Central limit theorems for network driven sampling

Xiao Li

Department of Statistics
367 Evans Hall
University of California, Berkeley, CA 94720-3860
e-mail: sxli@berkeley.edu

and

Karl Rohe

Department of Statistics
1220 Medical Sciences Center
1300 University Ave,
Madison, WI 53706
e-mail: karlrohe@stat.wisc.edu

Abstract: Respondent-Driven Sampling is a popular technique for sampling hidden populations. This paper models Respondent-Driven Sampling as a Markov process indexed by a tree. Our main results show that the Volz-Heckathorn estimator is asymptotically normal below a critical threshold. The key technical difficulties stem from (i) the dependence between samples and (ii) the tree structure which characterizes the dependence. The theorems allow the growth rate of the tree to exceed one and suggest that this growth rate should not be too large.

MSC 2010 subject classifications: Primary 60G17; secondary 62D05.
Keywords and phrases: Central limit theorem, respondent-driven sampling, stochastic blockmodel.

1. Introduction

Classical sampling requires a sampling frame, a list of individuals in the target population with a method to contact each individual (e.g. a phone number). For many populations, constructing a sampling frame is infeasible. Network driven sampling enables researchers to access populations of people, webpages, and proteins that are otherwise difficult to reach. These techniques go by many names: web crawling, Respondent-Driven Sampling, breadth-first search, snowball sampling, co-immunoprecipitation, and chromatin immunoprecipitation. In each application, the only way to reach the population of interest is by asking participants to refer friends.

Respondent-Driven Sampling (RDS) serves as a motivating example for this paper. The Centers for Disease Control, the World Health Organization, and

the Joint United Nations Programme on HIV/AIDS have invested in RDS to reach marginalized and hard-to-reach populations [6, 1]. Each individual i in the population has a corresponding feature y_i (e.g. $y_i \in \{0,1\}$ and $y_i = 1$ if i is HIV+). Using only the sampled individuals, we wish to make inferences about the average value of y_i across the entire population, denoted as μ (e.g. the proportion of the population that is HIV+). Extensive previous statistical research has proposed various estimators of μ which are approximately unbiased based upon various types of models for an RDS sample [16, 17, 4]. We note that in the papers cited above (except [4]), RDS is assumed to sample with replacement. Previous research has also explored the variance of these estimators [5, 13]. This paper studies the asymptotic distribution of statistics related to these estimators.

Results on asymptotic distributions for RDS are useful for two obvious reasons. First, they allow us to construct asymptotic confidence intervals for μ. Second, they provide essential tools to test various statistical hypotheses. The only central limit theorem considered in the RDS literature studied the case when the tree indexed process reduces to a Markov chain [5]; this presumes that each individual refers exactly one person. Previous research suggests that the number of referrals from each individual is fundamental in determining the variance of common estimators [13]. This paper establishes two central limit theorems in settings which allow for multiple referrals.

The main results apply to both the sample average and the Volz-Heckathorn estimator, which is an approximation of the inverse probability weighted estimator (cf Remark 1). Because the inverse probability weighted (IPW) estimator and its extensions are asymptotically unbiased, these estimators are often preferred to the sample average.

2. Notation

Following [5] and [13], the results below model the network sampling mechanism as a tree indexed Markov process on a graph. There are many assumptions in this model which are incorrect in practice. However, like the i.i.d assumption, it allows for tractable calculations. In the simulations, we show that the theory derived from this model provides a good approximation for a more realistic sampling model. [12] studies the sensitivities of the estimators to this model.

Let $G = (V, E)$ be a finite, undirected, and simple graph with vertex set $V = \{1, ..., N\}$ and edge set E. V contains the individuals in the population and E describes how they are related to one another. As discussed in the introduction, $y : V \to \mathbb{R}$ is a fixed real-valued function on the state space V; these are the node features that are measured on the sampled nodes. The target of RDS is to estimate $\mu = N^{-1} \sum_{i=1}^{N} y(i)$.

If each sampled node referred exactly one friend, then the Markov sampling procedure would be a Markov chain. Several classical central limit theorems exist for this model; see [8] for a review. The results herein allow for each sampled node to refer more than one node. This is a Markov process indexed not by

a chain, but rather by a tree. Denote the referral tree as $\mathbb{T}$. Where the node set of G indexes the population, the node set of $\mathbb{T}$ indexes the samples. That is, we observe a subset of the individuals in G with the sample $\{X_\tau\}_{\tau \in \mathbb{T}} \subset V$. An edge (σ, τ) in the referral tree denotes that sampled individual X_σ referred individual X_τ into the sample. Mathematically, $\mathbb{T}$ is a rooted tree–a connected graph with n nodes, no cycles, and a vertex 0 which indexes the seed node. To simplify notation, $\sigma \in \mathbb{T}$ is used synonymously with σ belonging to the vertex set of $\mathbb{T}$.

For each non-root node $\tau \in \mathbb{T}$, denote $p(\tau) \in \mathbb{T}$ as the parent of τ (i.e. the node one step closer to the root). This paper presumes that $\{X_\tau\}_{\tau \in \mathbb{T}}$ is a tree-indexed random walk on G, which was a model introduced by [2]. This model generalizes a Markov chain on G; each transition $X_{p(\tau)} \to X_\tau$ is an independent and identically distributed Markov transition with some transition matrix P that is defined below. Following [2], we will call this process a $(\mathbb{T}, P)$-walk on G. Unless stated otherwise, it will be presumed throughout that the root node of the random walk X_0 is initialized from the equilibrium distribution π of P. It follows that X_σ has distribution π for all $\sigma \in \mathbb{T}$.

Unless stated otherwise, this paper presumes throughout that the transition matrix P is constructed from a weighted graph G. Let w_{ij} be the weight of the edge $(i, j) \in E$; if $(i, j) \notin E$, define $w_{ij} = 0$. If the graph is unweighted, then let $w_{ij} = 1$ for all $(i, j) \in E$. Define the degree of node i as $deg(i) = \sum_j w_{ij}$. If the graph is unweighted, then $deg(i)$ is the number of connections to node i. Throughout this paper, the graph is undirected. So, $w_{ij} = w_{ji}$ for all pairs i, j. Given that $\{X_{p(\tau)} = i\}$, the probability of $\{X_\tau = j\}$ is proportional to w_{ij};

$$P_{ij} = P\left(X_\tau = j | X_{p(\tau)} = i\right) = \frac{w_{ij}}{deg(i)}. \tag{2.1}$$

We use the term **simple random walk** for the Markov chain constructed on the unweighted graph (i.e. $w_{i,j} \in \{0, 1\}$ for all i, j). The simple random walk presumes that each participant selects a friend uniformly and independently at random from their list of friends.

[10] serves as this paper's key reference for Markov processes. Following the notation in that text, define $E_\pi(y) = \sum_{i=1}^N \pi_i y(i)$ and $var_\pi(y) = E_\pi(y - E_\pi(y))^2$ for the function y. In order to estimate μ, we observe $y(X_\tau)$ for all $\tau \in \mathbb{T}$. Because G is undirected, P is reversible and has stationary distribution π with $\pi_i \propto deg(i)$ for all $i \in G$; this fact is helpful for creating an asymptotically unbiased estimator for μ, particularly under the simple random walk assumption [17].

Remark 1. *In general, the quantity of interest $\mu = N^{-1} \sum_{i=1}^N y(i)$ is not equal to $E_\pi(y)$. As such, the sample average of $y(X_\tau)$'s is a biased estimator for μ. With inverse probability weighting, define a new function $y'(i) = y(i)(N\pi_i)^{-1}$ and the respective estimator*

$$\hat{\mu}_{IPW} = \frac{1}{n} \sum_{\sigma \in \mathbb{T}} y'(X_\sigma) = \frac{1}{n} \sum_{\sigma \in \mathbb{T}} \frac{y(X_\sigma)}{N\pi_{X_\sigma}},$$

where $n = |\mathbb{T}|$ is the sample size. Then, $E_\pi(\hat{\mu}_{IPW}) = E_\pi(y') = \mu$. As such, the sample average of the $y'(X_\tau)$'s is an unbiased estimator of μ. Unfortunately, the values π_i are unknown. In practice, RDS participants are asked various questions to measure how many friends they have in G. Under the simple random walk assumption, π_i is proportional to the number of friends of i. Therefore the Volz-Heckathorn estimator

$$\hat{\mu}_{VH} = \sum_{\sigma \in \mathbb{T}} \frac{y(X_\sigma)/deg(X_\sigma)}{\sum_{\tau \in \mathbb{T}} 1/deg(X_\tau)}$$

is in essence a Hájek estimator based upon $deg(i)$ [17]. Under the simple random walk assumption, this estimator provides an asymptotically unbiased estimator of μ.

For each node $\tau \in \mathbb{T}$, let $|\tau|$ be the distance of the node from the root; this is also called the "wave" of τ. For every pair of node $\sigma, \tau \in \mathbb{T}$, define $d(\sigma, \tau)$ to be the distance between σ and τ on $\mathbb{T}$ (as a graph). For each non-leaf node $\sigma \in \mathbb{T}$, let $\eta(\sigma)$ be the number of offspring of σ. A tree is said to be an m-tree of height h if $\eta(\sigma) = m$ for all $\sigma \in \mathbb{T}$ with $|\sigma| < h$ and $\eta(\sigma) = 0$ for all $|\sigma| = h$. Here, both m and h are a natural numbers (i.e. $m, h \in \mathbb{N}$). $\mathbb{T}$ is said to be Galton-Watson if $\eta(\sigma)$ are i.i.d random variables in $\mathbb{N}$. While the theorems below only study 2-trees; the computational experiments in Section 5 suggest that the conclusions of the analytical results are highly robust to replacing the 2-tree with a Galton-Watson tree.

There are two primary concerns about the model and estimator used in the main results below. First, the Markov model allows for resampling. Second, the results below only apply to m-trees, not more general trees. The simulations in Section 5 suggest that the analytic results continue to hold under a more realistic setting that addresses both of these concerns.

3. Main results

Let $\mathbb{T}$ be an m-tree and λ_2 be the second largest eigenvalue of P. The variance of $\hat{\mu}_{IPW}$ decays at the standard rate if and only if $m < \lambda_2^{-2}$ [13]. In other words, if $m > \lambda_2^{-2}$, then

$$var\left(\frac{1}{\sqrt{|\sigma \in \mathbb{T} : |\sigma| \le h|}} \sum_{\sigma \in \mathbb{T} : |\sigma| \le h} y(X_\sigma) \right) \to \infty$$

as $h \to \infty$. As such, using the traditional scaling, no central limit theorem holds above the critical threshold. Because of this, the theorems focus on the case $m < \lambda_2^{-2}$. When $m > \lambda_2^{-2}$, the simulations in Section 5 suggest that the central limit theorem does not hold for any scaling.

Theorem 1 is a central limit theorem for an estimator constructed from the tree-indexed Markov chain. The theorem holds for any function y, any reversible transition matrix with second largest eigenvalue satisfying $|\lambda_2| \neq 1$, and any $m < \lambda_2^{-2}$.

Theorem 1. *Suppose that P is a reversible transition matrix with respect to the equilibrium distribution π, and that the eigenvalues of P are $1 = \lambda_1 > |\lambda_2| \geq ... \geq |\lambda_N|$. Without loss of generality, suppose that $E_\pi(y) = 0$. Define*

$$Y_i = \frac{1}{\sqrt{m^i}} \sum_{\tau \in \mathbb{T}:|\tau|=i} y(X_\tau).$$

If $\mathbb{T}$ is an m-tree with $m < \lambda_2^{-2}$, then

$$\frac{1}{\sqrt{h}} \sum_{i=1}^{h} Y_i \to N(0, \sigma_0^2)$$

in distribution, where $\sigma_0^2 = var_\pi((\sqrt{m}P - I)^{-1}y) - var_\pi(P(\sqrt{m}P - I)^{-1}y)$.

The sequence of random variables considered in Theorem 1 are not exactly sample averages, but a reweighted form of sample average. Samples in the same wave are equally weighted, while samples from different waves are not. The following theorem provides a theoretical guarantee on the distribution of sample average for a specific class of transition matrix and node feature. For a vector x, one of the conditions uses the notation $\|x\|_\infty = \max_i x_i$.

Theorem 2. *Let $\mathbb{T}$ be a $2-tree$. Without loss of generality, suppose that $E_\pi(y) = 0$. Define $\hat{\mu}_h = \frac{1}{\sqrt{2^h}} \sum_{\sigma \in \mathbb{T}, |\sigma| \leq h} y(X_\sigma)$. Suppose that*

(c1) $E(\hat{\mu}_h^{2k+1}) = 0$ for all $h, k \in \mathbb{N}$;
(c2) for any function f on V satisfying $E_\pi f = 0$, $\|Pf\|_\infty \leq |\lambda_2|\|f\|_\infty$;
(c3) $|\lambda_2| < \frac{1}{\sqrt{2}}$;
then

$$\hat{\mu}_h \to N(0, \sigma_0^2)$$

in distribution for some σ_0^2.

Remark 2. *Condition (c1) is a technical condition on the symmetry of $\hat{\mu}_h$ that is necessary in the proof. The following proposition provides a sufficient condition for (c1).*

Proposition 1. *Suppose that y is symmetric, i.e. for any $i \in V$ there exists j such that $y(j) = -y(i)$. If $p(u, v) = P(y(X_\sigma) = v|y(X_{p(\sigma)}) = u)$ is well-defined and $p(u, v) = p(-u, -v)$ for all $u, v \in y(V)$, then condition (c1) is satisfied.*

Proof. Under the conditions of the proposition, the distribution of $\hat{\mu}_h$ is symmetric with respect to 0. Thus $E(\hat{\mu}_h^{2k+1}) = 0$ for all $h, k \in \mathbb{N}$. □

Conditions (c2)–(c3) can be substituted by the following condition (c2'):
(c2') There exists $c < \frac{1}{\sqrt{2}}$ such that for any function f on V satisfying $E_\pi f = 0$, $\|Pf\|_\infty \leq c\|f\|_\infty$.
Condition (c2') is weaker than (c2) and (c3) combined, but is stronger than (c3) alone. To see this, let f be the eigenfunction of the second eigenvalue, and it follows that $|\lambda_2| < \frac{1}{\sqrt{2}}$. It can be easily seen that one necessary condition for (c2') is that

$$\sum_j |P_{ij} - \pi_j| < \frac{1}{\sqrt{2}}$$

for all $i \in V$. In other words, all the rows of P must be close to π. As previously discussed, condition (c3) is actually a necessary condition for the central limit theorem [13], in the sense that the variance of $\hat{\mu}_h$ tends to infinity if $|\lambda_2| \geq \frac{1}{\sqrt{2}}$.

For clarity in the exposition of the theorem and the proof, we have only proved the theorem for the 2-tree. Results for more general $m-$tree can be proved with a similar technique.

3.1. Extension to the Volz-Heckathorn estimator

When P is restricted to be the transition matrix of the simple random walk on G, the following corollary shows that Theorem 2 can be extended to the Volz-Heckathorn estimator [17].

Denote $\bar{d} = \frac{1}{N} \sum_{i \in V} deg(i)$ as the average node degree. Following Remark 1, the IPW estimator contains $1/(N\pi_i)$ which is equal to $\bar{d}/deg(i)$. The Volz-Heckathorn estimator first estimates $\bar{d}$ with the harmonic mean of the observed degrees. Because this harmonic mean converges to $\bar{d}$ in probability, the following corollary applies Slutsky's Theorem to give a central limit theorem for the Volz-Heckathorn estimator.

Corollary 1. *Let $\mathbb{T}$ be a 2-tree. Suppose in particular that P is the transition matrix of the simple random walk on G. Define a new node feature $y'(i) = y(i)/deg(i)$. Without loss of generality, suppose that $E_\pi y' = 0$ (this is not equivalent to $E_\pi y = 0$). Define*

$$\hat{\mu}_{h,VH} = \hat{\mu}_h \hat{\bar{d}} = \frac{1}{\sqrt{2^h}} \sum_{\sigma \in \mathbb{T}, |\sigma| \leq h} y'(X_\sigma)\hat{\bar{d}},$$

where

$$\hat{\bar{d}} = \frac{2^{h+1} - 1}{\sum_{\sigma \in \mathbb{T}, |\sigma| \leq h} 1/deg(X_\tau)}.$$

If the new node feature y' and the transition matrix P satisfy conditions (c1)–(c3) in Theorem 2, then

$$\hat{\mu}_{h,VH} \to N(0, \sigma_{0,VH}^2)$$

in distribution for some $\sigma_{0,VH}^2$.

3.2. Illustrating the conditions with a blockmodel

Consider G as coming from a blockmodel with two blocks [11]. Previously, [5] studied RDS with this model. It serves as an approximation to the Stochastic Blockmodel. In particular, suppose that each node $i = 1, \ldots, N$ is assigned to a block with $z(i) \in \{1, 2\}$. Suppose that each block contains $N/2$ nodes. For

$$B = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}, \quad \text{with } p \in (0,1),$$

suppose that every pair i,j has $w_{i,j} = B_{z(i),z(j)} \in (0,1)$. Thus, under the construction of P in Equation (2.1),

$$P_{ij} = (N/2)^{-1} B_{z(i),z(j)}.$$

Further suppose that if $z(i) = z(j)$, then $y_i = y_j$. Condition (c1) is satisfied if $y_i = -y_j$ is satisfied for i,j with $z(i) \neq z(j)$.

Given the structural equivalence of nodes within the same block, it is sufficient to study the conditions (c2) and (c3) with a Markov chain where the state space is reduced to the block labels $\{1,2\}$ and the transition matrix is $\mathcal{P} = B \in \mathbb{R}^{2 \times 2}$. See Section C in the Appendix for a further discussion of this fact.

Notice that $\lambda_2 = (p-r)/(p+r)$ is the second eigenvalue of both P and $\mathcal{P}$. So, if

$$\frac{1}{2} < p < \frac{1}{2} + \frac{1}{2\sqrt{2}},$$

then conditions (c2) and (c3) are satisfied.

This example can be expanded to study a blockmodel with $2K$ blocks, where each block contains $N/(2K)$ nodes. For $u,v \in 1,\ldots,2K$, define the blockmodel with $B \in \mathbb{R}^{2K}$,

$$B_{a,b} = pI(a=b) + rI(a \neq b) \quad \text{for } a,b \in \{1,\ldots,2K\}.$$

Suppose that the outcome y_i depends only on the block label, i.e. $y_i = \mathscr{Y}_{z(i)}$ for some vector $\mathscr{Y} \in \mathbb{R}^{2K}$. For $u \in 1,\ldots,K$, suppose that $\mathscr{Y}_{2u} = -\mathscr{Y}_{2u-1}$ Then, $\mathscr{P} \in \mathbb{R}^{2K}$ can be defined with $\tilde{p} = p/(p + r(K-1))$ as

$$\mathscr{P}_{a,b} = \tilde{p}I(a=b) + \frac{1-\tilde{p}}{2K-1}I(a \neq b), \quad \text{for } a,b \in \{1,\ldots,2K\},$$

Then, conditions (c1), (c2), and (c3) hold if

$$\frac{1}{2K} < \tilde{p} < \frac{1}{2K} + \frac{1}{2\sqrt{2}}.$$

4. Estimating the variance

For some node feature $\tilde{y}$ (e.g. HIV status y or the y' in Remark 1 that leads to the IPW estimator), let $\tilde{\mu}$ denote the sample average. Denote $\sigma_{\tilde{\mu}}^2$ as $Var_{\mathbb{T},P}(\tilde{\mu})$, where the subscript $\mathbb{T}, P$ denotes that the data is collected via a $(\mathbb{T}, P)$-walk on G. This subsection studies a simple plug-in estimator for $\sigma_{\tilde{\mu}}^2$.

The following function is essential to expressing $\sigma_{\tilde{\mu}}^2$ [13].

Definition 1. *Select two nodes I, J uniformly at random from the tree $\mathbb{T}$. Define the random variable $D = d(I,J)$ to be the graph distance in $\mathbb{T}$ between I and J. Define $\mathbb{G}$ as the probability generating function for D,*

$$\mathbb{G}(z) = E(z^D).$$

In practice, $\mathbb{T}$ is observed. So, the function $\mathbb{G}$ can be computed. In many studies there are multiple seed nodes. In these cases, we suggest computing $d(I, J)$ on a tree which has an artificial root node that connects to all of the seeds; this root node could be imagined as an individual that is responsible for finding the seed nodes. In this tree, two different seed nodes would be distance 2 apart.

Denote the autocorrelation at lag 1 of $\tilde{y}(X_\tau)$ as

$$R = \frac{Cov(\tilde{y}(X_{p(\tau)}), \tilde{y}(X_\tau))}{var_\pi(\tilde{y})}.$$

Both $Cov(\tilde{y}(X_{p(\tau)}), \tilde{y}(X_\tau))$ and var_π can be estimated with plug-in quantities. Because the data has been sampled proportional to π, the plug-in quantity for var_π should not explicitly adjust for π. Namely, we have

$$\widehat{var_\pi(\tilde{y})} = \frac{1}{n} \sum_{\tau \in \mathbb{T}} (\tilde{y}(X_\tau) - \hat{\mu}_{VH})^2,$$

and the estimation for $Cov(\tilde{y}(X_{p(\tau)}), \tilde{y}(X_\tau))$ is

$$Cov(\tilde{y}(\widehat{X_{p(\tau)}}), \tilde{y}(X_\tau)) = \frac{1}{n-1} \sum_{\tau \in \mathbb{T}\backslash 0} (\tilde{y}(X_{p(\tau)}) - \hat{\mu}_{VH})(\tilde{y}(X_\tau) - \hat{\mu}_{VH}),$$

where $\{\mathbb{T} \setminus 0\}$ contains all nodes except the root node 0 (because $p(0)$ does not exist). Using these plug-in quantities, define $\hat{R}$. Then, the estimator is

$$\hat{\sigma}_{\hat{\mu}}^2 = \mathbb{G}(\hat{R})\widehat{var_\pi(\tilde{y})}.$$

A popular bootstrap technique for estimating $\hat{\sigma}_{\hat{\mu}}^2$ resamples $y(X_\tau)$ as a Markov process (i.e. in addition to X_τ being a Markov process, the bootstrap procedure also assumes that $y(X_\tau)$ is Markov) [15]. This model is akin to the blockmodel with two blocks in Section 3.2. The following assumption is weaker than this assumption:

Assumption 1. $\tilde{y}(i) = \mu + \sigma f(i)$, where $\mu, \sigma \in \mathbb{R}$ and $f : V \to \mathbb{R}$ is an eigenfunction of P with $var_\pi(f) = 1$.

Proposition 2. *Under Assumption 1,*

$$\sigma_{\hat{\mu}}^2 = \mathbb{G}(R)var_\pi(\tilde{y}).$$

While Assumption 1 is weaker than the previous assumption in [15], the next proposition highlights the danger of this assumption. It uses a different assumption which is a rather weak assumption.

Assumption 2. $\mathbb{G}$ is convex on $[\lambda_{\min}, 1]$, where $\lambda_{\min}$ is the smallest eigenvalue of P.

Because $\mathbb{G}$ is a probability generating function, it is always convex on $[0, 1]$. As such, we only need to be worried about negative values. Recall that the

central limit theorems above only hold when $|\lambda_{\min}| < 1/\sqrt{2} \approx .7$ (the smallest possible value for $\lambda_{\min}$ is -1). Some simulated trees given in the appendix suggest that if $\mathbb{G}$ is not convex, it often fails in the neighborhood of -1. As such, the assumption that $|\lambda_{\min}| < 1/\sqrt{2} \approx .7$ is likely to imply Assumption 2. In practice, one observes the referral tree $\mathbb{T}$. Thus, one can compute the second derivative of $\mathbb{G}$. Eigenvalues of P close to negative one arise in antithetic sampling, where adjacent samples are dissimilar. For example, if the population in G was heterosexuals and edges in G represent sexual contacts, then men would only refer women and vice versa. In this case, $\lambda_{\min}$ would be exactly -1. While easily imagined, such settings are not current practice for RDS. As such, large an negative values are uncommon; $\lambda_{\min}$ is likely close to zero.

The following proposition follows from an application of Jensen's inequality. A proof is given in Appendix D.

Proposition 3. *Under Assumption 2,*

$$\sigma_{\tilde{\mu}}^2 \geq \mathbb{G}(R) var_\pi(\tilde{y}).$$

Because Assumption 2 is not very restrictive, the inequality in Proposition 3 highlights the danger in breaking Assumption 1 (and thus the Markov model in [15]); *breaking Assumption 1 leads to* $\hat{\sigma}_{\tilde{\mu}}^2$ *underestimating the variance.*

5. Numerical results

In this section, we illustrate the theoretical results on simulated data. The simulations are performed on networks simulated from the Stochastic Blockmodel [7]. The four colors in Figure 1 correspond to four different networks, from four different parameterizations of the model. Each of the four networks has $N = 5,000$ nodes, equally balanced between group zero and group one. The probability of a connection between two nodes in different blocks is r and the probability of connection between two nodes in the same block is p. To control the eigenvalues of the 5000×5000 transition matrix, consider the transition matrix between classes given by $\mathscr{P} = E(D)^{-1}E(A)$. The second eigenvalue of $\mathscr{P}$ is [14]

$$\lambda_2(\mathscr{P}) = \frac{p-r}{p+r},$$

where expectations are under the Stochastic Blockmodel. In our simulation, the second eigenvalue of the actual transition matrix is typically very close to $\lambda_2(\mathscr{P})$. We take $p + r = 0.01$ in all four Stochastic Blockmodels so that the average degree is about 25. As such, $\lambda_2(\mathscr{P})$ is actually controlled by $p - r$.

For each of the four networks, we carry out four different sampling designs. Let $\mathbb{T}$ be either a $2-$tree or a Galton-Watson tree with $E(\eta(\sigma)) = 2$. For the Galton-Watson tree, the distribution of $\eta(\sigma)$ is uniform on $\{1, 2, 3\}$. For each $\mathbb{T}$, we consider both with replacement sampling (i.e. the $(\mathbb{T}, P)$-walk on G) and without replacement sampling (i.e. referrals are sampled uniformly from the friends that have not yet been sampled). Note that the conditions of Theorem 2 may be violated when either the Galton-Watson tree or without-replacement

sampling is used. We take the first 8 waves of $\mathbb{T}$ as our sample. As such, the sample size is roughly $N/10$. For each social network and sampling design, we repeat the sampling process 2000 times and compute $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y(X_i)$ for each sample. The Quantile-Quantile (Q-Q) plot of $\hat{\mu}$ is shown in the left panel of Figure 1; note that the Q-Q plot centers and scales each distribution to have mean zero and standard deviation one. In addition, we repeat the above simulation for the Volz-Heckathorn estimator, and the Q-Q plot of $\hat{\mu}_{VH}$ is shown in the right panel of Figure 1.
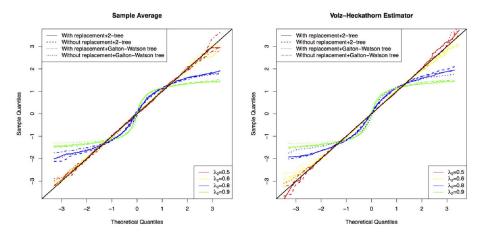

FIG 1. *Q-Q plots of the sample average (left panel) and the Volz-Heckathorn estimator (right panel) for different social network and sampling designs. For each scenario, we draw 2000 network driven samples of size $\approx$ 500 from a network containing 5,000 nodes. Here the threshold for λ_2 is $1/\sqrt{2} \approx 0.707$. For the two settings with $\lambda_2 < 1/\sqrt{2}$, the distributions appear normal. However, for the two settings with $\lambda_2 > 1/\sqrt{2}$, the distributions do not appear normal. Across all values of λ_2, there is no apparent difference between the four different designs (i.e. replacement sampling vs without replacement sampling and 2-tree vs Galton-Watson tree).*

It is clear from Figure 1 that there are two patterns of distribution: when $\lambda_2 < 1/\sqrt{m} \approx 0.7$, i.e. λ_2=0.5 or 0.6, the Q-Q plots of $\hat{\mu}$ and $\hat{\mu}_{VH}$ approximately lie on the line $y = x$ for all sampling design; when $\lambda_2 > 1/\sqrt{m} \approx 0.7$, i.e. λ_2=0.8 or 0.9, the Q-Q plot of $\hat{\mu}$ and $\hat{\mu}_{VH}$ departs from the line $y = x$. Taken together, Figure 1 suggests that the distribution of $\hat{\mu}$ and $\hat{\mu}_{VH}$ converges to Gaussian distribution if and only if $m < \lambda_2^{-2}$. In fact, when $m > \lambda_2^{-2}$, the distribution of the estimators has two modes. The relationship between the expectation of the offspring distribution and the second eigenvalue of the social network determines the asymptotic distribution of RDS estimators, regardless of the node feature, the particular structure of the tree or the way we handle replacement.

6. Discussion

A recent review of the RDS literature counted over 460 studies which used RDS [18]. Many of these studies seek to estimate the prevalence of HIV or

other infectious diseases; for these studies, a point estimate of the prevalence is insufficient. These studies have used confidence intervals constructed from bootstrap procedures and from estimates of the standard error. These standard error intervals implicitly rely on a central limit theorem and this paper provides a partial justification for such techniques, so long as $m \leq 1/\lambda_2^2$. Figure 1 suggests that if m is larger than $1/\lambda_2^2$, then the simple estimators ($\hat{\mu}$ and $\hat{\mu}_{VH}$) are no longer normally distributed.

The theorems in this paper do not apply to general trees, only to m-trees. If $\mathbb{T}$ is a Galton-Watson tree with $E(\eta(\sigma)) < \lambda_2^{-2}$, then the simulations support the following conjecture:

$$\frac{1}{\sqrt{n}} \sum_{\sigma \in \mathbb{T}} (y(X_\sigma) - E_\pi(y)) \to N(0, \sigma^2),$$

where σ^2 can be computed from the results in [13]. To prove this result requires a more careful study of the structure of $\{X_\sigma\}_{\sigma \in \mathbb{T}}$. We leave this problem to future investigation.

Appendix A: Proof of Theorem 1

In the appendix, we give a proof of the theorems and propositions in the paper. First, we give an outline of the proof of our main theorem. Consider the martingale

$$\sum_h E(Y_h | \mathscr{F}_{h-1}) - Y_h,$$

where $\{\mathscr{F}_h\}$ is a filtration to be defined later. Using the Markov property and the estimation of $var(Y_h)$, we show that the martingale difference sequence satisfies the condition of the martingale central limit theorem. In this section, P will be a reversible transition matrix with eigenvalues $1 = \lambda_1 \geq |\lambda_2| \geq ... \geq |\lambda_N|$ and corresponding eigenfunctions $f_1, ..., f_N$ satisfying $\sum_k f_i(k) f_j(k) \pi_k = \delta_{ij}$ for any i, j. We refer to [10] for the existence of such eigen-decomposition. Unless stated otherwise, expectations are calculated with respect to the tree indexed random walk on the graph.

We begin with some lemmas.

Lemma 1. *(Lemma 12.2 in [10]) Let P be a reversible Markov transition matrix on the nodes in G with respect to the stationary distribution π. The eigenvectors of P, denoted as $f_1, \ldots, f_N$, are real-valued functions of the nodes $i \in G$ and orthonormal with respect to the inner product*

$$\langle f_a, f_b \rangle_\pi = \sum_{i \in G} f_a(i) f_b(i) \pi_i. \tag{A.1}$$

If λ is an eigenvalue of P, then $|\lambda| \leq 1$. The eigenfunction f_1 corresponding to the eigenvalue 1 is taken to be the constant vector $\mathbf{1}$. If $X(0), \ldots, X(t)$ represent t steps of a Markov chain with transition matrix P, then the probability of a transition from $i \in G$ to $j \in G$ in t steps can be written as

$$P(X(t) = j | X(0) = i) = P_{ij}^t = \pi_j + \pi_j \sum_{\ell=2}^{N} \lambda_\ell^t f_\ell(i) f_\ell(j). \tag{A.2}$$

Lemma 2. *For any nodes* σ, τ *in* $\mathbb{T}$,

$$cov(y(X_\sigma), y(X_\tau)) = \sum_{l=2}^{N} \lambda_\ell^{d(\sigma,\tau)} \langle y, f_\ell \rangle_\pi^2,$$

where $\langle y, f_\ell \rangle_\pi = \sum_{i=1}^{N} y(i) f_\ell(i) \pi_i$.

Proof. From the reversibility of the Markov chain and Lemma 1, we have

$$P(X_\sigma = j | X_\tau = i) = P_{ij}^{d(\sigma,\tau)} = \pi_j + \pi_j \sum_{l=2}^{N} \lambda_\ell^{d(\sigma,\tau)} f_\ell(i) f_\ell(j).$$

Therefore,

$$cov(y(X_\sigma), y(X_\tau)) = \sum_{i,j} y(i) y(j) \pi_i P(X_\sigma = j | X_\tau = i) - \left(\sum_i \pi_i y_i \right)^2$$

$$= \sum_{i,j} y(i) y(j) \pi_i \pi_j \sum_{l=2}^{N} \lambda_\ell^{d(\sigma,\tau)} f_\ell(i) f_\ell(j)$$

$$= \sum_{l=2}^{N} \lambda_\ell^{d(\sigma,\tau)} \langle y, f_\ell \rangle_\pi^2,$$

and the lemma is proved.

The next lemma gives the expression of $var(Y_h)$. $\qquad \square$

Lemma 3 (Variance of Y_h). *Suppose that* $|\lambda_2| > 0$. *Then as* $h \to \infty$,

$$var(Y_h) = \begin{cases} O(1) & if \quad m < \lambda_2^{-2}, \\ O(h) & if \quad m = \lambda_2^{-2}, \\ O((m\lambda_2^2)^h) & if \quad m > \lambda_2^{-2}. \end{cases}$$

Proof. For $k = 0, 1, ..., h$, denote by s_{hk} the number of ordered pairs (σ, τ) such that $|\sigma| = |\tau| = h$ and $d(\sigma, \tau) = 2k$. Then $s_{h0} = m^h$, and

$$s_{hk} = m^{h-k} m(m-1)(m^{k-1})^2 = m^{h+k} - m^{h+k-1}$$

for $k \geq 1$. By Lemma 2,

$$var(Y_h) = \frac{1}{m^h} \sum_{k=0}^{h} \left(s_{hk} \sum_{l=2}^{N} \lambda_\ell^{2k} \langle y, f_\ell \rangle_\pi^2 \right)$$

$$= \sum_{l=2}^{N} \left(\langle y, f_\ell \rangle_\pi^2 \left(1 + \sum_{k=0}^{h} m^{k-1} (m-1) \lambda_\ell^{2k} \right) \right)$$

$$\leq \sum_{l=2}^{N}(\langle y, f_\ell\rangle_\pi^2 \sum_{k=0}^{h}(m\lambda_2^2)^k).$$

Thus

$$var(Y_h) = \begin{cases} O(1) & \text{if} \quad m < \lambda_2^{-2}, \\ O(h) & \text{if} \quad m = \lambda_2^{-2}, \\ O((m\lambda_2^2)^h) & \text{if} \quad m > \lambda_2^{-2}. \end{cases} \qquad \square$$

Corollary 2. *For any function y on the state space, $\frac{1}{\sqrt{m^h}}Y_h \to 0$ in probability.*

Proof. It follows from Lemma 3 that

$$var(\frac{1}{\sqrt{m^h}}Y_h) = O(\lambda_2^{2h}) \to 0. \qquad \square$$

The next lemma is a convergence argument which we will use in the proof of Theorem 1.

Lemma 4 (Slutsky's lemma). *If $X_h \to X$ in distribution and $Y_h \to 0$ in probability, then $X_h + Y_h \to X$ in distribution.*

The following theorem from [3] is essential to the proof of our main theorem.

Theorem 3 (Martingale central limit theorem). *Suppose that $\{Z_h\}_{h\geq 1}$ is adapted to the filtration $\{\mathscr{F}_h\}_{h\geq 1}$ and that $E(Z_{h+1}|\mathscr{F}_h) = 0$ for all $h \geq 1$. Let $S_h = \sum_{i=1}^{h} Z_i$ and $V_h = \sum_{i=1}^{h} E(Z_i^2|\mathscr{F}_{i-1})$. If*
(1) $V_h/h \to \sigma^2 > 0$ in probability and
(2) $h^{-1}\sum_{i=1}^{h} E(Z_i^2 1_{\{|Z_i|>\epsilon\sqrt{h}\}}) \to 0$ for every $\epsilon > 0$,
then $S_h/\sqrt{h} \to N(0, \sigma^2)$ in distribution.

Now we are ready to prove our main theorems.

Proof of Theorem 1. Define Y_h in the same way as Theorem 1. Without loss of generality, suppose that $E_\pi(y) = 0$. Since $m < \lambda_2^{-2}$, $\sqrt{m}P - I$ is invertible. Let $y' = (\sqrt{m}P - I)^{-1}y$. Then y' is also a function on the state space. We will first argue on the new node feature y' and then convert back to y. Define

$$Y_h' = \frac{1}{\sqrt{m^h}} \sum_{\tau \in \mathbb{T}:|\tau|=h} y'(X_\tau).$$

Let $z_{hk} = \sum_{\sigma:|\sigma|=h} 1_{\{X_\sigma=k\}}$ for $h \geq 0$ and $k = 1, ..., N$. Define $z_h = \{z_{h1}, ..., z_{hN}\}$, and

$$\mathscr{F}_h = \sigma(X_\tau : |\tau| \leq h)$$

for $h \geq 1$. It is obvious that $\{Y_h\}_{h\geq 1}$ is adapted to the filtration $\{\mathscr{F}_h\}_{h\geq 1}$. Let

$$Z_h = E(Y_h'|\mathscr{F}_{h-1}) - Y_h'.$$

Then $\{Z_h, \mathscr{F}_h\}_{h\geq 1}$ is a martingale difference sequence. We will verify that $\{Z_h, \mathscr{F}_h\}_{h\geq 1}$ satisfies (1) and (2) in Theorem 3.

We have

$$Z_h = m z_{h-1} P y' / \sqrt{m^h} - Y_h' = \frac{z_{h-1}^T (\sqrt{m} P) y'}{\sqrt{m^{h-1}}} - \frac{z_h^T y'}{\sqrt{m^h}}.$$

For any $\sigma \in \mathbb{T}$, denote by $p(\sigma)$ the parent node of σ. Z_h can also be expressed as

$$Z_h = \sum_{\sigma: |\sigma| = h} E\left(\frac{y'(X_\sigma)}{\sqrt{m^h}} \Big| \mathscr{F}_{h-1}\right) - \frac{y'(X_\sigma)}{\sqrt{m^h}}$$

$$= \sum_{\sigma: |\sigma| = h} \frac{1_{p(\sigma)} P y' - y'(X_\sigma)}{\sqrt{m^h}} = \sum_{\sigma: |\sigma| = h} W_\sigma,$$

where $1_{p(\sigma)}$ is the $1 \times N$ vector with $1_{p(\sigma), i} = 1$ if $X_{p(\sigma)} = i$ and 0 otherwise.

For any σ with $|\sigma| = h$, we have

$$E(W_\sigma | \mathscr{F}_{h-1}) = 0$$

and

$$E(W_\sigma^2 | \mathscr{F}_{h-1}) = \frac{Var_{p_i}(y')}{m^h}$$

for $i = X_{p(\sigma)}$, where p_i is the ith row of the transition matrix P and $Var_{p_i}(y') = \sum_j p_{ij}(y(j) - \sum_j p_{ij} y(j))^2$. From the definition of tree indexed Markov process, if $|\sigma| = |\tau| = h$, then W_σ, W_τ are independent given $\{X_\sigma : |\sigma| = h - 1\}$. Using $E(W_\sigma | \mathscr{F}_{h-1}) = 0$, we have

$$E(Z_h^2 | \mathscr{F}_{h-1}) = \sum_{\sigma: |\sigma| = h} E(W_\sigma^2 | \mathscr{F}_{h-1}) = \sum_{i=1}^N \frac{z_{hi}}{m^h} Var_{p_i}(y').$$

From Corollary 2, $var(\frac{z_{hi}}{m^h}) = O(\lambda_2^{2h}) \to 0$ for every i. Thus $var(E(Z_h^2 | \mathscr{F}_{h-1})) = O(\lambda_2^{2h}) \to 0 (h \to \infty)$ and $\sum_{i=1}^h var(E(Z_i^2 | \mathscr{F}_{i-1}))$ converges. It follows from the definition of V_h and the Cauchy-Schwarz inequality that

$$\lim_{h \to \infty} var(V_h / h) = \lim_{h \to \infty} var\left(\frac{1}{h} \sum_{i=1}^h E(Z_i^2 | \mathscr{F}_{i-1})\right)$$

$$\leq \lim_{h \to \infty} \frac{1}{h} \sum_{i=1}^h var(E(Z_i^2 | \mathscr{F}_{i-1})) = 0.$$

Therefore

$$V_h / h \to \sigma^2$$

in probability, where

$$\sigma^2 = E(E(Z_h^2 | \mathscr{F}_{h-1})) = \sum_{i=1}^N \pi_i Var_{p_i}(y') = \sum_{i=1}^N \pi_i \left(\sum_{j=1}^N p_{ij} y'(j)^2 - \left(\sum_{j=1}^N p_{ij} y'(j)\right)^2\right)$$

$$= var_\pi(y') - var_\pi(Py'),$$

and condition (1) in Theorem 3 is satisfied.

Similarly, we have

$$E(Z_h^4|\mathscr{F}_{h-1}) = \sum_{\sigma:|\sigma|=h} E(W_\sigma^4|\mathscr{F}_{h-1}) + \sum_{\sigma,\tau:|\sigma|=|\tau|=h} E(W_\sigma^2 W_\tau^2|\mathscr{F}_{h-1})$$

$$\leq \frac{C_0}{m^h} + C_1 \frac{m^h(m^h-1)}{m^{2h}} \leq C,$$

where C_0, C_1, C are constants. Thus $E(Z_h^4) \leq C$ for any h, and

$$h^{-1} \sum_{i=1}^h E(Z_i^2 1_{\{|Z_i|>\epsilon\sqrt{h}\}}) \leq h^{-1} \sum_{i=1}^h E(Z_i^2 \frac{Z_i^2}{\epsilon^2 h}) = \frac{1}{\epsilon^2 h^2} \sum_{i=1}^h E(Z_i^4) \leq \frac{C}{h} \to 0.$$

Condition (2) is also satisfied. From Theorem 3, we have

$$\frac{1}{\sqrt{h}} \sum_{i=1}^h Z_i = \frac{1}{\sqrt{h}} \sum_{i=1}^h \left(\frac{z_{i-1}^T(\sqrt{m}P)y'}{\sqrt{m^{i-1}}} - \frac{z_i^T y'}{\sqrt{m^i}} \right) \to N(0,\sigma^2)$$

in distribution. If $m < \lambda_2^{-2}$, then from Lemma 3, $\frac{z_h^T(\sqrt{m}P)y'}{\sqrt{m^{h-1}}} \to 0$ in probability. From Lemma 4 and the definition of y',

$$\frac{1}{\sqrt{h}} \sum_{i=1}^h \frac{z_i^T(\sqrt{m}P - I)y'}{\sqrt{m^i}} = \frac{1}{\sqrt{h}} \sum_{i=1}^h Y_i \to N(0,\sigma^2)$$

in distribution, where $\sigma^2 = var_\pi(y') - var_\pi(Py') = var_\pi((\sqrt{m}P - I)^{-1}y) - var_\pi(P(\sqrt{m}P - I)^{-1}y)$. The proof is now complete. □

Appendix B: Proof of Theorem 2 and Corollary 1

We provide a proof of the central limit theorem using the moment method. It involves a careful study of all the moments of $\hat{\mu}_h$. The following proposition is essential to our proof.

Proposition 4. *(Moment continuity theorem) Let X_h be a sequence of uniformly subgaussian real random variables, and let X be another subgaussian random variable. Then the following statements are equivalent:*

(1)$EX_h^k \to EX^k$ for all k

(2)$X_h \to X$ in distribution

Following this proposition, we can break our proof into two parts. We will first prove that all the moments of $\hat{\mu}_h$ converge to the moments of some normal distribution. Then we will verify that $\hat{\mu}_h$ is a uniformly subgaussian sequence.

B.1. Proof of moments convergence

Let X_r be the root of the 2-tree. Define

$$\gamma_{k,h}(i) = E[\hat{\mu}_h^k | X_r = i],$$

and

$$\gamma_{k,h} = E[\hat{\mu}_h^k].$$

Let $\rho = \sqrt{2}|\lambda_2| < 1$. We prove that there exist γ_k, $k \geq 1$ such that $|\gamma_{k,h}(i) - \gamma_k| = O(\rho^h)$ for all k, i, h and that $\gamma_k = E(\xi^k)$ for $\xi \sim N(0, \gamma_2)$. Our key observation is that the left and right subtree can be seen as i.i.d copies of the whole tree given the left and right child of the seed, which makes it possible to build a relationship between $\gamma_{k,h}(i)$ and $\gamma_{k,h-1}(i)$. Only condition (c3) is needed throughout the proof.

We need the following Lemma.

Lemma 5. *Let* $\{a_h\}$ *be a sequence satisfying*

$$a_h = c_h(ca_{h-1} + C + d_h),$$

where $|1 - c_h| = O(\rho^h)$, $|d_h| = O(\rho^h)$, C *is a constant and* $c < \rho < 1$. *Then* $|a_h - C/(1 - c)| = O(\rho^h)$.

Proof. Without loss of generality, suppose that $c_h \neq 0$ and $C = 0$. Since $|1 - c_h| = O(\rho^h)$ and $|d_h| = O(\rho^h)$, there exists M such that $\prod_{k=1}^h |c_k| \leq M$ and $|d_h| \leq M\rho^h$ for all h. Therefore,

$$|a_h| \leq \sum_{k=0}^h c^{h-k} d_k \prod_{i=h-k}^h |c_i| \leq \frac{M^2}{\rho - c}\rho^h,$$

and the lemma is proved. $\square$

We use an induction on k. First, we will prove that $\gamma_1 = 0$. In fact, from Lemma 1,

$$E(y(X_\sigma)|X_r = i) = \sum_{j=1}^N y(j) P_{ij}^{|\sigma|} = O(|\lambda_2|^{|\sigma|}).$$

Therefore

$$E(\hat{\mu}_h | X_r = i) = \frac{1}{\sqrt{2}^h} \sum_{k=1}^h 2^k O(|\lambda_2|^k) = O(\rho^h)$$

for all i, and $|\gamma_{1,h}(i) - \gamma_1| = O(\rho^h)$ for $\gamma_1 = 0$.

Now we move from $k - 1$ to k. Without loss of generality, suppose that $\gamma_{2,h}(i) > 1$ for all h, i (or we can multiply y with a large constant). It follows that $\gamma_{2k,h}(i) \geq (\gamma_{2,h}(i))^k > 1$ for all k. We can decompose $\gamma_{k,h}(i)$ into

$$\gamma_{k,h}(i) = E[\hat{\mu}_h^k | X_r = i] - E[(\frac{1}{\sqrt{2}^h} \sum_{\sigma \in \mathbb{T}, 0 < |\sigma| \leq h} X_\sigma)^k | X_r = i]$$

$$+ E[(\frac{1}{\sqrt{2}^h} \sum_{\sigma \in \mathbb{T}, 0 < |\sigma| \le h} X_\sigma)^k | X_r = i] \tag{B.1}$$

$$:= I_1 + I_2.$$

If k is even, then we know that

$$
\begin{aligned}
|I_1| =& |E[\hat{\mu}_h^k | X_r = i] - E[(\hat{\mu}_h - \frac{i(s)}{\sqrt{2}^h})^k | X_r = i]| \\
=& \sum_{m=1}^{k} \binom{k}{m} (-i\sqrt{2}^{-h})^m E[\hat{\mu}_h^{k-m} | X_r = i] \\
\le& \sum_{m=1}^{k} \binom{k}{m} (|i|\sqrt{2}^{-h})^m (E[\hat{\mu}_h^k | X_r = i])^{\frac{k-m}{k}} \\
\le& \sum_{m=1}^{k} \binom{k}{m} (|i|\sqrt{2}^{-h})^m (E[\hat{\mu}_h^k | X_r = i]) \\
\le& [(1 + M\sqrt{2}^{-h})^k - 1] E[\hat{\mu}_h^k | X_r = i],
\end{aligned}
\tag{B.2}
$$

where the first inequality follows from Jensen's inequality and the second inequality follows from our assumption that $\gamma_{2k,h}(i) > 1$ for all k. Likewise, if k is odd, then we have

$$
\begin{aligned}
|I_1| =& E[\hat{\mu}_h^k | X_r = i] - E[(\hat{\mu}_h - \frac{i}{\sqrt{2}^h})^k | X_r = i] \\
\le& [(1 + M\sqrt{2}^{-h})^k - 1] E[\hat{\mu}_h^{k-1} | X_r = i].
\end{aligned}
\tag{B.3}
$$

Since k is fixed, $E[\hat{\mu}_h^{k-1} | X_r = i]$ is bounded from our assumption on $\gamma_{k-1,h}(i)$, and $(1 + M\sqrt{2}^{-h})^k - 1 = O(\sqrt{2}^{-h}) = O(\rho^h)$. Hence,

$$|I_1| \le [(1 + M\sqrt{2}^{-h})^k - 1] E[\hat{\mu}_h^{k-1} | X_r = i] = O(\rho^h).$$

Let X_{lc} and X_{rc} be the left and right child of the root and $\mathbb{T}_l$ and $\mathbb{T}_r$ the left and right subtree, and we have

$$
\begin{aligned}
& E[(\hat{\mu}_h - \frac{i}{\sqrt{2}^h})^k | X_r = i] \\
=& E((\frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}^{h-1}} \sum_{\sigma \in \mathbb{T}_r, |\sigma| \le h} X_\sigma + \frac{1}{\sqrt{2}^{h-1}} \sum_{\sigma \in \mathbb{T}_l, |\sigma| \le h} X_\sigma))^k | X_r = i) \\
=& \sum_{u,v} p_{iu} p_{iv} E((\frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}^{h-1}} \sum_{\sigma \in \mathbb{T}_r, |\sigma| \le h} X_\sigma \\
& \qquad\qquad + \frac{1}{\sqrt{2}^{h-1}} \sum_{\sigma \in \mathbb{T}_l, |\sigma| \le h} X_\sigma))^k | X_r = i, X_{lc} = u, X_{rc} = v)
\end{aligned}
$$

$$= \sum_{u,v} p_{iu}p_{iv}E((\frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}^{h-1}} \sum_{\sigma \in \mathbb{T}_r, |\sigma| \leq h} X_\sigma$$

$$+ \frac{1}{\sqrt{2}^{h-1}} \sum_{\sigma \in \mathbb{T}_l, |\sigma| \leq h} X_\sigma))^k | X_{lc} = u, X_{rc} = v)$$

$$= \sum_{u,v} p_{iu}p_{iv}\frac{1}{\sqrt{2}^k}(\gamma_{k,h-1}(u) + \gamma_{k,h-1}(v)) + S_1$$

$$= \frac{1}{\sqrt{2}^{k-2}} \sum_{u} p_{iu}\gamma_{k,h-1}(u) + S_1, \tag{B.4}$$

where

$$S_1 = \sum_{m=1}^{k-1} \frac{1}{\sqrt{2}^k} \binom{k}{m} \sum_{u,v} p_{iu}p_{iv}\gamma_{m,h-1}(u)\gamma_{k-m,h-1}(v).$$

If $k = 2$, Equation B.2 and B.4 reduce to

$$\gamma_{k,h}(i) = \sum_u p_{iu}\gamma_{k,h-1}(u) + \delta_h(i),$$

where

$$\delta_h(i) = \frac{y(i)^2}{2^h} - \frac{2y(i)}{\sqrt{2}^h}\gamma_{1,h}(i) + (\sum_u p_{iu}\gamma_{1,h}(u))^2 = O(\rho^h).$$

Write $\nu_h = \{\gamma_h^2\}'$ and $\delta_h = \{\delta_h\}'$. For $n \geq 2$,

$$\nu_h = P\nu_{h-1} + \delta_h.$$

Thus by setting $\delta_1 = 0$ we have $\nu_h = P^h\nu_1 + \sum_{k=1}^h P^k\delta_{h-k}$, and it is not hard to verify that all the components of ν_h (i.e., every $\gamma_{k,h}(i)$) converge to $\gamma_2 = \pi^t\nu_1 + \sum_{h=1}^\infty \pi^t\delta_h$ with rate ρ^h.

Now suppose that $k > 2$. Since k is fixed, there are a fixed number of terms in S_1 as h goes to infinity. Since $|\gamma_{l,h}(i) - \gamma_l| = O(\rho^h)$ for all $i \in S$ and $l < k-1$, we have

$$|S_1 - \sum_{m=1}^{k-1} \frac{1}{\sqrt{2}^k} \binom{k}{m}\gamma_m\gamma_{k-m}| = O(\rho^h).$$

Thus,

$$I_2 = \frac{1}{\sqrt{2}^{k-2}} \sum_u p_{iu}\gamma_{k,h-1}(u) + \sum_{m=1}^{k-1} \frac{1}{\sqrt{2}^k} \binom{k}{m}\gamma_m\gamma_{k-m} + O(\rho^h). \tag{B.5}$$

Combining Equations B.2, B.3 and B.5 we arrive at the final equation for $\gamma_{k,h}(i)$:

$$\gamma_{k,h}(i) = c_{k,h}I_2$$

$$= c_{k,h}\left(\frac{1}{\sqrt{2}^{k-2}}\sum_u p_{iu}\gamma_{k,h-1}(u) + \sum_{m=1}^{k-1}\frac{1}{\sqrt{2}^k}\binom{k}{m}\gamma_m\gamma_{k-m} + O(\rho^h)\right),$$

(B.6)

where $c_{k,h} = 1$ if k is odd and $c_{k,h} = 1 + O(\rho^h) \in [2 - (1 + M\sqrt{2}^{-h})^k, (1 + M\sqrt{2}^{-h})^k]$ if k is even. Since $\frac{1}{\sqrt{2}^{k-2}} < \rho$ and $\prod_{i=1}^h c_{k,h}$ converges, we conclude from Lemma 5 that

$$|\gamma_{k,h}(i) - \gamma_k| = O(\rho^h),$$

where $\gamma_k = \sum_{m=1}^{k-1}\frac{1}{\sqrt{2}^k}\binom{k}{m}\gamma_m\gamma_{k-m}/(1 - \frac{1}{\sqrt{2}^{k-2}})$.

We proved that $|\gamma_{k,h}(i) - \gamma_k| = O(\rho^h)$ for all k, i, h. Let h tend to infinity in Equation (B.5), we have

$$\gamma_k = \frac{1}{\sqrt{2}^{k-2}}\gamma_k + \sum_{m=1}^{k-1}\frac{1}{\sqrt{2}^k}\binom{k}{m}\gamma_m\gamma_{k-m}.$$

(B.7)

Now suppose that $\xi_1 \sim N(0, \gamma_2)$ and $\tilde{\gamma}_k = E\xi_1^k$. Let ξ_2 be an i.i.d copy of ξ_1. Then

$$\tilde{\gamma}_k = E\xi_1^k = E\left(\frac{\xi_1 + \xi_2}{\sqrt{2}}\right)^k = \frac{1}{\sqrt{2}^{k-2}}\tilde{\gamma}_k + \sum_{m=1}^{k-1}\frac{1}{\sqrt{2}^k}\binom{k}{m}\tilde{\gamma}_m\tilde{\gamma}_{k-m}.$$

Hence $\{\tilde{\gamma}_k\}, k \in \mathbb{N}$ also follows Equation (B.7). Since $\gamma_1 = \tilde{\gamma}_1 = 0$ and $\gamma_2 = \tilde{\gamma}_2$, we have $\gamma_k = \tilde{\gamma}_k$ for every k, and the argument is proved.

B.2. Proof of uniform sub-gaussianity

To prove that $\hat{\mu}_h$ are uniformly sub-gaussian for all h, we need to show that there exists some θ such that

$$\gamma_{2\ell,h}(i) \leq \theta^{2\ell}\gamma_{2\ell}$$

for all ℓ.

Let c_1 be a large constant to be defined later and $c_{h+1} = (1 + M(2\lambda_2')^{-h})(1 + (\sqrt{2}\lambda_2')^h)c_h$, where $\lambda_2' = max\{|\lambda_2|, 2/3\}$ and $M = ||y||_\infty$. Let $s_{\ell,h} = ||\gamma_{\ell,h}(i)||_\infty = max_i |\gamma_{\ell,h}(i)|$. Since $0 < \sqrt{2}\lambda_2' < 1 < 2\lambda_2'$, $c_{h+1} > c_h$ and $\theta = \lim_{h\to\infty} c_h$ exists. Thus it suffices to prove that

$$s_{2\ell,h} \leq c_h^{2\ell}\gamma_{2\ell}$$

(B.8)

and

$$s_{2\ell-1,h} \leq c_h^{2\ell-1}(\sqrt{2}\lambda_2')^h\gamma_{2\ell}$$

(B.9)

for all ℓ and h.

Again we use an induction on ℓ. Since $\gamma_{1,h}(i) = O(|\lambda_2|^h)$, we can choose c_1 large enough such that the inequalities in Equations B.8 and B.9 hold for all

(h, ℓ) with $h = 1$ or $\ell = 1$. Suppose that Equations B.8 and B.9 are verified for all $\ell \leq k$. We will prove that they are also true for $\ell = k + 1$.

By condition (c1) and (c2), we know that

$$|| \sum_u p_{iu} \gamma_{2k+1,h}(u) ||_\infty \leq \lambda_2 ||\gamma_{2k+1,h}(i)||_\infty = \lambda_2 s_{2k+1,h}.$$

From our assumption of induction we have

$$s_{2k+1,h} \leq [(1 + M\sqrt{2}^{-h})^{2k+1} - 1]c_{h-1}^{2k}\gamma_{2k} + c_{h-1}^{2k+1}\frac{(|\sqrt{2}\lambda_2'|)^h}{\sqrt{2}^{2k+2}}(\binom{2k+1}{0}\gamma_{2k+2}$$

$$+ \sum_{m=1}^{k}(\binom{2k+1}{2m-1} + \binom{2k+1}{2m})\gamma_{2k+2-2m}\gamma_{2m} + \binom{2k+1}{2k+1}\gamma_{2k+2})$$

$$= [(1 + M\sqrt{2}^{-h})^{2k+1} - 1]c_{h-1}^{2k}\gamma_{2k} + c_{h-1}^{2k+1}$$

$$\frac{(|\sqrt{2}\lambda_2'|)^h}{\sqrt{2}^{2k+2}}(\binom{2k+2}{0}\gamma_{2k+2} + \sum_{m=1}^{k}\binom{2k+2}{2m}\gamma_{2k+2-2m}\gamma_{2m} + \gamma_{2k+2})$$

$$= [(1 + M\sqrt{2}^{-h})^{2k+1} - 1]c_{h-1}^{2k}\gamma_{2k} + c_{h-1}^{2k+1}(|\sqrt{2}\lambda_2'|)^h\gamma_{2k+2}$$

$$\leq ([(1 + M\sqrt{2}^{-h})^{2k+1} - 1](|\sqrt{2}\lambda_2'|)^{-h} + 1)c_{h-1}^{2k+1}(|\sqrt{2}\lambda_2'|)^h\gamma_{2k+2}$$

$$\leq (1 + M(2\lambda_2')^{-h})^{2k+1}c_{h-1}^{2k+1}(|\sqrt{2}\lambda_2'|)^h\gamma_{2k+2}$$

$$\leq c_h^{2k+1}(|\sqrt{2}\lambda_2'|)^h\gamma_{2k+2},$$

(B.10)

and Equation (B.9) is true for $2k + 1$.

Now we move from $2k + 1$ to $2k + 2$. Recall that

$$\gamma_{2k+2,h}(i) = c_{2k+2,h}\sum_{m=0}^{2k+2}\frac{1}{\sqrt{2}^{2k+2}}\binom{2k+2}{m}\sum_{u,v}p_{iu}p_{iv}\gamma_{m,h-1}(u)\gamma_{2k+2-m,h-1}(v)$$

$$\leq (1 + M2^{-h})^{2k+2}\sum_{m=0}^{2k+2}\frac{1}{\sqrt{2}^{2k+2}}\binom{2k+2}{m}s_{m,h-1}s_{2k+2-m,h-1}.$$

Thus,

$$s_{2k+2,h} \leq (1 + M2^{-h})^{2k+2}(\sum_{m=0}^{2k+2}\frac{1}{\sqrt{2}^{2k+2}}\binom{2k+2}{m}s_{m,h-1}s_{2k+2-m,h-1}).$$

Let

$$I_1 = \sum_{m=0}^{k+1}\frac{1}{\sqrt{2}^{2k+2}}\binom{2k+2}{2m}s_{2m,h-1}s_{2k+2-2m,h-1}$$

and

$$I_2 = \sum_{m=0}^{k}\frac{1}{\sqrt{2}^{2k+2}}\binom{2k+2}{2m+1}s_{2m+1,h-1}s_{2k+1-2m,h-1}.$$

Then $s_{2k+2,h} \leq (1 + M2^{-h})^{2k+2}(I_1 + I_2)$.

We have

$$
\begin{aligned}
I_1 &\leq \sum_{m=0}^{k+1} \frac{1}{\sqrt{2}^{2k+2}} \binom{2k+2}{2m} c_{h-1}^{2m} \gamma_{2m} c_{h-1}^{2k+2-2m} \gamma_{2k+2-2m} \\
&= c_{h-1}^{2k+2} \sum_{m=0}^{k+1} \frac{1}{\sqrt{2}^{2k+2}} \binom{2k+2}{2m} \gamma_{2m} \gamma_{2k+2-2m} \\
&= c_{h-1}^{2k+2} \gamma_{2k+2},
\end{aligned}
\tag{B.11}
$$

where the last equality follows from Equation (B.7). On the other hand

$$
\begin{aligned}
I_2 &\leq \sum_{m=0}^{k} \frac{1}{\sqrt{2}^{2k+2}} \binom{2k+2}{2m+1} c_{h-1}^{2m+1} (\sqrt{2}\lambda_2')^{h-1} \\
&\quad \times \gamma_{2m+2} c_{h-1}^{2k+1-2m} (\sqrt{2}\lambda_2')^{h-1} \gamma_{2k+2-2m} \\
&= c_{h-1}^{2k+2} (\sqrt{2}\lambda_2')^{2(h-1)} \sum_{m=0}^{k} \frac{1}{\sqrt{2}^{2k+2}} \binom{2k+2}{2m+1} \gamma_{2m+2} \gamma_{2k+2-2m}.
\end{aligned}
\tag{B.12}
$$

It can be directly verified that for all m,

$$
\binom{2k+2}{2m+1} \gamma_{2m+2} \gamma_{2k+2-2m} \leq 2(k+1) \binom{k}{m} \gamma_{2k+2}.
$$

Thus,

$$
\begin{aligned}
I_2 &\leq c_{h-1}^{2k+2} (\sqrt{2}\lambda_2')^{h-1} \sum_{m=0}^{k} \frac{1}{\sqrt{2}^{2k+2}} 2(k+1) \binom{k}{m} \gamma_{2k+2} \\
&= c_{h-1}^{2k+2} (\sqrt{2}\lambda_2')^{2(h-1)} (k+1) \gamma_{2k+2} \\
&\leq c_{h-1}^{2k+2} (\sqrt{2}\lambda_2')^h 2k \gamma_{2k+2}.
\end{aligned}
\tag{B.13}
$$

Combining Equation (B.11) and B.13, we have

$$
\begin{aligned}
I_1 + I_2 &\leq c_{h-1}^{2k+2} \gamma_{2k+2} (1 + 2k * (\sqrt{2}\lambda_2')^h) \\
&\leq c_{h-1}^{2k+2} \gamma_{2k+2} (1 + (\sqrt{2}\lambda_2')^h)^{2k+2}.
\end{aligned}
\tag{B.14}
$$

Therefore,

$$
\begin{aligned}
s_{h,2k+2} &\leq c_{2k+2,h}(I_1 + I_2) \\
&\leq c_{h-1}^{2k+2} (1 + M2^{-h})^{2k+2} (1 + (\sqrt{2}\lambda_2')^h)^{2k+2} \gamma_{2k+2} \\
&\leq c_h^{2k+2} \gamma_{2k+2},
\end{aligned}
\tag{B.15}
$$

and the theorem is proved.

B.3. Proof of Corollary 1

By Theorem 2 and Slutsky's lemma, it suffices to prove that $\hat{\bar{d}} \to \bar{d}$ in probability. Let $D = max_{1 \leq i \leq N} deg(X_i)$. For any $\sigma \in \mathbb{T}$, $E_\pi \frac{1}{deg(X_\sigma)} = \frac{1}{\bar{d}}$. Thus $E\frac{1}{\hat{\bar{d}}} = \frac{1}{\bar{d}}$, and it follows from Theorem 2 that $\hat{\bar{d}} \to \bar{d}$ in probability. Since $\hat{\bar{d}}, \bar{d} \leq D$, we have $P(|\hat{\bar{d}} - \bar{d}| > \epsilon) \leq P(|\frac{1}{\hat{\bar{d}}} - \frac{1}{\bar{d}}| > \epsilon/D^2) \to 0$ for all $\epsilon > 0$, and the corollary is proved.

Appendix C: Reducing the state space of the Markov chain

This section justifies the simplification in Section 3.2. Recall that $P \in \mathbb{R}^{N \times N}$ is a Markov transition matrix on N nodes, where each node i is assigned to one of two classes $z(i) \in \{1, 2\}$, and

$$P_{ij} = (N/2)^{-1} B_{z(i),z(j)}.$$

Let $X_t \in \{1, \dots, N\}$ for $t \in 0, 1, \dots$ be a Markov chain with transition matrix P that is initialized from the stationary distribution.

One can construct a Markov chain $\{Z_t\}_t$ on the block labels $\{1, 2\}$ that is equal in distribution to $\{z(X_t)\}_t$. Define $Z_t \in \{1, 2\}$ for $t = 0, 1, \dots$ as a Markov chain with transition matrix $\mathscr{P} = B$ and initialize Z_0 from the stationary distribution of $\mathscr{P}$. For $a, b \in \{1, 2\}$ and any $t \geq 0$,

$$
\begin{aligned}
P(Z_t = a, Z_{t+1} = b) &= P(Z_t = a)P(Z_{t+1} = b | Z_t = a) \\
&= (1/2)\mathscr{P}_{a,b} \\
&= \sum_{i,j} \mathbf{1}\{z(i) = a, z(j) = b\}\pi_i P_{i,j} \\
&= \sum_{i,j} \mathbf{1}\{z(i) = a, z(j) = b\}P(X_t = i, X_{t+1} = j) \\
&= P(z(X_t) = a, z(X_{t+1}) = b).
\end{aligned}
$$

Induction shows that $\{Z_t\}_t$ is equal in distribution to $\{z(X_t)\}_t$.

Appendix D: Proofs of Propositions 2 and 3

Before the proofs, some more notation is necessary. Let $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_N|$ denote the eigenvalues of P. Denote the $f_1, \dots, f_N : V \to \mathbb{R}$ as the corresponding eigenfunctions of P.

First, a lemma from [10].

Lemma 6.
$$\sum_{\ell=2}^{N} \frac{\langle \tilde{y}, f_\ell \rangle_\pi^2}{var_\pi(y)} = 1.$$

The following is a proof of Lemma 6 from [10].

Proof of Lemma 6.

$$var_\pi(\tilde{y}) = E_\pi(\tilde{y}^2) - (E_\pi \tilde{y})^2 = \sum_{j=1}^N \langle \tilde{y}, f_j \rangle_\pi^2 - (E_\pi \tilde{y})^2 = \sum_{j=2}^N \langle \tilde{y}, f_j \rangle_\pi^2. \qquad \square$$

The following is a proof of Proposition 3.

Proof of Proposition 3. From Theorem 1 in [13],

$$\sigma_{\hat{\mu}}^2 = \sum_{\ell=2}^N \langle \tilde{y}, f_\ell \rangle_\pi^2 \mathbb{G}(\lambda_\ell) = var_\pi(\tilde{y}) \sum_{\ell=2}^N \frac{\langle \tilde{y}, f_\ell \rangle_\pi^2}{var_\pi(\tilde{y})} \mathbb{G}(\lambda_\ell). \qquad (\text{D.1})$$

From Lemma 6, this is a convex combination of $\mathbb{G}(\lambda_2), \ldots, \mathbb{G}(\lambda_N)$. Applying Jensen's inequality yields

$$\sigma_{\hat{\mu}}^2 = var_\pi(\tilde{y}) \sum_{\ell=2}^N \frac{\langle \tilde{y}, f_\ell \rangle_\pi^2}{var_\pi(\tilde{y})} \mathbb{G}(\lambda_\ell) \geq var_\pi(\tilde{y}) \mathbb{G}\left(\sum_{\ell=2}^N \frac{\langle \tilde{y}, f_\ell \rangle_\pi^2}{var_\pi(y)} \lambda_\ell \right). \qquad (\text{D.2})$$

Moreover, from Proposition 1 in [9],

$$\frac{\sum_{\ell=2}^N \langle \tilde{y}, f_\ell \rangle_\pi^2 \lambda_\ell}{var_\pi(y)} = \frac{Cov(\tilde{y}(X_{p(\tau)}), \tilde{y}(X_\tau))}{var_\pi(y)} = R.$$

Thus,

$$\sigma_{\hat{\mu}}^2 \geq \mathbb{G}(R) var_\pi(\tilde{y}) \qquad \square$$

The proof of Proposition 2 is similar.

Proof of Proposition 2. In the case when $\tilde{y}(i) = \mu + \sigma f(i)$, this implies that $\tilde{y} = \mu f_1 + \sigma f_j$. By the orthonormality of the eigenvectors, $\frac{\langle \tilde{y}, f_\ell \rangle_\pi^2}{var_\pi(\tilde{y})} = \mathbf{1}\{j = \ell\}$ for $\ell > 1$. As such, the inequality in equation (D.2) holds with equality. $\square$

Proposition 3 presumes that $\mathbb{G}$ is convex. Figure 2 plots $\mathbb{G}$ for twenty different Galton-Watson trees with offspring distribution $p(0) = .1$, $p(1) = .1$, $p(2) = .3$, $p(3) = .5$. This offspring distribution has expected value 2.2. The construction of each tree was stopped when it reached 5000 nodes; if it failed to reach 5000 nodes, then the process was started over. In these simulations and in others not shown, $\mathbb{G}$ is often convex. When it is not convex, the second derivative of $\mathbb{G}(z)$ is positive when z is away from -1. This simulation was selected because it shows that even when the trees are sampled from the same distribution, even when there is nothing strange about the offspring distribution (e.g. all moments are finite), even when it is a very big tree, even under all of these nice conditions, some of the trees have a convex $\mathbb{G}$ and some of the trees have a non-convex $\mathbb{G}$. Similar results hold when the trees have 500 nodes; the only thing that changes is that the red regions extend slightly further away from -1.
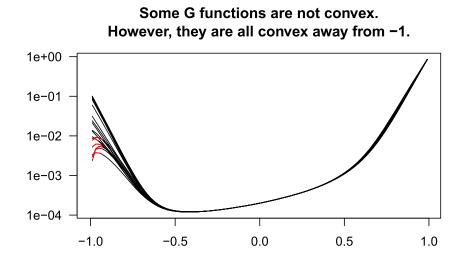
Some G functions are not convex.
However, they are all convex away from −1.

FIG 2. *Each line corresponds to a function* $\mathbb{G}$ *for a random Galton-Watson tree. Some of the curves are not convex; the regions of these functions which have a negative second derivative are highlighted in red. While some of the black lines appear to have a negative second derivative, they do not; this illusion is due to the log transformation on the vertical axis.*

Acknowledgement

References

[1] (2013). *Introduction To HIV/AIDS And Sexually Transmitted Infection Surveillance Module 4: Introduction to Respondent-drive Sampling.* World Health Organization & UNAIDS. http://applications.emro.who.int/dsaf/ EMRPUB_2013_EN_1539.pdf.

[2] Benjamini, I. and Peres, Y. (1994). Markov chains indexed by trees. *The Annals of Probability*, pages 219–243. MR1258875

[3] Durrett, R. (2010). *Probability: theory and examples. Edition 4.1.* Cambridge university press. MR2722836

[4] Gile, K. J. (2011). Improved inference for respondent-driven sampling data with application to hiv prevalence estimation. *Journal of the American Statistical Association*, 106(493). MR2816708

[5] Goel, S. and Salganik, M. J. (2009). Respondent-driven sampling as markov chain monte carlo. *Statistics in medicine*, 28(17):2202–2229. MR2751515

[6] Heckathorn, D. D. (1997). Respondent-driven sampling: a new approach to the study of hidden populations. *Social problems*, pages 174–199.

[7] Holland, P., Laskey, K., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137. MR0718088

[8] Jones, G. L. et al. (2004). On the markov chain central limit theorem. *Probability surveys*, 1(299–320):5–1. MR2068475

[9] Khabbazian, M., Hanlon, B., Russek, Z., and Rohe, K. (2016). Novel sampling design for respondent-driven sampling. *arXiv preprint arXiv:1606.00387*. MR3729659

[10] Levin, D. A., Peres, Y., and Wilmer, E. L. (2009). *Markov chains and mixing times*. American Mathematical Soc. MR2466937

[11] Lorrain, F. and White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, 1(1):49–80.

[12] Lu, X., Bengtsson, L., Britton, T., Camitz, M., Kim, B. J., Thorson, A., and Liljeros, F. (2012). The sensitivity of respondent-driven sampling. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(1):191–216. MR2873802

[13] Rohe, K. (2015). Network driven sampling; a critical threshold for design effects. *arXiv preprint arXiv:1505.05461*.

[14] Rohe, K., Yu, B., and Chatterjee, S. (2011). Spectral clustering and the high dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915. MR2893856

[15] Salganik, M. J. (2006). Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *Journal of Urban Health*, 83(1):98–112.

[16] Salganik, M. J. and Heckathorn, D. D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological methodology*, 34(1):193–240.

[17] Volz, E. and Heckathorn, D. D. (2008). Probability based estimation theory for respondent driven sampling. *Journal of Official Statistics*, 24(1):79.

[18] White, R. G., Hakim, A. J., Salganik, M. J., Spiller, M. W., Johnston, L. G., Kerr, L., Kendall, C., Drake, A., Wilson, D., Orroth, K., et al. (2015). Strengthening the reporting of observational studies in epidemiology for respondent-driven sampling studies:"strobe-rds" statement. *Journal of clinical epidemiology*, 68(12):1463–1471.