the parameters defined for the models in (a) and (b) and using the equation (2.4) and we shall call it } \hat{P}_0(R) \). Also, for the network, the unknown integral parameter for the subgraph \( R \) is \( \tilde{P}(R) \). Formally, the hypothesis becomes

\[ H_0 : \tilde{P}(R) = \hat{P}_0(R) \text{ vs } \tilde{P}(R) < \hat{P}_0(R) \]

for each \( R \) and for each of the models (a) and (b). We use the results of Theorem 4.1 to form the asymptotic test. The results are given in Table 1. We see in the results that, according to the two simple models, it is extremely unlikely for 3-cycles and 4-cycles to occur in the graph. In fact, the original network has too many 4-cycles short cycles not too few. This is an interesting observation coming out of our simple exploratory analysis.

So, our simple models do not capture the probabilistic mechanism of the original network correctly and we need to analyze the short cycles in the network more closely to understand their formation.
Table 1

The normalized subgraph counts, their standard deviation and the expected counts from the stochastic block model (SBM) and preferential attachment model (PFA) for the whole high school network.

| Subgraph     | Normalized Count | Standard Deviation | Count (SBM) | Count (PFA) |
|--------------|------------------|--------------------|-------------|-------------|
| (1,2)-wheel  | 2.27             | 0.17               | 1.01        | 2.97        |
| 3-cycle      | 1.31             | 0.10               | 0.01        | 1.04        |
| 4-cycle      | 9.47             | 3.16               | 0.63        | 3.06        |

Note that, this is a very small and sparse network. For this network, the use of Theorem 1 from [6] would have sufficed, but we give the example as an example of the use of count statistics and their quantitative behavior. Here in the paper [3] simulation based tests were used. We use asymptotic Gaussian tests and we can directly answer the questions without the possible theoretical pitfalls of permutation and simulation based tests.

6.2. Facebook Network. In this application, we try to quantitatively analyze the behavior of some of the known descriptive statistics for Facebook collegiate networks. The networks were presented in the paper by Traud et.al. (2011) [30]. The network is formed by Facebook users acting as nodes and if two Facebook users are “friends” there is an edge between the corresponding nodes. Along with the network structure, we also have the data on covariates of the nodes. Each node has covariates: gender, class year, and data fields that represent (using anonymous numerical identifiers) high school, major, and dormitory residence. We try to answer two very basic questions quantitatively for these networks -

1. Can the node covariates act as cluster identifiers?
2. Can two college networks be distinguishable in terms of some basic descriptive statistics?

In order to address the first question, we consider the network of a specific college (Caltech). We consider the covariates class year, major and dormitory residence as our covariates of interest. We take the induced network created by levels of each of these covariates and try to see if those networks have different clustering properties. For example, consider class year and major as the covariates of interest. We consider the nodes belonging two different class years and find their induced network from the whole collegiate network. Similarly, we consider the nodes belonging two different majors and find their induced network from the whole collegiate network. Now, we have two different networks, one of which has nodes coming exclusively from two different class years and another has nodes coming exclusively from
two different majors. We now try to find which of two networks is more “clustered” by comparing transitivity of the two networks. We can repeat the same exercise for any two covariates and choosing a subset of their levels. For the two networks, the unknown integral parameter for transitivity is $\tilde{P}_{Tr}^1$ and $\tilde{P}_{Tr}^2$ respectively. Formally, the hypothesis becomes

$$H_0 : \tilde{P}_{Tr}^1 = \tilde{P}_{Tr}^2 \text{ vs } \tilde{P}_{Tr}^1 \neq \tilde{P}_{Tr}^2.$$

The second question can also be answered in the similar spirit as the first one. We consider the full collegiate network of two different colleges (Caltech and Princeton). Then, we try to compare the transitivity of these two collegiate networks. For the two networks, the unknown integral parameter for transitivity is $\tilde{P}_{Tr}^1$ and $\tilde{P}_{Tr}^2$ respectively. Formally, the hypothesis becomes

$$H_0 : \tilde{P}_{Tr}^1 = \tilde{P}_{Tr}^2 \text{ vs } \tilde{P}_{Tr}^1 \neq \tilde{P}_{Tr}^2.$$

These comparisons could in principle be possible using the results given in Bickel et al. (2011) [6], but computationally intractable. Using bootstrap estimators, we can estimate the variance of the estimators and thus perform hypothesis testing in reasonable time.

In Tables 6.2, 3 and 4, we present an excerpt of the result of our analysis and answer both the questions. The results give a better understanding about the network statistics reported in [30], like those reported in Table of [30], as results relying only on estimates without any indication of variance of estimates can be misleading and unreliable.

| Class Year(CY) | Dormitory(DM) | Major(MJ) |
|----------------|---------------|-----------|
| Estimated Transitivity | 0.15 | 0.22 | 0.12 |

Table 2

Transitivity of induced networks formed by considering only two levels of a specific covariate of a specific collegiate network.

| Difference     | CY and DM | DM and MJ |
|----------------|-----------|-----------|
| Estimated      | 0.07      | 0.1       |
| Estimated SD   | 0.05      | 0.035     |

Table 3

The Difference between Class Year and Dorm is not significant but difference between Dorm and Major is significant by asymptotic normal test at 5% level. The data was presented in Traud et. al. (2011) SIAM Review.

Now, without finding the bootstrap estimate of count statistics and its variance, finding the asymptotic distribution of these count statistics will
|                      | Network 1 | Network 2 |
|----------------------|-----------|-----------|
| Estimated Transitivity| 0.29      | 0.16      |
| Estimated Difference  | 0.13      |           |
| Estimated Difference SD| 0.11      |           |

**Table 4**

The difference of transitivity between two networks is not significant by asymptotic normal test at 5% level. Therefore Network 1 can not be said to be more 'clusterable'.

The data was presented in Traud et. al. (2011) SIAM Review.

not have been possible. So, with the help of the bootstrap based estimates we can perform hypothesis testing on the count statistics and provide the estimates of their asymptotic distribution.

7. **Conclusion and Future Works.** In this paper, we have considered two known subsampling schemes of networks and tried to show situations, where, they are applicable to find the asymptotic distribution of certain count statistics of the network. We showed that the normalized bootstrap subsample estimates of the count statistics and their smooth functions have asymptotic normal distribution. We proposed bootstrap schemes by which we could efficiently compute the asymptotic mean and variance of these count statistics. We also showed that the Subgraph sampling bootstrap scheme seemed most stable and we recommend that scheme for use as bootstrap subsampling scheme in most cases.

We also use the estimated asymptotic mean and variances of the count statistics to construct hypothesis tests. These hypothesis tests can serve several purposes, such as

(a) Distinguish between the count statistics of two different networks
(b) Distinguish between parts of same network
(b) Testing whether a network has been generated from a specific model by comparing the empirical and population version of the count statistic.
(c) Testing how close parameters of two different network models can become.

All of these different qualitative tests can be made quantitative by using hypothesis tests using the count statistics. Using subsample bootstrap estimates of count statistics, we show from simulations that transitivity of networks from stochastic block models becomes easier to differentiate from transitivity of preferential attachment model as average degree grows. Similarly, in real networks, such as Facebook collegiate network, we show that certain covariate based subnetworks have more ‘cluster’ structure than others. Also, even in large networks, conclusions based on means only as opposed
to confidence statements using variances could be unreliable.

7.1. Future Works. One natural generalization can be use of bootstrap scheme to get asymptotic distribution of global statistics - such as graph cut, conductance, functionals of graphon (non-integral functionals) and such parameters. Sample and bootstrap estimates of such parameters are sometimes obtainable, but their theoretical properties are still unknown. It would be a nice future endeavor to extend bootstrap subsampling scheme to estimate such global characteristics of the networks.

Acknowledgements. We thank Aiyou Chen, Dave Choi and Liza Levina for helpful discussions and comments.

Appendix.

A1. Variance of $\hat{P}_{b1}(R)$. The variance of $\hat{P}_{b1}(R)$ is

$$
\text{Var}_b \left[ \frac{1}{m\left|\text{Iso}(R)\right|} \sum_{S \subseteq K_m, S \supseteq R} \mathbf{1}(S \subseteq H) \bigg| G \right] = \left( \frac{1}{m\left|\text{Iso}(R)\right|} \right)^2 \text{Var}_b \left[ \sum_{S \subseteq K_m, S \supseteq R} \mathbf{1}(S \subseteq H) \bigg| G \right]
$$

$$
\text{Var}_b \left[ \sum_{S \subseteq K_m, S \supseteq R} \mathbf{1}(S \subseteq H) \bigg| G \right] = \mathbb{E}_b \left[ \left( \sum_{S \subseteq K_m, S \supseteq R} \mathbf{1}(S \subseteq H) \right)^2 \bigg| G \right] - \left( \mathbb{E}_b \left[ \sum_{S \subseteq K_m, S \supseteq R} \mathbf{1}(S \subseteq H) \bigg| G \right] \right)^2
$$

$$
\mathbb{E}_b \left[ \left( \sum_{S \subseteq K_m, S \supseteq R} \mathbf{1}(S \subseteq H) \right)^2 \bigg| G \right] = \mathbb{E}_b \left[ \sum_{S \subseteq K_m, S \supseteq R} \mathbf{1}(S \subseteq H) \bigg| G \right] + \mathbb{E}_b \left[ \sum_{S,T \subseteq K_m, S \supseteq R, S \neq T} \mathbf{1}(S, T \subseteq H) \bigg| G \right]
$$

$$
= I + II \quad \text{(Suppose)}
$$
Thus,

\[
I = \sum_{S \subseteq K_n, S \not\subseteq R} \binom{n-p}{m-p} \mathbf{1}(S \subseteq G)
\]

\[
II = \mathbb{E}_b \left[ \sum_{S,T \subseteq K_n \atop S,T \not\subseteq R, S \neq T} \mathbf{1}(S,T \subseteq H) \bigg| G \right]
\]

Now, a host of subgraphs can be formed by the intersection of two copies of \( R \). The number of intersected vertices can range from 0 to \( p - 1 \). Let us consider, that for number of vertices in intersection as \( k \) (\( k = 1, \ldots, (p - 1) \)), the number of graph structures that can be formed is \( g_k \) and we represent that graph structure by \( W_{jk} \), where, \( j = 1, \ldots, g_k \). Thus,

\[
II = \sum_{k=0}^{p-1} g_k \sum_{j=1}^{g_k} \sum_{S,T \subseteq K_n \atop S,T \not\subseteq W_{jk}} \binom{n-(2p-k)}{m-(2p-k)} \binom{n}{m} \mathbf{1}(S \subseteq G)
\]

So,

\[
\mathbb{E}_b \left[ \left( \sum_{S \subseteq K_n, S \not\subseteq R} \mathbf{1}(S \subseteq H) \right)^2 \bigg| G \right] = \sum_{S \subseteq K_n, S \not\subseteq R} \binom{n-p}{m-p} \mathbf{1}(S \subseteq G) + \sum_{k=0}^{p-1} g_k \sum_{j=1}^{g_k} \sum_{S \subseteq K_n, S \not\subseteq W_{jk}} \binom{n-(2p-k)}{m-(2p-k)} \binom{n}{m} \mathbf{1}(S \subseteq G)
\]

\[
\text{Var}_b \left[ \sum_{S \subseteq K_n, S \not\subseteq R} \mathbf{1}(S \subseteq H) \bigg| G \right] = \sum_{S \subseteq K_n, S \not\subseteq R} \binom{n-p}{m-p} \binom{n}{m} \mathbf{1}(S \subseteq G) + \sum_{k=0}^{p-1} g_k \sum_{j=1}^{g_k} \sum_{S \subseteq K_n, S \not\subseteq W_{jk}} \binom{n-(2p-k)}{m-(2p-k)} \binom{n}{m} \mathbf{1}(S \subseteq G) - \left( \sum_{S \subseteq K_n, S \not\subseteq R} \binom{n-p}{m-p} \mathbf{1}(S \subseteq G) \right)^2
\]
\[ \text{Var}_b \left[ \frac{1}{\binom{m}{p}|\text{Iso}(R)|} \sum_{S \subseteq K_n, S \not\subseteq R} 1(S \subseteq H) \right] \\
= \left( \frac{1}{\binom{m}{p}|\text{Iso}(R)|} \right)^2 \left[ \sum_{S \subseteq K_n, S \not\subseteq R} \frac{(n-p)}{\binom{n}{m}} 1(S \subseteq G) \right] \\
- \left( \frac{1}{\binom{m}{p}|\text{Iso}(R)|} \right)^2 \left[ \left( \sum_{S \subseteq K_n, S \not\subseteq R} \frac{(n-p)}{\binom{n}{m}} 1(S \subseteq G) \right)^2 \right] \\
+ \left( \frac{1}{\binom{m}{p}|\text{Iso}(R)|} \right)^2 \left[ \sum_{k=0}^{p-1} \sum_{j=1}^{g_k} \sum_{S \subseteq K_n, S \not\subseteq W_{jk}} \frac{(n-(2p-k))}{\binom{n}{m}} 1(S \subseteq G) \right] \\
\]

So,

\[ \text{Var}_b \left[ \hat{P}_{b1}(R) \right] = \left( \frac{1}{\binom{m}{p}|\text{Iso}(R)|} \right)^2 \left[ \sum_{S \subseteq K_n, S \not\subseteq R} \frac{(n-p)}{\binom{n}{m}} 1(S \subseteq G) \right] \\
- \left( \frac{1}{\binom{m}{p}|\text{Iso}(R)|} \right)^2 \left[ \left( \sum_{S \subseteq K_n, S \not\subseteq R} \frac{(n-p)}{\binom{n}{m}} 1(S \subseteq G) \right)^2 \right] \\
+ \left( \frac{1}{\binom{m}{p}|\text{Iso}(R)|} \right)^2 \left[ \sum_{k=0}^{p-1} \sum_{j=1}^{g_k} \sum_{S \subseteq K_n, S \not\subseteq W_{jk}} \frac{(n-(2p-k))}{\binom{n}{m}} 1(S \subseteq G) \right] \\
\]

**A2. Proof of Theorem 3.1.**

**Proof.** (i) Now, let us try to try to find the expectation of \( \hat{P}_{b1}(R) \)
under the sampling distribution conditional on the given data $G$. 

\[
\mathbb{E}_b \left[ \frac{1}{\binom{m}{p} |\text{Iso}(R)|} \sum_{S \subseteq K_m, S \supseteq R} 1(S \subseteq H) \mid G \right] = \frac{1}{\binom{m}{p} |\text{Iso}(R)|} \mathbb{E}_b \left[ \sum_{S \subseteq K_m, S \supseteq R} 1(S \subseteq H) \mid G \right] = \frac{1}{\binom{m}{p} |\text{Iso}(R)|} \sum_{S \subseteq K_m, S \supseteq R} \frac{1}{\binom{n}{m}} 1(S \subseteq H)
\]

\[
= \frac{1}{\binom{m}{p} |\text{Iso}(R)|} \sum_{S \subseteq K_n, H \supseteq S, H \subseteq G} \frac{1}{\binom{n}{m}} 1(S \subseteq G)
\]

\[
= \frac{1}{\binom{m}{p} |\text{Iso}(R)|} \sum_{S \subseteq K_n, S \supseteq R} \frac{1}{\binom{n}{m}} 1(S \subseteq G)
\]

So, we have,

\[
\mathbb{E}_b[\hat{P}_{B1}(R) \mid G] = \mathbb{E}_b[\hat{P}_{b1}(R) \mid G] = \hat{P}(R)
\]

(ii) Given $G$,

\[
\text{Var}_b[\rho^{-e} \hat{P}_{B1}(R) \mid G] = \rho^{-2e} \frac{1}{B^2} \left( \sum_{b=1}^{B} \text{Var}_b[\hat{P}_{b1}(R)] + \sum_{b,b' = 1, b \neq b'} \text{Cov}_b(\hat{P}_{b1}(R), \hat{P}_{b'1}(R)) \right)
\]

Now, the formula for $\text{Var}_b[\hat{P}_{b1}(R)]$ from A1 we get that

\[
\text{Var}_b \left[ \rho^{-e} \hat{P}_{b1}(R) \right] = \left( \frac{\rho_n^{-e}}{\binom{m}{p} |\text{Iso}(R)|} \right)^2 \sum_{S \subseteq K_n, S \supseteq R} \frac{1}{\binom{n}{m}} 1(S \subseteq G)
\]

\[
- \left( \frac{\rho_n^{-e}}{\binom{m}{p} |\text{Iso}(R)|} \right)^2 \left[ \sum_{S \subseteq K_n, S \supseteq R} \frac{1}{\binom{n}{m}} 1(S \subseteq G) \right]^2
\]

\[
+ \left( \frac{\rho_n^{-e}}{\binom{m}{p} |\text{Iso}(R)|} \right)^2 \sum_{k=0}^{p-1} \sum_{j=1}^{g_k} \sum_{S \subseteq K_n, S \supseteq W_{jk}} \frac{1}{\binom{n}{m}} 1(S \subseteq G)
\]


Here, we use properties of the underlying model. Let us condition on similar steps as variance in Appendix A1. If we consider the univ-

So,

\[ W = \sum_{b} \rho_{n}^{\xi} \]

\[ \mathbb{E}(\rho^{-e} \hat{P}(R) | \xi) = \frac{1}{\binom{m}{p}} \sum_{S \subseteq K_{n}, S \subseteq R} \frac{1}{|S|} \prod_{(i,j) \in E(S)} w(\xi_{i}, \xi_{j}) + O(n^{-1} \lambda_{n}). \]

We shall use the same decomposition as used in [6] of \((\rho_{n}^{-e} \hat{P}_{B1}(R) - \hat{P}(R))\) into

\[
(\rho_{n}^{-e} \hat{P}_{B1}(R) - \hat{P}(R)) = \rho_{n}^{-e} (\hat{P}_{B1} - \mathbb{E}[\hat{P}_{b1}(R)|G]) \\
+ \rho_{n}^{-e} (\hat{P}(R) - \mathbb{E}(\hat{P}(R)|\xi)) \\
+ \mathbb{E}(\hat{P}(R)|\xi) \rho_{n}^{-e} - \hat{P}(R)
\]

Let us define,

\[ U_{3} = \mathbb{E}(\hat{P}(R)|\xi) \rho_{n}^{-e} - \hat{P}(R) \]

\[ U_{2} = \rho_{n}^{-e} (\hat{P}(R) - \mathbb{E}(\hat{P}(R)|\xi)) \]

\[ U_{1} = \rho_{n}^{-e} (\hat{P}_{B1} - \mathbb{E}[\hat{P}_{b1}(R)|G]) \]

Now, it is easy to see that

\[
\text{Var}(\rho^{-e} \hat{P}_{B1}(R)) = \mathbb{E}(\text{Var}(\rho^{-e} (\hat{P}_{B1}(R)|G))) + \text{Var}(\mathbb{E}(\rho^{-e} \hat{P}_{B1}(R)|G)) \\
= \mathbb{E}(\text{Var}(\rho^{-e} \hat{P}_{B1}(R) - \hat{P}(R)|G)) + \text{Var}(\hat{P}(R)) \\
= \mathbb{E}(\text{Var}(U_{1}|G)) + \text{Var}(\mathbb{P}(R)|\xi)) + \text{Var}(\mathbb{E}(\hat{P}(R)|\xi)) \\
= \mathbb{E}(\text{Var}(U_{1}|G)) + \text{Var}(\mathbb{E}(U_{2}|\xi)) + \text{Var}(U_{3})
\]

We shall try to see the behavior of \(\text{Var}(U_{1}|G) = \text{Var}_{b}[\rho^{-e} \hat{P}_{B1}(R)|G].\)

From (ii) we get that, \(\text{Var}_{b}[\rho_{n}^{-e} \hat{P}_{b1}(R)] = O \left( \frac{1}{m^{3}\rho_{n}} \vee \frac{1}{m} \right).\) Similarly,

\[ \text{Cov}_{b}[\rho_{n}^{-e} \hat{P}_{b1}(R), \rho_{n}^{-e} \hat{P}_{b1}(R)] = O(\frac{1}{m}) \] for acyclic and \(k\)-cycle \(R\) following similar steps as variance in Appendix A1. If we consider the uniform probability for bootstrap to be \(\gamma\), then, \(B = O(\gamma n^{p}).\) Note that,
if \( E(H_b) \cap E(H_{b'}) = \phi \), then, \( \text{Cov}_b(\hat{P}_{b1}(R), \hat{P}_{b'1}(R)) = 0 \). The number of pairs such that \( E(H_b) \cap E(H_{b'}) \neq \phi \) is \( O(m^2 \gamma^2 n^{2m-2}) \). Also, the number of edges for the leading term in the covariance is equal to or more than \( 2e \). So,

\[
\text{E}(\text{Var}_b[\rho - e \hat{P}_{B1}(R)|G]) = O \left( \frac{1}{B(m^p \rho_n \land m)} + O \left( \frac{m^2 \gamma^2 n^{2m-2}}{m \gamma^2 n^{2m}} \right) \right)
\]

\[
= O \left( \frac{1}{B(m^p \rho_n \land m)} + \frac{m}{m^2} \right) = O \left( \frac{1}{B(m^p \rho_n \land m)} \right) + o(n^{-1})
\]

The second equality follows since we have \( m/n \to 0 \) as \( n \to \infty \). So, since, \( B(m^p \rho_n \land m) > O(n) \), we have, \( \text{E}(\text{Var}(U_1|G)) = o(n^{-1}) \).

Now, by proof of Theorem 1 in [6], we have,

\[
\text{Var}(U_2) = o(n^{-1})
\]

\[
\text{Var}(U_3) = o(n^{-1})
\]

So, we get, \( \text{Var}(\rho - e \hat{P}_{B1}(R)) = o(n^{-1}) \). Since, we already know \( \sqrt{n} \)-consistency of \( \rho_n^{-e} \hat{P}(R) - \hat{P}(R) \), this proves the \( \sqrt{n} \)-consistency of \( \rho_n^{-e} \hat{P}_{B1}(R) \) to \( \rho_n^{-e} \hat{P}(R) \).

\( \square \)

**A3.** Proof of Theorem 3.2. For variance calculation, we also need the joint inclusion probability of two items, \( S, S' \in \mathcal{S}_p \), which are subgraphs of \( G \) induced by the set of vertices \( \{w_1, \ldots, w_p\} \) and \( \{w'_1, \ldots, w'_p\} \) respectively, where, we take that \( w_{i+1} \succ w_i \) and \( w'_{i+1} \succ w'_i \), \( i = 1, \ldots, p - 1 \). So,

\[
\pi_{SS'} \equiv \text{Inclusion Probability of } S \text{ and } S' = P[(w_1, \ldots, w_p) \text{ is selected and } (w'_1, \ldots, w'_p) \text{ is selected}]
\]

\[
= \prod_{d=1}^{p} (q_d) \frac{z_{1d}}{z_{2d}} \prod_{d=1}^{p} (q_d) \frac{z_{2d}}{z_{2d}}
\]

where,

\[
z_{1d} = \begin{cases} 1(w_d = w'_d), & \text{for } d = 1 \\ 1((w_d, w_{d-1}) = (w'_d, w'_{d-1})), & \text{for } d = 2, \ldots, p \end{cases}
\]

\[
z_{2d} = \begin{cases} 1(w_d \neq w'_d), & \text{for } d = 1 \\ 1((w_d, w_{d-1}) \neq (w'_d, w'_{d-1})), & \text{for } d = 2, \ldots, p \end{cases}
\]
(i) We know that \( \hat{P}_{b2}(R) \) is a Horvitz-Thompson estimator with inclusion probability of each population unit to be \( \pi = \prod_{d=1}^{p} q_d \). So, according to the sampling theory [28], \( \hat{P}_{b2}(R) \) is an unbiased estimator of \( \hat{P}(R) \) given the network \( G \), if \( \Pr(\hat{P}_{b2}(R) = 0 | \hat{P}(R)) \rightarrow 0 \) as \( n \rightarrow \infty \). Now, \( \Pr(\hat{P}_{b2}(R) = 0 | \hat{P}(R)) \leq (1 - q_d) \lambda_n \) for all \( d = 1, \ldots, p \). For all \( d = 1, \ldots, p \), \( (1 - q_d) \lambda_n \rightarrow 0 \) if \( \lambda_n q_d \rightarrow \infty \). So, under the condition, \( \lambda_n q_d \rightarrow \infty \) and \( q_d \rightarrow 0 \) as \( n \rightarrow \infty \), we have, \( \hat{P}_{b2}(R) \) is an asymptotically unbiased estimator of \( \hat{P}(R) \).

(ii) The variance of \( \hat{P}_{b2}(R) \) coming from the bootstrap sampling only is given by

\[
\text{Var}_b \left[ \hat{P}_{b2}(R) \right] = \frac{1}{N^2} \left[ \frac{1 - \pi}{\pi} \sum_{S \in S_p} 1(S \equiv R) \right.

+ \left. \sum_{S, S' \in S_p, S \neq S'} \frac{\pi_{SS'} - \pi^2}{\pi^2} 1(S \equiv R, S' \equiv R) \right]
\]

where,

\[ N = \rho_n \left( \frac{n}{p} \right) |Iso(R)| \]

From the formula of \( \text{Var}_b[\hat{P}_{b2}(R)|G] \), we see that, the covariance terms vanishes when \( \pi_{SS'} = \pi^2 \). Now, if \( q_1 = 1 \), then, \( \pi_{SS'} = \pi^2 \) if \( E(S) \cap E(S') = \phi \). The number of pairs such that \( E(S) \cap E(S') \neq \phi \) is \( O(p^2 n^{2p-2}) \).

Now, the condition of \( q_1 = 1 \) is a bit restrictive. In stead, if we have \( q_1 \rightarrow 1 \) as \( n \rightarrow \infty \), then, the highest order term of covariance term comes from the case when \( E(S) \cap E(S') \neq \phi \) but the root nodes are same that is \( w_1 = w'_1 \). So, for some constant \( C > 0 \),

\[
\frac{1}{N^2} \sum_{S, S' \in S_p, S \neq S'} \frac{\pi_{SS'} - \pi^2}{\pi^2} 1(S \equiv R, S' \equiv R)
\]

\[
\leq \frac{C}{N^2} \sum_{S, S' \in S_p, S \neq S'} \frac{q_1 - \frac{q_2^2}{q_1}}{q_1} 1(S \equiv R, S' \equiv R)
\]

\[
= O \left( \left( \frac{1}{q_1} - 1 \right) \frac{n^{2p-1}}{n^{2p}} \right)
\]

\[
= O \left( \left( \frac{1}{q_1} - 1 \right) \frac{1}{n} \right)
\]
Now, for the variance term to vanish we need the conditions \( q_1 = 1 \) or \( q_1 \to 1 \) and \( q_d \to 0 \) and \( \lambda_n q_d \to \infty \) for \( d = 2, \ldots, p \) as \( n \to \infty \). Since, we know that \( O(1) \leq \lambda_n O(n) \), we get \( nq_d \to \infty \) for \( d = 2, \ldots, p \) as \( n \to \infty \). So, we have

\[
\frac{1}{N^2} \frac{1 - \pi}{\pi} \sum_{S \in \mathcal{S}_p} 1(S \cong R) = \left( \frac{1}{\pi} - 1 \right) O \left( \frac{1}{n^p \rho^e} \right)
\]

So,

\[
E \left( \text{Var}_b \left[ \rho_n^{-e} \hat{P}_{B2}(R) \right] \right) = O \left( \left( \frac{1}{q_1} - 1 \right) \frac{1}{n} \right) + O \left( \frac{1}{n \rho_n^{-e-p+1}} \prod_{d=2}^p \frac{1}{\lambda_n q_d} \right)
\]

(iii) We shall use the same decomposition as used in [6] of \( (\rho_n^{-e} \hat{P}_{B2}(R) - \hat{P}(R)) \) into

\[
(\rho_n^{-e} \hat{P}_{B2}(R) - \hat{P}(R)) = \rho_n^{-e} (\hat{P}_{B2} - E_b[\hat{P}_{B2}(R)|G])
\]

\[
+ \rho_n^{-e} (\hat{P}(R) - E(\hat{P}(R)|\xi))
\]

\[
+ E(\hat{P}(R)|\xi) \rho_n^{-e} - \hat{P}(R)
\]

Let us define,

\[
U_3 = E(\hat{P}(R)|\xi) \rho_n^{-e} - \hat{P}(R)
\]

\[
U_2 = \rho_n^{-e} (\hat{P}(R) - E(\hat{P}(R)|\xi))
\]

\[
U_1 = \rho_n^{-e} (\hat{P}_{B2} - E_b[\hat{P}(R)|G])
\]

Now, it is easy to see that

\[
\text{Var}(\rho^{-e} \hat{P}_{B2}(R)) = E(\text{Var}(\rho^{-e} \hat{P}_{B2}(R)|G)) + \text{Var}(E(\rho^{-e} \hat{P}_{B2}(R)|G))
\]

\[
= E(\text{Var}(\rho^{-e} \hat{P}_{B2}(R) - \hat{P}(R)|G)) + \text{Var}(\hat{P}(R))
\]

\[
= E(\text{Var}(U_1|G)) + E(\text{Var}(\hat{P}(R)|\xi)) + \text{Var}(E(\hat{P}(R)|\xi))
\]

\[
= E(\text{Var}(U_1|G)) + E(\text{Var}(U_2|\xi)) + \text{Var}(U_3)
\]

We shall try to see the behavior of \( \text{Var}_b[r_n^{-e} \hat{P}_{b2}(R)|G] \).

\[
E \left( \text{Var}_b \left[ \rho_n^{-e} \hat{P}_{b2}(R) \right] \right) = O \left( \left( \frac{1}{q_1} - 1 \right) \frac{1}{n} \right) + O \left( \frac{1}{n \rho_n^{-e-p+1}} \prod_{d=2}^p \frac{1}{\lambda_n q_d} \right)
\]
Now, since the bootstrap samples for subgraph sampling are selected independently, we have that,

$$\mathbb{E}(\text{Var}_b [\rho_n^{-e} \hat{P}_{B_2}(R)]) = O\left(\left(\frac{1}{q_1} - 1\right) \frac{1}{nB}\right) + O\left(\frac{1}{Bn\rho_n^{e-p+1}} \prod_{d=2}^p \frac{1}{\lambda_n^d q_d}\right)$$

Now, under the condition $\frac{1}{B} \left(\frac{1}{q_1} - 1\right) \rightarrow 0$, $q_d \rightarrow 0$ for all $d = 1, \ldots, p$ and $B \prod_{d=2}^p q_d \geq \frac{1}{n^{p-1} \rho_n}$, we have

$$\mathbb{E}(\text{Var}(U_1|G)) = \mathbb{E}(\text{Var}_b [\rho_n^{-e} \hat{P}_{B_2}(R)|G]) = o(n^{-1})$$

Now, by proof of Theorem 1 in [6], we have,

$$\text{Var}(U_2) = o(n^{-1})$$
$$\text{Var}(U_3) = o(n^{-1})$$

So, we get, $\text{Var}(\rho_n^{-e} \hat{P}_{B_2}(R)) = o(n^{-1})$. Since, we already know $\sqrt{n}$-consistency of $(\rho_n^{-e} \hat{P}(R) - \bar{P}(R))$, this proves the $\sqrt{n}$-consistency of $\rho_n^{-e} \hat{P}_{B_2}(R)$ to $\rho_n^{-e} \bar{P}(R)$.

**B1. Proof of Proposition 6.**

$$\text{Var} [\rho^{-e} \hat{P}(R)] = \text{Var} \left[ \sum_{S \subseteq K_n, S \supseteq R} \frac{1}{\rho^e(n)} |\text{Iso}(R)| \right]$$

$$= \frac{1}{(\rho^e(n)|\text{Iso}(R)|)^2} \mathbb{E} \left[ \sum_{S \subseteq K_n, S \supseteq R} 1(S \subseteq G) \right]^2 - (\hat{P}(R))^2$$

$$= \frac{1}{(\rho^e(n)|\text{Iso}(R)|)^2} \mathbb{E} \left[ \sum_{S,T \subseteq K_n, S \cap T \neq \emptyset} 1(S \subseteq G) \right] - \left(1 - \frac{(n-p)}{n}\right) (\hat{P}(R))^2$$

If $R$ is a connected subgraph, then, we can write,

$$= \frac{1}{(\rho^e(n)|\text{Iso}(R)|)^2} \mathbb{E} \left[ \sum_{S,T \subseteq K_n, S \cap T \neq \emptyset} 1(S \subseteq G) \right]$$

$$= \frac{1}{(\rho^e(n)|\text{Iso}(R)|)^2} \sum_{W: W \subseteq G} 1(W \subseteq G)$$

$$= \frac{1}{(\rho^e(n)|\text{Iso}(R)|)^2} \sum_{W: W \subseteq G} 1(W \subseteq G)$$
\[
\text{Cov}(\rho^{-e_1}\hat{P}(R_1), \rho^{-e_2}\hat{P}(R_2)) = \left(\frac{1}{\rho^{e_1+e_2}(\binom{n}{p_1})(\binom{n}{p_2})|\text{Iso}(R_1)||\text{Iso}(R_2)|}\right) \times \\
\mathbb{E}\left[\left(\sum_{S \subseteq K_n, \ S \ni R_1} 1(S \subseteq G)\right) \left(\sum_{S \subseteq K_n, \ S \ni R_2} 1(S \subseteq G)\right) - \hat{P}(R_1)\hat{P}(R_2)\right]
\]

If \( R \) is a connected subgraph, then, we can write,

\[
\frac{1}{\rho^{e_1+e_2}(\binom{n}{p_1})(\binom{n}{p_2})|\text{Iso}(R_1)||\text{Iso}(R_2)|} \mathbb{E}\left[\sum_{S,T \subseteq K_n, \ S \ni R_1, T \ni R_2, S \cap T \neq \emptyset} 1(S,T \subseteq G)\right] - \left(1 - \frac{\binom{n-p_1}{p_2}}{\binom{n}{p_2}}\right)\hat{P}(R_1)\hat{P}(R_2)
\]

\[B2. \text{ Proof of Lemma 7.} \text{ Let us define}
\]

\[
\hat{\sigma}^2(R) = \frac{1/(1-x)}{\left(\frac{n}{p}\right)^{e_1+e_2}(\binom{n}{p_1})(\binom{n}{p_2})|\text{Iso}(R)|^2} \sum_{W \subseteq K_n, W = S \cup T, S \ni R, |S \cap T| = 1, p} 1(W \subseteq G) - \frac{x\rho^{-2e}\hat{P}(R)^2}{(1-x)}
\]

where, \( x = \left(1 - \frac{(n-p)!}{n!(n-2p)!}\right) \) and

\[
\hat{\sigma}(R_1, R_2) = \frac{1/(1-y)}{\left(\frac{n}{p}\right)^{e_1+e_2}(\binom{n}{p_1})(\binom{n}{p_2})|\text{Iso}(R_1)||\text{Iso}(R_2)|} \sum_{W \subseteq K_n, W = S \cup T, S \ni R_1, T \ni R_2, |S \cap T| = 1, p} 1(W \subseteq G) - \frac{y\rho^{-e_1+e_2}\hat{P}(R_1)\hat{P}(R_2)}{(1-y)}
\]

where, \( y = \left(1 - \frac{(n-p_1)!/(n-p_2)!}{n!(n-p_1-p_2)!}\right) \).
Now,

\[
\mathbb{E} \left[ \hat{\sigma}^2(R) \right] = \frac{1}{(1-x)} \frac{\mathbb{E} \left[ \sum_{W \subseteq K_n, W = S \cup T, S, T \geq R, |S \cap T| = 1, p} 1(W \subseteq G) \right]}{\mathbb{P}(\sigma^2(R))} - x \rho_n^{-2e} \mathbb{E} \left[ \hat{P}(R) \right]^2 - x \rho_n^{-2e} \mathbb{E} \left[ \hat{P}(R) \right] \\
= \frac{1}{(1-x)} \frac{1}{\mathbb{P}(\sigma^2(R))} \mathbb{E} \left[ \sum_{W \subseteq K_n, W = S \cup T, S, T \geq R, |S \cap T| = 1, p} 1(W \subseteq G) \right] - x \rho_n^{-2e} \mathbb{E} \left[ \hat{P}(R) \right] \\
= \frac{1}{1-x} \frac{1}{\mathbb{P}(\sigma^2(R))} \mathbb{E} \left[ \sum_{W \subseteq K_n, W = S \cup T, S, T \geq R, |S \cap T| = 1, p} 1(W \subseteq G) \right] - x \rho_n^{-2e} \mathbb{E} \left[ \hat{P}(R) \right] \\
= \frac{1}{1-x} \frac{1}{\mathbb{P}(\sigma^2(R))} \mathbb{E} \left[ \sum_{W \subseteq K_n, W = S \cup T, S, T \geq R, |S \cap T| = 1, p} 1(W \subseteq G) \right] - x \rho_n^{-2e} \mathbb{E} \left[ \hat{P}(R) \right] \\
= \frac{\mathbb{E} \left[ \sigma^2(R) \right]}{1-x} - x \rho_n^{-2e} \mathbb{E} \left[ \hat{P}(R) \right] - o \left( \mathbb{E} \left[ \sigma^2(R) \right] \right)
\]

Similarly, we get,

\[
\mathbb{E} \left[ \hat{\sigma}(R_1, R_2) \right] = \mathbb{Cov} \left( \rho_n^{-e_1} \hat{P}(R_1), \rho_n^{-e_2} \hat{P}(R_2) \right) - o \left( \mathbb{Cov} \left( \rho_n^{-e_1} \hat{P}(R_1), \rho_n^{-e_2} \hat{P}(R_2) \right) \right)
\]

Now, from the Theorem 1(a) in [6], we know that as \( n \to \infty \), if \( \hat{\rho}_n = \frac{D}{n-1} \) as defined in (2.6),

\[
\frac{\hat{\rho}_n}{\rho_n} \to 1
\]

So, using the estimate \( \hat{\rho}_n \), we get that,

\[
\frac{\hat{\sigma}^2(R)}{\sigma^2(R)} \to 1, \quad \frac{\hat{\sigma}(R_1, R_2)}{\sigma(R_1, R_2)} \to 1
\]
**B3. Proof of Lemma 8.** Given $G$,

\[ \mathbb{E} \left[ \hat{\sigma}^2_{Bi}(R) \big| G \right] = \sum_{W = S \cup T, S \neq T, |S \cap T| = 1, p} \left( \frac{\hat{\rho}_{n}^{\text{is}}(\frac{n}{p}) | \text{Iso}(R) \right)}{(1 - x) \left( \frac{\hat{\rho}_{n}^{\text{is}}(\frac{n}{p}) | \text{Iso}(R) \right)} \right)^2 \mathbb{E} \left[ \hat{P}_{Bi}(W) \big| G \right] - \frac{x \mathbb{E} \left[ \hat{\rho}^{2e}_{n} \hat{P}_{Bi}(R)^2 \big| G \right]}{(1 - x)} \]

\[ = \hat{\sigma}^2(R) - \frac{x \text{Var} \left( \hat{T}_{Bi}(R) \right)}{1 - x} = \hat{\sigma}^2(R) - o \left( \hat{\sigma}^2(R) \right) \]

where the last inequality follows since $x = O \left( \frac{1}{n} \right)$ and Theorem 3.1 and Theorem 3.2 for $i = 1, 2$.

Similarly, we get,

\[ \mathbb{E} \left[ \hat{\sigma}_{Bi}(R_1, R_2) \big| G \right] = \hat{\sigma}(R_1, R_2) - o \left( \hat{\sigma}^2(R) \right) \]

So using Lemma 7, we have that,

\[ \frac{\hat{\sigma}^2_{Bi}(R)}{\sigma^2(R)} \xrightarrow{P} 1, \quad \frac{\hat{\sigma}_{Bi}(R_1, R_2)}{\sigma(R_1, R_2)} \xrightarrow{P} 1 \text{ for } i = 1, 2 \]

**References.**

[1] Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* 11 581–598. MR637937 (82m:60022)

[2] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* 286 509–512.

[3] Bearman, P. S., Moody, J. and Stovel, K. (2004). Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. *American Journal of Sociology* 110 44–91.

[4] Bhattacharyya, S. and Bickel, P. J. (2013). Supplement to “Subsampling Bootstrap of Count Features of Networks”.

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

[6] Bickel, P. J., Chen, A. and Levina, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* 39 2280–2301. MR2906868

[7] Bickel, P. J., Götze, F. and van Zwet, W. R. (1997). Resampling fewer than $n$ observations: gains, losses, and remedies for losses. *Statist. Sinica* 7 1–31. Empirical Bayes, sequential analysis and related topics in statistics and probability (New Brunswick, NJ, 1995). MR1441142 (98g:62079)

[8] Bollobás, B., Janson, S. and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* 31 3–122. MR2337396 (2008e:05124)
[9] Chung, F. and Lu, L. (2002). Connected components in random graphs with given expected degree sequences. *Ann. Comb.* 6 125–145. MR1955514 (2003k:05123)

[10] Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)* 28 33–61. MR2463439 (2010a:60127)

[11] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* 7 1–26. MR515681 (80b:62021)

[12] Frank, O. (2005). Network sampling and model fitting. *Models and methods in social network analysis* 31–56.

[13] Handcock, M. S. and Gile, K. J. (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics* 4 5–25.

[14] Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* 97 1090–1098.

[15] Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks* 5 109–137.

[16] Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Institute for Advanced Study, Princeton, NJ.*

[17] Kallenberg, O. (2005). *Probabilistic symmetries and invariance principles. Probability and its Applications (New York).* Springer, New York. MR2161313 (2006i:60002)

[18] Kolaczyk, E. D. (2009). *Statistical analysis of network data. Springer Series in Statistics.* Springer, New York. Methods and models. MR2724362 (2011j:05308)

[19] Leskovec, J., Kleinberg, J. and Faloutsos, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* 177–187. ACM.

[20] Lovász, L. (2012). *Large networks and graph limits* 60. American Mathematical Soc.

[21] Middendorf, M., Ziv, E. and Wiggins, C. H. (2005). Inferring network mechanisms: the Drosophila melanogaster protein interaction network. *Proceedings of the National Academy of Sciences of the United States of America* 102 3192–3197.

[22] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298 824–827.

[23] Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheiffer, M. and Alon, U. (2004). Superfamilies of evolved and designed networks. *Science* 303 1538–1542.

[24] Nowicki, K. and Sniders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* 96 1077–1087. MR1947255

[25] Picard, F., Daudin, J.-J., Koskas, M., Schbath, S. and Robin, S. (2008). Assessing the exceptionality of network motifs. *Journal of Computational Biology* 15 1–20.

[26] Przytycka, T. M. (2006). An important connection between network motifs and parsimony models. In *Research in Computational Molecular Biology* 321–335. Springer.

[27] Shen-Orr, S. S., Milo, R., Mangan, S. and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics* 31 64–68.

[28] Thompson, S. K. (2012). *Sampling, third ed. Wiley Series in Probability and Statistics.* John Wiley & Sons Inc., Hoboken, NJ. MR2894042 (2012k:60204)

[29] Thompson, S. K. and Frank, O. (2000). Model-based estimation with link-tracing sampling designs. *Survey Methodology* 26 87–98.
[30] Traud, A. L., Kelsic, E. D., Mucha, P. J. and Porter, M. A. (2011). Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.* **53** 526–543. MR2834086 (2012g:91133)

[31] Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications* 8. Cambridge university press.

[32] Wernicke, S. (2006). Efficient detection of network motifs. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on* **3** 347–359.