Thirty Years of The Network Scale-up Method

Ian Laga, Le Bao, and Xiaoyue Niu
Department of Statistics, Pennsylvania State University

June 7, 2021

Abstract

Estimating the size of hard-to-reach populations is an important problem for many fields. The Network Scale-up Method (NSUM) is a relatively new approach to estimate the size of these hard-to-reach populations by asking respondents the question, "How many X's do you know," where X is the population of interest (e.g. "How many female sex workers do you know?"). The answers to these questions form Aggregated Relational Data (ARD). The NSUM has been used to estimate the size of a variety of subpopulations, including female sex workers, drug users, and even children who have been hospitalized for choking. Within the Network Scale-up methodology, there are a multitude of estimators for the size of the hidden population, including direct estimators, maximum likelihood estimators, and Bayesian estimators. In this article, we first provide an in-depth analysis of ARD properties and the techniques to collect the data. Then, we comprehensively review different estimation methods in terms of the assumptions behind each model, the relationships between the estimators, and the practical considerations of implementing the methods. We apply many of the models discussed in the review to one canonical data set and compare their performance and unique features, presented in the supplementary materials. Finally, we provide a summary of the dominant methods and an extensive list of the applications, and discuss the open problems and potential research directions in this area.

Keywords: Size estimation, small area estimation, key populations, aggregated relational data.

1   Introduction

Estimating the size of hard-to-reach populations is an important problem in a variety of contexts. Governments and humanitarian organizations which aim to eradicate infectious diseases and improve the lives of citizens through treatment programs are interested in population sizes because the treatment target needs to be clear and funds need to be allocated correctly. The Joint United Nations Programme on HIV and AIDS (UNAIDS) aims to limit the spread of HIV by locating large HIV populations. Groups that are particularly vulnerable are known as target- or key-populations. For example, female sex workers (FSW) are among the highest subpopulation living with HIV. It is difficult to estimate the size of FSW directly because of social stigma around sex work and because FSW comprise a relatively small percentage of the general population. Existing approaches to estimate these hard-to-reach populations include mark-recapture, mapping, and venue-based surveys. See Bernard et al. (2010) for a detailed list of population size estimation methods and Sabin et al. (2016) for a comparison of the availability and quality of different data types when estimating certain key subpopulations.

A relatively new method for estimating the size of key populations is the Network Scale-up Method (NSUM), based on the basic scale-up model (Bernard et al., 1989). The authors were in Mexico soon after an earthquake and were interested in estimating the number of people who had died in the earthquake. In this case, where the target population is people who have died from an earthquake, many existing methods were impossible to implement. One author asked people around Mexico City how many people they knew who had died in the earthquake. By leveraging only the responses about how many people each respondent knows, they were able to estimate the number of people who died from the earthquake. The NSUM uses questions from "How many X's do you know?" surveys to estimate both

average network size and subpopulation sizes.

The method provides a relatively cheap, easy, and powerful tool for researchers and can still be applied when it is impossible or difficult to reach the target population directly. Furthermore, the NSUM respects the privacy of the respondents since respondents are not asked about their own characteristics or identifying information of their alters. UNAIDS and WHO published guidelines for the NSUM and noted that the NSUM is advantageous because respondents do not need to reveal their own status, questions can be added to existing household surveys, and sizes of multiple subpopulations can be estimated from one survey (UNAIDS and WHO, 2010). Furthermore, obtaining data through NSUM can be $70 - 80\%$ cheaper than collecting traditional network data (Breza et al., 2017). However, due to biases from violating method assumptions, the UNAIDS and WHO designated the NSUM as a method "under-development" for estimating hard-to-reach populations (UNAIDS and WHO, 2010), prompting the need for further development.

Here we standardize the notation we will use for the rest of the paper and provide a brief introduction to relevant network and NSUM terminology and notation to streamline reading of later sections. Aggregated relational data (ARD) refers to the data collected using "How many X's do you know?" questions, while NSUM refers to the process of estimating network size or subpopulation size using ARD. Let N be the size of the general population, N_k be the size of subpopulation k, and n be the number of respondents. We denote the ARD responses by y_{ik}, the number of people that respondent i reports knowing in subpopulation k. The network size, or degree, for person i is d_i and for ARD, typically represents the total number of individuals that the respondent recognizes by sight or name, has contacted in the last several years, and can still contact. The degree is typically equal to the number of edges connected to the respondent node in an undirected graph

since the connections are generally assumed to be symmetric, although more precisely is equal to the number of out-going edges from the respondent node in a directed graph (e.g. respondents can still know people who have died in an earthquake). Let L be the number of subpopulations with known sizes. For simplicity, we assume only one subpopulation size is unknown, denoted by N_u, but stress that any number of subpopulations can be unknown in practice. The popular McCarty ARD dataset (McCarty et al., 2001), contained 29 known subpopulations which included 12 names (Michael, James, Kimberly), Native Americans, commercial pilots, and homicide victims. The three unknown subpopulations were HIV positive individuals, women raped in the past 12 months, and homeless individuals. Thus, the responses from person i are denoted $y_{i.} = (y_{i1}, \ldots, y_{iL}, y_{iu})$. The respondents to ARD surveys are also known as the ego and the people to whom the ego can form ties are called alters (Salganik et al., 2011b).

The basic idea behind NSUM relies on the assumption that the proportion of the subpopulation to the general population is equal to the proportion of the person's network that belongs to the subpopulation, i.e.

$$\frac{y_{ik}}{d_i} = \frac{N_k}{N} \tag{1}$$

If the degree d_i was known, then it would be easy to solve for N_k using Equation (1). However, estimating the degree is a difficult problem in and of itself (Rogerson, 1997; Dasgupta et al., 2014). A related topic is that of the "small-world problem," which states that only a small number of connections can connect any two people (Rogerson, 1997). Estimating the average network size is also difficult because it is nearly impossible for someone to recount their entire social network without substantial effort and network size varies dramatically between individuals.

The model in Equation (1) works well under three strong conditions: 1. everyone in

4

the population is equally likely to know someone in subpopulation k, 2. for every person in a respondent's network, the respondent knows every subpopulation the person belongs to, and 3. respondents are able to fully recall everyone in their social network in the allotted time. These assumptions are commonly violated in practice. Barrier effects cause individuals to be more or less likely to know individuals in certain subpopulations, violating condition 1. Transmission errors block individuals from knowing everything about people in their social networks, violating condition 2. Recall errors result when people are unable to quickly count or remember everyone in their social network that belongs to a certain group, violating condition 3 (Bernard et al., 1989; Johnsen et al., 1995; Killworth et al., 1998b; McCarty et al., 2001). We further explore each of these violations in later sections.

There are currently two software packages to analyze ARD using the NSUM, both implemented in R (R Core Team, 2019). The first, NSUM (Maltiel and Baraff, 2015), implements the models proposed by Maltiel et al. (2015). The second, networkreporting (Feehan and Salganik, 2014), fits the generalized network scale-up model proposed by Feehan and Salganik (2016).

During the writing of this review, another review of the network scale-up method was published in The Oxford Handbook of Social Networks (McCormick, 2020). Like the review presented here, the author introduces the NSUM in the context of ARD, delineates the assumptions and pitfalls of the method, and introduces the proposed models to improve the method. While McCormick (2020) walks through the history of the NSUM primarily through the lens of NSUM applications, we instead reflect on NSUM through the modeling perspective. To this end, our review allocates more space to model properties and the complete network models while McCormick (2020) spends more time discussing designing and analyzing an ARD survey. In addition, we apply many of the models discussed in

5

this review to one canonical data set and compare their performance and unique features. Read together, these two reviews provide a near-complete picture of the current NSUM and ARD literature.

The rest of this manuscript is organized as follows. First, we offer a background to ARD and explore the features of the data and problems that arise when collecting ARD. Then, in Section 3, we provide introductions to all significant NSU models. This section is divided into three subsections. In Section 3.1, we discuss the frequentist network scale-up estimators. These basic estimators are the most frequently used in practice due to their ease of use, but also include more complex models with increased flexibility. In Section 3.2, we introduce the Bayesian estimators. These models are reported to have improved the basic methods by better accounting for the sources of bias, but are also more difficult to use since they rely on Bayesian sampling algorithms. In Section 3.3, we review recent estimators that estimate complete network properties using only ARD. After discussing specific models, Section 4 introduces modifications to model estimates. These approaches recognize limitations of the modeling procedures and further calibrate estimates using empirical studies. Final discussion is found in Section 5.

2   Properties of Aggregated Relational Data

Before discussing the various NSUM models, it is important to look at the properties of ARD in more detail. Specifically, there are four key biases that plague ARD: 1) Transmission errors, 2) Barrier effects, 3) Recall errors, and 4) Response biases. As with any survey, ARD can suffer from poor sampling behavior. Recall errors and response biases depend on the survey implementation, while transmission errors and barrier effects depend on the

6

subpopulations and respondents.

2.1 Transmission Error

A response suffers from transmission error when the respondent is unaware that someone in their network belongs to a subpopulation (Killworth et al., 2006). The transmission error violates the assumption that respondents have perfect knowledge of which subpopulations their alters belong to and varies widely between different subpopulations (Killworth et al., 2006; Zheng et al., 2006; Maltiel et al., 2015).

Shelley et al. (1995) studied transmission error by interviewing respondents in key subpopulations. HIV-positive respondents reported that only 49% of their relatives were aware of their HIV-status. However, note that transmission error also exists for easy-to-reach populations. A large percentage of diabetics, twins, Native Americans, and widows and widowers all reported not revealing their status to some members of their social network (Killworth et al., 2006). Focus has primarily been on accounting for the transmission rate of the hard-to-reach subpopulations, although ignoring the transmission rate of the known subpopulation may also significantly influence estimates.

In some cases, researchers can estimate the *transmission rate*, τ, of a subpopulation. Also frequently called the *visibility factor*, τ represents the fraction of a respondent's network that is aware the respondent is in the hidden subpopulation. The methods to estimate the visibility factor include expert opinion, comparison of NSU with proxy respondent method, social respect, coming-out ratio, and game of contacts (Haghdoost et al., 2018). Thus, in addition to collecting the ARD, researchers can also collect additional data to estimate the visibility factor. To the best of our knowledge, there has been no study that compares the accuracy of the different methods to estimate the visibility factor.

2.2 Barrier Effect

Barrier effects violate the constant proportion assumption because respondents can be more or less likely to know someone in a subpopulation due to their own characteristics, and are not limited to hard-to-reach populations. (Shelley et al., 1995; Killworth et al., 2006; Shelley et al., 2006; Salganik et al., 2011a). For example, respondents were more likely to know people with whom they shared their race.

The reasons for barrier effects can be both geographical as well as social. For example, the number of Native Americans that a respondent knows is highly correlated to the state in which they reside. On the other hand, doctors are much more likely to know other doctors, regardless of where they live (Killworth et al., 2006). The influence that barrier effects can have on model estimates can be reduced by obtaining a representative sample of the population of interest. Killworth et al. (2006) cited a study that estimated an unusually high HIV-positive prevalence when compared to national surveys because the original study interviewed only Florida residents, increasing the influence of barrier effects. Unlike transmission error, there does not appear to be any feasible way of estimating barrier effects directly without studying every characteristic of each respondent, since barrier effects depend on both the respondent and the subpopulation considered.

2.3 Recall Error

Recall error occurs when respondents inaccurately recall the number of alters they know in a subpopulation (Killworth et al., 2003, 2006; McCormick et al., 2007; Maltiel et al., 2015). In many studies, respondents have only about 30 seconds to recall everyone in their social network that belongs to a subpopulation (Killworth et al., 2003). Thus, respondents will

8

undercount or overestimate the true number of alters. Little if any research has been done one how many people respondents can recall (Killworth et al., 2003). Furthermore, given that a respondent has recalled i of n possible alters, the probability that they recall another member decreases as i increases. Recall bias may also increase as the survey progressing, meaning later questions will suffer from larger recall error (McCarty et al., 2001).

Researchers found that respondents typically overcounted the number of alters in small subpopulations and undercounted the number of alters in large subpopulations (Killworth et al., 2003; Zheng et al., 2006; McCormick et al., 2007). Killworth et al. (2003) proposed a formula to study the relationship between the estimated degree and the rates at which respondents over and undercount. For large subpopulations, respondents have also been observed to round their responses, typically answering in multiples of 5, but enumerate the individuals in their social network for small subpopulations (McCarty et al., 2001; Killworth et al., 2003). For the large subpopulations, respondents relied on "feel" (McCarty et al., 2001). This means that even for subpopulations with low transmission error, like those based on names, the recall error may be larger than for other subpopulations. Little work has been done to reduce these recall errors when collecting ARD.

2.4   Response Bias

Response bias refers to respondents deliberately misreporting the number of individuals they know in the subpopulation. Respondents may be hesitant to report members of stigmatized subpopulations are in their social network. For example, respondents were hesitant to admit knowing FSW (female sex workers) in household settings (Jing et al., 2018). Thus, response bias can reasonably be reduced at the data-collection stage by making respondents feel comfortable enough to truthfully answer the survey.

Snidero et al. (2009) studied how the survey design and implementation affected response bias. The authors showed that question order likely did not seem to be an important factor in producing reliable estimates of degree and subpopulation size. However, individual interviewers did have a significant effect on whether interviews were interrupted, refused, or contained missing fields. Therefore, in order to reduce variability of the ARD and provide consistent estimates, it is most important to provide the interviewers with sufficient training.

NSUM was also combined with the randomized response technique (RRT) to reduce response bias (Jing et al. (2018)). RRT aims to increase the likelihood that a respondent answers sensitive questions by splitting respondents into two groups and randomly asking the respondents either a sensitive or unrelated survey question and ensuring the respondents that only the respondent knows which question they are answering. Since only the respondents knew which question was asked, the researcher must use only the proportion of the sensitive and unrelated questions to calculate the mean response to the sensitive question. Respondents were much more likely to answer the sensitive question truthfully under RRT, leading to reliable NSUM estimates. (Jing et al., 2018). Note, however, that the RRT provided only the average number of FSW that the respondents knew. Thus, the RRT can only be combined with the scale-up estimates that rely on the average number of individuals known in a subpopulation, rather than the number known for each respondent. Based on the current trend of NSU estimators in Section 3, the RRT is useful only for the most basic estimators.

3 Models

In this section, we will discuss the different ARD models. These models can be loosely categorized into three groups: 1) frequentist models which provide subpopulation size estimates, 2) Bayesian models which handle the biases through the prior distributions, and 3) complete network models which focus on estimating network properties using only ARD. Within each subsection, models will be introduced chronologically. Key theoretical properties of the models will also be discussed, although for brevity, full model properties are left to the original publications.

3.1 Frequentist Models

3.1.1 First NSUM Model

Bernard et al. (1989) first proposed the Network Scale-up Model to study the number of people who had died in the 1985 Mexico City earthquake. The authors derived bounds for the average network size and point estimates for the unknown subpopulation size and Bernard et al. (1991) provided additional empirical results from a larger survey. We focus here on the subpopulation size estimates. While the estimator for subpopulation size is limited in application, it provided a powerful stepping stone for future estimators. The first probability estimator makes no assumption for the distribution of the ARD responses, but notes that the probability of the event that a random respondent knows no one in subpopulation k (denote this event W_k) is given by

$$P(W_k) = \sum_{m=d_{min}}^{d_{max}} P(W_k|y_{ik} = m)P(y_{ik} = m), \tag{2}$$

11

where the degrees can vary over the integers from d_{min} to d_{max}. The authors then assume that either for a random respondent not in the subpopulation k, the respondent's social network is equally likely to have been any subset of size d_i from the general population, or that all subsets of the general population of size N_k were equally likely to be k. Using either of these assumptions and several steps of algebra, the authors show that there exists a real number g, $1 \leq g \leq d_{max}$, such that

$$P(W_k) = \sum_{m=d_{min}}^{d_{max}} (1 - N_k/(N-g))^m P(y_{ik} = m) \approx \sum_{m=d_{min}}^{d_{max}} (1 - N_k/N)^m P(y_{ik} = m), \quad (3)$$

where the approximation holds since g is small with respect to N. Now consider (3) for subpopulations 1 to L, with corresponding values $\epsilon_1 = 1 - N_1/N$ to $\epsilon_L = 1 - N_L/N$. Since $P(W_k)$ is an increasing function with respect to $\epsilon_k = 1 - N_k/N$, if $1 > \epsilon_1 > \epsilon_2 > \cdots > \epsilon_L$, then $1 > P(W_1) > P(W_2) > \cdots > P(W_L)$. Therefore, if $P(W_u)$ for the unknown subpopulation u is such that $P(W_{j-1}) > P(W_u) > P(W_j)$, the subpopulation size is also bounded, where

$$1 - \frac{N_{j-1}}{N} > 1 - \frac{N_k}{N} > 1 - \frac{N_j}{N} \implies \frac{N_{j-1}}{N} < \frac{N_k}{N} < \frac{N_j}{N}. \quad (4)$$

Thus, this procedure provides an upper and lower bound for the size of the unknown subpopulation.

3.1.2 Maximum Likelihood Estimator Models

In order to derive more precise size estimates, Killworth et al. (1998a) proposed a Binomial likelihood based estimator. Overall, six different estimators were proposed to model both personal network size and subpopulation size, but the maximum likelihood based estimator proved the most useful for subpopulation size estimation and was later extended into

12

the most frequently used NSU estimator. The estimator, which we call the *plug-in MLE* (PIMLE), works by first maximizing the following likelihood with respect to d_i:

$$L(d_i; \mathbf{y}) = \prod_{k=1}^{L} \binom{d_i}{y_{ik}} \left(\frac{N_k}{N} \right)^{y_{ik}} \left(1 - \frac{N_k}{N} \right)^{d_i - y_{ik}}.$$

<div align="right">(5)</div>

When N_k are small relative to N and y_{ik} are small relative to d_i, the maximum likelihood estimate of d_i is given by

$$\hat{d}_i = N \cdot \frac{\sum_{k=1}^{L} y_{ik}}{\sum_{k=1}^{L} N_k}.$$

<div align="right">(6)</div>

Plugging in these $\hat{d}_i$ into (1) yields respondent-level subpopulation estimates $\hat{N}_u^{\{i\}} = N \cdot y_{iu}/\hat{d}_i$. The results are then averaged into a single estimate for N_u,

$$\hat{N}_u^{PIMLE} = \frac{1}{n} \sum_{i=1}^{n} N \cdot \frac{y_{iu}}{\hat{d}_i},$$

<div align="right">(7)</div>

where each term inside the summation is the estimated unknown subpopulation size from respondent i. Note that the degree estimate for each respondent depends only on the responses from that respondent and the known subpopulation sizes. Then, each respondent is weighted equally in the final summation. It was shown that $\hat{d}_i$ is unbiased. Furthermore, Monte Carlo simulations showed $1/\hat{d}_i$ is essentially unbiased for $1/d_i$ and the back-estimates for N_u were essentially unbiased when more than 20 subpopulations with known sizes were used for verification (Killworth et al., 1998a). Note that no statement about the bias or standard error of $\hat{N}_u^{PIMLE}$ can be made since the final plug-in does not make any distributional assumptions.

Killworth et al. (1998b) proposed a modified version of the estimate in (7), which has become the most frequently used NSU estimator for unknown subpopulation size. Studies typically refer to this as the *maximum likelihood estimator* (MLE). Instead of

back-estimating N_u using (1), the MLE method instead maximizes the binomial likelihood

$$L(N_k; \mathbf{y}, \{d_i\}) = \prod_{i=1}^{n} \binom{d_i}{y_{iu}} \left(\frac{N_u}{N}\right)^{y_{iu}} \left(1 - \frac{N_u}{N}\right)^{d_i - y_{iu}} \qquad (8)$$

with respect to N_u, where the d_i are fixed at the estimated $\hat{d}_i$ from (6). Thus, the final estimate is still uses the known subpopulation to estimate the degrees, but estimates N_k by maximizing a likelihood instead of solving the equality of two ratios. The new MLE for N_u is then easily found to be

$$\hat{N}_u^{MLE} = N \cdot \frac{\sum_{i=1}^{n} y_{iu}}{\sum_{i=1}^{n} d_i} = \sum_{i=1}^{n} y_{iu} \frac{\sum_{k=1}^{L} N_k}{\sum_{i=1}^{n} \sum_{k=1}^{L} y_{ik}}. \qquad (9)$$

The critical difference in the unknown subpopulation estimate is that the PIMLE averages N_k estimates from each respondent while the MLE maximizes a likelihood using all respondent data simultaneously. The estimate $\hat{N}_u$ is unbiased and assuming the N_u/N is sufficiently small, the authors show the standard error is given by

$$SE(\hat{N}_u^{MLE}) = \sqrt{\frac{N \cdot N_u}{\sum_{i=1}^{n} \hat{d}_i}}, \qquad (10)$$

which decreases as the network sizes d_i increases. When the degrees are small or prevalence is relatively large, the standard error in (10) will be inaccurate.

3.1.3 Weighted Estimators

The MLE estimate of Killworth et al. (1998b) implicitly values information from large known subpopulations more than small known subpopulations. In order to weight the subpopulations equally, the mean of sums (MoS) estimator first estimates the d_i based on each subpopulation, averages those $\hat{d}_i$'s, and then back-estimates N_u based on the $\hat{d}_i$'s

14

(Habecker et al., 2015). Thus, the estimate for network size d_i is given by $\hat{d}_i = (N/L) \cdot \sum_{k=1}^{L} y_{ik}/N_k$, which then yields the back-estimate for $\hat{N}_u$,

$$\hat{N}_u^{MoS} = \frac{N}{n} \cdot \sum_{i=1}^{n} \frac{y_{iu}}{\hat{d}_i}. \tag{11}$$

Note that $\hat{N}_u^{MoS}$ has the same form as $\hat{N}_u^{PIMLE}$, but relies on a different method to calculate the $\hat{d}_i$. The MoS estimator was first proposed in Killworth et al. (1998a), but the authors noted that the variance of the estimate is extremely large when one or more of the known subpopulations are small. Habecker et al. (2015) proposed controlling this variance by choosing subpopulations of similar size.

Habecker et al. (2015) also proposed improving the the MLE and the MoS estimator by incorporating weights to adjust for survey characteristics, like probability of selection. Thus, the weighted MLE (9) and the weighted MoS estimator (11) are given by

$$\hat{N}_u^{WMLE} = N \cdot \frac{\sum_{i=1}^{n} y_{iu} w_i}{\sum_{i=1}^{n} d_i} \tag{12}$$

and

$$\hat{N}_u^{WMoS} = \frac{N}{n} \cdot \sum_{i=1}^{n} (y_{iu} w_i)/\hat{d}_i, \tag{13}$$

respectively.

3.1.4 Generalized Scale-up Estimators

Feehan and Salganik (2016) developed the *generalized scale-up estimator* (GNSUM) to estimate the size of a hard-to-reach subpopulation by using the network property that the total number of in-reports equals the total number of out-reports (i.e. if person i reports they know trait k about person j, then person j reports that person i knows trait k about them). The new estimator also requires additional data collected from the hard-to-reach

15

subpopulations called enriched ARD. Enriched ARD is collected by asking members of the unknown population "How many X's do you know?" and then "How many of these X's are aware that you belong to population G_u?" Clearly one limitation of the generalized network scale-up estimator is that it requires directly sampling from the hard-to-reach subpopulation. However, by leveraging the additional information from the enriched ARD, the generalized scale-up estimator can provide more accurate network and subpopulation size estimates.

First, let the hard-to-reach population G_u of size N_u be a subset of the entire population G. In practice, researchers sample from G_F, which is a separate subset of G called the frame population. Then, let y_{iu} be the number of out-reports from respondent i to subpopulation G_u, i.e. how many people respondent i knows in subpopulation G_u. Furthermore, let $v_{i,G}$ be the number of in-reports to respondent i from the entire population, also known as the *visibility* of person i to people in G. Thus, if we define $y_{G,u} = \sum_{i \in G} y_{iu}$ to be the total number of in-reports and $v_{G,G} = \sum_{i \in G} v_{i,u}$ to be the total number of out-reports, then $y_{G,u} = v_{G,G}$. Multiplying both sides by N_u, we can write $N_u = y_{G,u}/(v_{G,G}/N_u)$. This must also be the case for the frame population, i.e. $N_u = y_{F,u}/(v_{G,F}/N_u)$. Then, the estimate $\hat{N}_u^{GNSUM}$ is found by estimating the numerator and denominator separately, i.e.

$$\hat{N}_u^{GNSUM} = \frac{\hat{y}_{F,u}}{\hat{\bar{v}}_{u,F}}. \tag{14}$$

The numerator is found using the ARD, where $\hat{y}_{F,u} = \sum_{i \in s_F} y_{iu}/\pi_i$, s_F is the sample population, and π_i is the probability that respondent i is included in the sample from the frame population G_F. The enriched ARD is used to find $\hat{\bar{v}}_{u,F}$, but the exact details of the estimator are complicated, and for brevity we leave the description of the estimator to the original manuscript.

Feehan and Salganik (2016) make several important connections between their model

16

and the MLE estimator. First, the authors show that the GNSUM is equal to the MLE times several adjustment factors, and the MLE is only correct when all adjustments factors are equal to 1 while the generalized scale-up is correct regardless. Procedures for estimating the adjustments factors are provided in the original manuscript. This decomposition also leads an expression for the bias of the MLE. This bias expression can be used to adjust MLE estimates if the adjustment factors are known or estimated.

Recently, Verdery et al. (2019) extended the generalized network scale-up estimator to allow venue-based sampling designs instead of the link-tracing samples from the original approach, producing the venue-based generalized scale-up estimator (VB-GNSUM). The generalized NSU estimator is difficult to use with venue-based sampling because estimating $\hat{\bar{v}}_{u,F}$ requires the probability that an individual was included in the sample and because venue-based sampling produces only one sample, rather than the two samples needed for the traditional estimator. Verdery et al. (2019) use the same estimator in Equation 14, but develop new ways to estimate the numerator and denominator.

3.2   Bayesian Models

While the original NSUM literature acknowledged that network sizes varied greatly between individuals and the biases influenced the results, the basic models made it difficult to account for these factors and thus relied on averaging estimates. After the initial development of the NSUM, a wave of Bayesian models dominated the topic, allowing the parameters to vary between individuals and subpopulations and relying on posterior estimates. The Bayesian approach inherently allows us to also consider the joint distribution between N_u and the other model parameters.

3.2.1 The Overdispered Model

Zheng et al. (2006) proposed the first Bayesian model for ARD. Noting that there was large overdispersion in the data, likely a result from barrier effects and varying network sizes, Zheng et al. (2006) proposed the overdispersed model, given by

$$y_{ik} \sim \text{Poisson}(e^{\alpha_i + \beta_k + \gamma_{ik}}). \tag{15}$$

The parameter $a_i = e_i^{\alpha}$ represents the expected degree of respondent i, $b_k = e_k^{\beta}$ represents the proportion of total links that involve subpopulation k, and γ_{ik} allows for extra variability in the model not accounted for by α_i and β_k. Thus, if the γ_{ik} are constant, the number of people that respondent i knows in subpopulation k is dependent only on the number of people that respondent i knows and the relative prevalence of subpopulation k. If, however, γ_{ik} varies widely across the k, then this suggests the presence of one or more of the ARD biases. The flexibility and hierarchical formulation differentiates the Bayesian models from the frequentist models by allowing for more variation in the responses. It is difficult for the frequentist models to accommodate large changes in how likely each respondent is to know someone in each subpopulation. The addition of an overdispersion parameter allows two respondents with equal degrees to have very different responses, which is often seen in the data.

The authors let $g_{ik} = e^{\gamma_{ik}}$ follow a gamma distribution with mean 1 and shape parameter $1/(\omega_k - 1)$, which integrated to the following negative binomial model:

$$y_{ik} \sim \text{negative-binomial} \left(\text{mean} = e^{\alpha_i + \beta_k}, \text{overdispersion} = \omega_k \right), \tag{16}$$

where $E(y_{ik}) = e^{\alpha_i + \beta_k}$ and $Var(y_{ik}) = \omega_k e^{\alpha_i + \beta_k}$.

Note that the α_i's and β_k's are nonidentifiable.Left untouched, this means that an increase in the expected degree of person i is equivalent to a decrease in the proportion of

total links that involve group k. The authors chose to leave the nonidentifiability in the model and instead renormalize the $log(\beta_k)$'s using the rare names (those believed to have the least bias) after running the MCMC chain. Full details of the renormalization process can be found in Zheng et al. (2006).

The main motivation and utility of this paper is the relationship between the subpopulations and the overdispersion, ω_k. Figure 4 of the original manuscript provides a nice visual summary of the overdispersion estimates. For the considered data-set, the model showed that the "homeless" and "member of Jaycees" populations had some of the highest overdispersions while the names had some of the lowest. This means that the number of homeless and Jaycees that a respondent knew varied highly between respondents, while the propensity for respondents to know someone named "Stephanie," for example, was roughly equal. The authors pointed out that homeless populations are both "geographically and socially localized," explaining the largest range of propensities between respondents.

3.2.2 The Latent Profile Models

McCormick et al. (2010) note that the normalization procedure of Zheng et al. (2006) does not ensure the degrees are estimated accurately since the transmission errors and the barrier effects can still bias the degree estimates. Thus, the authors propose introducing latent nonrandom mixing to account for the biases. Unlike previous models, the McCormick et al. (2010) model estimates the propensity for respondents in ego group e to know members of alter group a. In their case-study, the authors chose ego and alter groups by crossing gender and age (e.g. males aged 25-64 comprised one ego group), but the ego and alter groups need not match. The initial model is given by

$$y_{ik} \sim \text{negative-binomial}\,(\text{mean} = \mu_{ike}, \text{overdispersion} = \omega_k') \tag{17}$$

where $\mu_{ike} = d_i \sum_{a=1}^{A} m(e,a) N_{ak}/N_a$ is the mean, N_{ak}/N_a is the proportion of subpopulation k within alter group a, and A is the total number of alter groups. $m(e,a)$ is the mixing coefficient between group e and alter group a, where

$$m(e,a) = \mathrm{E}\left(\frac{d_{ia}}{d_i = \sum_{a=1}^{A} d_{ia}} \middle| i \text{ in ego group } e \right), \tag{18}$$

and d_{ia} is the number of respondent i's alters that belong to subpopulation a. Note that the ego groups and the alter are both exhaustive and mutually exclusive. So for any ego group e, we have $\sum_{a=1}^{A} m(e,a) = 1$. The proposed model in Equations (17)-(18) accounts for the barrier effects but still suffers from recall bias. To account for recall bias, the authors add a calibration curve derived in McCormick et al. (2007) to the mean of their negative binomial model, i.e. replace the previous mean with

$$\mu_{ike} = d_i f \left(\sum_{a=1}^{A} m(e,a) \frac{N_{ak}}{N_a} \right), \tag{19}$$

where the full details of the calibration curve $f(x)$ can be found in the original manuscript.

The authors also provide suggestions for designing future ARD surveys. The Killworth et al. (1998a) degree estimates are equivalent to the degree estimates from the above model in expectation if either (1) there is random mixing; or (2) the known subpopulations represent a *scaled-down* population, i.e.

$$\frac{\sum_{k=1}^{L} N_{ak}}{\sum_{k=1}^{K} N_k} = \frac{N_a}{N}. \tag{20}$$

In words, McCormick et al. (2010) explain that "if 20% of the general population is females under age 30, then 20% of the people with the [subpopulations] used also must be females under age 30." The utility of this requirement is that if the ARD survey is well-designed, then simple models can have the same accuracy as more complicated models. Strategies

for designing such a survey and guidelines for understanding the standard error of the estimates are also provided in the original manuscript.

Note, however, that the McCormick et al. (2010) model does not estimate the size of unknown subpopulations, focusing instead on estimating the individual degree and the distribution of network sizes. In a follow-up paper, McCormick and Zheng (2012) extended the McCormick et al. (2010) model to estimate the unknown subpopulation sizes via MCMC. Via a two-stage estimation procedure, the authors first use the subpopulations where size is known to fit the model in (17) and (19) and then estimate the latent profiles for unknown subpopulation conditional on the estimated values.

3.2.3   Maltiel et al. (2015)

Maltiel et al. (2015) introduced five additional models of increasing complexity to estimate unknown subpopulation sizes from NSUM data and implemented the models in the `NSUM` `R` package (Maltiel and Baraff, 2015; R Core Team, 2019). Figure 1 in Maltiel et al. (2015) contains a helpful flowchart detailing the four basic proposed models. We discuss only the most complex version of the model and direct the reader to the original manuscript for the other models.

In order to address barrier effects, transmission effects, and recall error, Maltiel et al. (2015) proposed the recall adjustment model, given by

$$y_{ik} \sim \text{Binom}\left(d_i, e^{r_k} \tau_k q_{ik}\right), \tag{21}$$

where $d_i \sim \text{log-normal}(\mu, \sigma^2)$, $r_k \sim N(a + b \log(N_k), \sigma_r^2)$, $q_{ik} \sim \text{Beta}(\text{mean} = m_k, \text{dispersion} = \rho_k)$, and $\tau_k \sim \text{Beta}(\text{mean} = \eta_k, \text{dispersion} = \nu_k)$. The parameters r_k handle the recall error, q_{ik} handle the barrier effects, τ_k (fixed at 1 for known subpopulations) handle the transmission effects, and d_i represents the varying degree between respondents. Furthermore, the

21

hyperparameters m_k are set to $m_k = N_k/N$. Full prior settings for the hyperparameters are omitted here for clarity. The advantage of this model is that the parameters are clearly separated into the degree and the bias terms. Thus, the response y_{ik} is influenced both by the number of people that respondent i knows as well as the subpopulation size and the biases. Estimating the transmission effect τ_k for the unknown subpopulation requires some estimate of the visibility factor from additional data sources, for example via the game of contacts.

Ultimately, the authors noted that this complex model was difficult to estimate and proposed removing the r_k parameters and accounting for the recall error post hoc. As others have shown (Killworth et al., 2003; Zheng et al., 2006; McCormick et al., 2007, 2010), respondents seemed to over-report the number of people they knew in small subpopulations and under-report for large subpopulations. The post hoc adjustments are similar to approaches used by others, so details of the adjustments are provided in Section 4.2.

3.2.4 Teo et al. (2019)

In order to account for transmission error and barrier effects, Teo et al. (2019) proposed two new models that include respondent demographics as regression coefficients. Some of the covariates capture overall respondent characteristics, like age and sex. These covariates can be used to capture additional trends in the response. Other covariates measure the perception that each respondent has for each subpopulation. The main idea is if a respondent views a subpopulation poorly, then they are less likely to know individuals from that subpopulation. The transmission error model assumes a Poisson distribution for the ARD given by

$$y_{ik} \sim \text{Poisson}\left(\lambda \alpha_i \exp\{\beta_k[x_{i,k} - U_k]\} N_k\right), \tag{22}$$

while transmission error and barrier effect model is then given by

$$y_{ik} \sim \text{Poisson}\left(\lambda\alpha_i \exp\{\beta_k[x_{i,k} - U_k]\}\exp\left\{\sum_{j=1}^{p}\gamma_{j,k}\mathbf{z}_i\right\}N_k\right),\qquad(23)$$

where $\mathbf{U}$ represents the upper bound of the Likert scale, $x_{i,k}$ is respondent i's response to the question, and $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ represents the column-centered covariate matrix. Like McCormick et al. (2010); McCormick and Zheng (2012), the model relies on respondent characteristics, but treats them more as predictors in a regression framework rather than discretizing the predictors jointly into groups.

3.3   Complete Network Models

Full network data is common in a variety of disciplines. For example, economists can use full social networks to determine whether someone is more likely to save more money when another individual monitors their saving progress (Breza and Chandrasekhar, 2019). Recently, researchers have focused on using ARD as a substitute for full social network studies since ARD is significantly cheaper and easier to collect. ARD has been estimated to be 70-80% cheaper to collect that full network surveys (Breza et al., 2017). The question that arises is if and when ARD can be used in place of full social network data. While this line of research deviates significantly from the predominately population size estimation focus of previous methods, we want to make the reader aware of this new research area. Furthermore, to the best of our knowledge, ARD has not been studied in connection with exponential random graph models.

3.3.1 Latent Surface Model

ARD are ultimately partially observed or sampled network data. Thus, McCormick and Zheng (2015) developed a latent surface model, a popular model for complete network data, to analyze incomplete networks like ARD. The proposed model is based on the "projection model" of Hoff et al. (2002), relaxing the complete graph model to handle the incomplete nature of the ARD. In this framework, the propensity for person i and j to know each other is proportional to the distance between person i and j in the latent geometry. The latent geometry, in their case, is a $p+1$-dimensional hypersphere. Furthermore, for subpopulation G_k, the collected ARD represents $y_{ik} = \sum_{j \in G_k} \delta_{ij}$, where δ_{ij} equals 1 if person i and j know each other and 0 otherwise. Thus, if we denote the latent positions of i and j as z_i and $z_{j \in G_k}$, then the distribution of y_{ik} is approximately Poisson distributed with rate $\lambda_{ik} = \sum_{j \in G_k} P(\delta_{ik} = 1 | \mathbf{Z_i}, \mathbf{Z_j} \in \mathbf{G_k})$. In the complete network case, $j \in G_k$ are observed and known, while for ARD they are unobserved, making it impossible to calculate λ_{ik} directly. Instead, the rate is approximated by

$$\lambda_{ik} \approx N_k \int_{\mathbf{Z_j} \in \mathbf{G_k}} P(\delta_{ik} = 1 | \mathbf{Z_i}, \mathbf{Z_j} \in \mathbf{G_k}) \mathbf{P}(\mathbf{Z_j} \in \mathbf{G_k}) d\mathbf{Z_j} \in \mathbf{G_k}, \tag{24}$$

where N_k is the size of G_k. After computing the expectation of the observed data and reparameterizing the model in terms of d_i, the likelihood of the latent surface model for ARD is found to be

$$y_{ik} | d_i, \beta_k, \zeta, \eta_k, \theta_{(z_i, \nu_k)} \sim \text{Poisson} \left(d_i \beta_i \left(\frac{C_{p+1}(\zeta) C_{p+1}(\eta_k)}{c_{p+1}(0) C_{p+1} \left(\sqrt{\zeta^2 + \eta_k^2 + 2\zeta \eta_k \cos(\theta_{(z_i, \nu_k)})} \right)} \right) \right),$$
$$\tag{25}$$

where $C_{p+1}(\cdot)$ is the normalizing constant of the von-Mises Fisher distribution and $\theta_{(z_i, \nu_k)}$ is the angular distance between respondent i and the center of subpopulation k. The authors propose a Metropolis MCMC algorithm to sample draws from the posterior distribution.

The latent surface model is closely related to the overdispersed model in Zheng et al. (2006), and McCormick and Zheng (2015) offer a more detailed comparison of the two models. As opposed to the overdispersed model, the latent surface model can put two subpopulations on opposite sides of a latent sphere despite the two subpopulations having similar sizes and dispersion. McCormick and Zheng (2015) observed from one data set that individuals who reported knowing more people with AIDS also reported knowing more religious individuals. This ability to view the relationship between subpopulations is the main benefit of the latent surface model, but comes at the cost of increased computation.

3.3.2   Network Statistics

Breza et al. (2017) extended the latent surface models and showed that ARD can be used to estimate node- or graph-level statistics under certain situations and provided insight into when ARD is sufficient. Examples of these statistics are individual centrality and average path length of the graph. The authors showed that they could reproduce the findings of complete network studies using only ARD. As the focus of this paper is remarkably different from our previous discussions, we refer the reader to the original manuscript for model details.

Breza et al. (2019) further developed the theory behind why and when ARD is sufficient to estimate model parameters for complete networks. One key result from their manuscript is that under certain graphs and given a sufficiently large graph, certain parameter estimates from ARD are consistent. The authors further develop a system to identify when ARD is sufficient to recover graph statistics. While there are too many results to include here, ARD proves to be an extremely useful tool to estimate graph statistics, especially considering the cost savings and ease of implementation. If the interest is statistical inference of various

25

population parameters of networks, network sampling methods (Bhattacharyya et al., 2015; Green and Shalizi, 2017; Levin and Levina, 2019; Lin et al., 2020) might be an alternative to ARD and NSUM.

4   Model Calibration

A significant portion of the NSU literature focuses on calibrating the crude estimates from the NSU models in Section 3 through post hoc adjustments. The calibrations scale the model estimates to correct for ARD biases. Calibrations exist for transmission errors and recall errors, but no adjustments exist for barrier effects because of the aforementioned difficulties in estimating the barrier effects. Note, the original adjustments discussed in this section were typically developed for specific models, but the ideas can be often applied to others, so we discuss them in a general setting.

4.1   Transmission Calibration

The two general approaches to account for transmission error are to use known subpopulations that have low transmission error and to correct crude estimates using the visibility factors. Using only low transmission error subpopulations for estimation does not require any additional data sets and can be applied to most NSU studies. The visibility factor, however, requires an additional sample to estimate and can generally only be used for the hidden subpopulation of interest.

Transmission error can be reduced by using only known subpopulations that are unlikely to have transmission error, like names (McCormick et al., 2010). It is unlikely that a respondent reports knowing someone without also knowing their name, thereby remov-

ing bias from the estimates of degree size. However, the approach does not eliminate the transmission error present in the hard-to-reach subpopulation, where the unknown subpopulation size estimates can still be biased.

More sophisticated methods for selecting the known subpopulations have also been proposed. One approach is to back-estimate each known subpopulation using a leave-one-out procedure and remove all subpopulations which are poorly estimated based on the ratio between the back-estimate and the known size (Guo et al., 2013). Similarly, Habecker et al. (2015) proposed trimming one subpopulation at a time by removing only the worst performing subpopulation ratio in the leave-one-out back-estimates. This step-wise trimming addresses the fact that all model estimates will change after removing any subpopulation, so subpopulations that originally had poor performing ratios might actually perform well after removing one other subpopulation.

The most common method to account for transmission bias is to scale the crude subpopulation size estimates by some scaling factor estimated from an additional data set. The visibility factors are used to directly scale the subpopulation size estimates from the NSU procedure by dividing the crude NSU estimate by the visibility factor. For example, if only 50% of a female sex worker's social network is aware that they are a female sex worker, than the FSW subpopulation estimate is divided by 0.5 to account for the transmission bias. Given a estimate of the visibility factor, this approach can be easily applied to any estimator. However, only the unknown subpopulation size estimates are scaled and not the degree estimates. Combining the visibility factor method with the only names approach in McCormick et al. (2010) would likely lead to better degree and subpopulation size estimates.

4.2 Calibration Curve

While Shelley et al. (1995), Killworth et al. (2003), and Shelley et al. (2006) investigated recall bias in ARD, accounting for the recall bias in the models proved difficult. However, several approaches have been proposed, none of which require any additional data sets. McCormick et al. (2007) were the first to propose a method to account for recall bias. In order to account for models overestimating small subpopulations and underestimating large subpopulations, the authors constructed a "calibration curve" which the authors believed to match the relationship between subpopulation size and recall bias. The calibration curve attempts to scale the *recalled* number of alters to be equal to the true number of alters in subpopulation k. As defined in McCormick et al. (2007), let e^{β_k} be the proportion of ties in the social network for subpopulation k, $e^{\beta'_k}$ be for the recalled social network, and then define $\beta'_k = f(\beta_k)$ to be the calibration curve. The calibration curve was first defined as

$$f(\beta_k) = b + \frac{1}{2}(\beta_k - b) + \frac{1}{2a}\left(1 - e^{-a(\beta_k - b)}\right), \tag{26}$$

where a controls how fast the derivative of this curve approaches $1/2$ and b controls at what value of β_k the curve changes from correcting for over-reporting to under-reporting. The parameters a and b are then estimated using the subpopulation size estimates without any bias correction.

As mentioned in Section 3.2.3, Maltiel et al. (2015) proposed a similar calibration curve. Their calibration works by first treating each known subpopulation as unknown one at a time and estimating the size of that population, $\hat{N}_k$. Then, the errors-in-variables model $\log\left(\hat{N}_k\right) = a + b\log(N_k) + \delta_k + \epsilon_k$ is fit via maximum likelihood to estimate a, b, and the variances of δ_k and ϵ_k, where $\delta_k \sim N(0, s_k^2)$ and $\epsilon_k \sim N(0, \sigma_\epsilon^2)$. Estimates of $\log(N_k)$ from the posterior are then transformed using the estimated a, b, and an additional random

28

Table 1: Brief summary of NSUM models.

Model	Primary Objective	Requires External Data	Requires Respondent Covariates	Distributional Assumptions	Bayesian
Bernard et al. (1989) and **Bernard et al. (1991)**	Degree Estimation			N/A	
Killworth et al. (1998a) and **Killworth et al. (1998b)**	Size Estimation			Binomial	
Zheng et al. (2006)	Overdispersion/social structure			Negative Binomial	
McCormick et al. (2010)	Degree Estimation		✓	Negative Binomial	✓
McCormick and Zheng (2012)	Size Estimation		✓	Negative Binomial	✓
Habecker et al. (2015)	Size Estimation			Binomial	
Maltiel et al. (2015)	Size Estimation	✓		Binomial	✓
Feehan and Salganik (2016)	Size Estimation	✓		N/A	
Verdery et al. (2019)	Size Estimation	✓		N/A	
Teo et al. (2019)	Size Estimation		✓	Poisson	✓
McCormick and Zheng (2015)	Complete Network Statistics			Poisson	✓
Breza et al. (2017)	Complete Network Statistics			Poisson	✓
Breza et al. (2019)	Complete Network Statistics			N/A	

noise. The transformation is given by

$$\frac{Y_k^{[t]} - a}{b} + Z, \tag{27}$$

where $Y_k^{[t]}$ is the t-th MCMC sample from the posterior distribution of $\log(N_K)$ and $Z \sim N(0, \sigma_\epsilon^2/b^2)$.

5 Discussion

In this manuscript, we discussed the properties of ARD, explored the wide range of models for ARD, and summarized common modifications to these model estimates. For a succinct summary, Table 1 lists these models, the primary objectives, and additional modeling properties. Many of the models have multiple objectives and estimate several properties,

so we record what appears to be the primary objective of the original manuscript. ARD is an increasingly popular survey type for estimating the size of unknown subpopulations due to its relatively cheap method of collecting network data and that individuals from hard-to-reach populations do not need to be surveyed. There are several biases that make modeling ARD difficult, but extensive research has been performed to improve the accuracy and precision of ARD. Recently, ARD has been used in place of full social networks to estimate network properties, and this is a promising area of research that deserved more attention. We include an extensive list of implemented and some proposed NSU studies in the Supplementary Material. The additional abbreviations we use for the target subpopulations are MCFSW (male client of a female sex worker), MSM (men who have sex with men), RWOS (relationship with opposite sex), EPMS (extra/pre-marital sex), and PED (performance enhancing drugs).

There are a variety of popular methods for estimating key populations, including census/enumeration, capture-recapture, multiplier, and respondent-driven sampling. Population size estimates from these methods and NSUM can vary widely, and in a study of heavy drug users in Curitiba, Brazil, Salganik et al. (2011a) found the NSUM and the generalized NSUM estimates were multiple times larger than the size estimates from comparable direct and multiplier methods. While several reasons for the discrepancy were discussed, there was no clear conclusion.

Respondent-driven sampling and NSUM are related through their reliance on respondents' social networks, albeit it this connection takes a different form in each method. Respondent-driven sampling relies on the social networks of members of the target population and specifically how connections the respondent knows that are also in the target population. Thus, future responses depend highly on each respondent. This procedure

makes it easy to sample directly from the target population, but may come at a cost. For respondent-driven sampling, the standard error of the population size estimate may converge at a much slower rate than expected as the number of participants increases Rohe et al. (2019). Meanwhile, the NSUM relies on social networks of individuals from the general population and each respondent is assumed to be independent of the others. These factors suggest that the NSUM does not suffer from the same problems of convergence.

There are many desirable features of the more traditional estimation techniques, but like NSUM, they all rely heavily on data assumptions. While multiple studies have compared estimates from the more traditional estimation techniques and NSUM, there is still not a clear consensus that any model out-performs the other. The main advantage of the NSUM is respondents do not need to be from the target population. This is beneficial both because it is often difficult simply to survey members from the target population, but also because ARD surveys are cheap and can be implemented quickly. Furthermore, the NSUM can easily estimate the size of impossible-to-reach population like those who have died in an earthquake and estimate the size of multiple unknown subpopulations from the same ARD survey with little extra cost. However, these features do not necessarily lead to more accurate results, and estimates depend significantly on the design of the ARD survey as well as the details of the implementation, like question order. Furthermore, NSUM relies more heavily both on the truthfulness and accuracy of the respondent answers than some of the other methods. Despite these limitations, NSUM models have developed to handle many of the biases present in the data and additional information can be included in the models, making NSUM an attractive and trustworthy method.

The models for ARD range from very simple MLE estimators to more complex latent surface approaches. Each method has different purposes and assumptions in place, so

31

we refrain from recommending any specific model. Furthermore, the performance of each model has been shown to depend heavily on the sample and place of study. The practical challenge of implementing these methods is also of importance. The frequentist estimators are much easier to implement and have remained the most popular approaches. However, the more complex models offer potential gains in accuracy as well as potentially additional insight into the subpopulations. For example, at the cost of additional computation, the latent surface model proposed in McCormick and Zheng (2015) can place two subpopulations with similar dispersion parameters on opposite sides of the latent surface, indicating that the subpopulations are socially distinct, while the Zheng et al. (2006) model estimates only the dispersion parameters. This is not to say that the latent surface model outperforms the overdispersed model, but rather that each model has a different niche. We implement many of the models on a canonical ARD data set and present the results in the Supplementary Material. In there, we compare the performance of various methods and comment on their unique features.

ARD, and the extension of NSUM, is a promising area of research, and the models can be extended beyond size estimation. One interesting approach already considered was to invert the problem and estimate the ARD from known subpopulation sizes and degrees. Moody (2005) estimated how many people know someone affected by the United States war on terror. Furthermore, there are many hard-to-reach populations that can benefit from the NSUM. Shelton (2015) proposed using the NSUM to estimate the prevalence of sex trafficked individuals, noting how difficult it is to obtain a direct estimate. Another promising extension of ARD is to consider other forms of "How many X's do you know?" questions. The generalized NSUM estimator requires enriched ARD. Asking additional questions to the respondents may yield helpful information that can be used in a new esti-

mator. Finally, Breza et al. (2017) suggest that ARD should move beyond social networks. For example, the authors consider the question "How many links does the firm have to firms with trait k?" This question takes a step away from social networks and generalizes ARD to weighted and directed graphs.

The data and models still have limitations, of course. Estimates for known subpopulation sizes are often significantly incorrect. Approaches for estimating the transmission error are often expensive or impossible for certain subpopulations. Many approaches have reduced response bias, but the success of these methods depends heavily on the unknown subpopulation as well as where the study is implemented. Some papers report eliminating the recall error in the data, but it is difficult to validate these approaches since little research has been done on just recall error. Adjustments are based on leave-one-out estimates of known subpopulations, but the discrepancy between these estimates and the truth are not shown to be due to recall error.

With enough improvement, the NSUM will hopefully drop its "under-development" label and be a useful and accurate method for estimating hard-to-reach populations. The models have been used in a large number of real-world studies and have offered promising results in the field of size estimation. The simplicity of the method is attractive and we hope this review inspires new and exciting developments in the field.

Acknowledgment

Supplementary Material to "Thirty Years of The Network Scale-up Method"

A   Applied Study

In this section, we apply many of the NSUM models listed in Table 1 of the main manuscript to a standard aggregated relational data (ARD) data set. For this study, we consider the Curitiba, Brazil dataset, originally studied in Salganik et al. (2011a), which contains responses from 500 residents of Curitiba about 20 known subpopulations and the one unknown subpopulation of heavy drug users. The 20 known subpopulations and their respective known sizes are shown in Table 2.

For readability and conciseness, we apply models from Killworth et al. (1998b) (Killworth MLE), Zheng et al. (2006) (Zheng Overdispersed), Maltiel et al. (2015) (Maltiel Random Degree and Maltiel Barrier Effects), Habecker et al. (2015) (Habecker MoS), and Teo et al. (2019) (Teo Basic and Teo Barrier Effects) and report additional estimates and conclusions from Maltiel et al. (2015) (Maltiel Transmission Error and Maltiel Combined) and Feehan and Salganik (2016) (Feehan Generalized). The estimates from the Feehan and Salganik (2016) method are reported in Salganik et al. (2011a) based on an earlier technical report of their manuscript. We exclude models from before Killworth et al. (1998b) since Killworth et al. (1998b) is overwhelmingly the most popular method and considerably improved the earlier estimators. We also exclude McCormick et al. (2010) and McCormick and Zheng (2012) because the methods require additional data to construct the relative sizes needed in the modeling. The Curitiba data set does not come from venue sampling, so Verdery et al. (2019) is not applicable. Finally, we exclude the complete network statistic

34

models (McCormick and Zheng, 2015; Breza et al., 2017, 2019) as these methods move beyond the NSUM method.

We also briefly describe some of the modeling choices for our implementations. For the Zheng et al. (2006) model, to offer a more equal comparison to the other models and because there are no subpopulations corresponding to names in the Curitiba data, we let $C = C_1$ and P_{G_1} be the proportion of people in all known subpopulations (page 414 of Zheng et al. (2006)). In practice, a different set of reference populations could be used for scaling, and we've found that the choice of reference populations can have a significant effect on estimates.

For the Habecker et al. (2015) models, we implement only the unweighted mean of sums (MoS) estimator, since there is no natural choice of weights in the Curitiba data set. Again to offer a more equal comparison, we do not perform the suggested recursive trimming of populations.

For the Teo et al. (2019) models, we implement a modified version of their model. The Curitiba data set does not contain a measure of each respondent's perception of each population, which the authors denote $x_{i,l}$. However, the data set does include demographic information. Thus, we fit the basic model with no covariates, as well as a barrier effect model with no transmission error. In our case and adopting the authors notation, the resulting barrier effect model looks like

$$N_{i,l}^H \sim Po((\lambda \alpha_i \exp\{\beta_l\} \exp\{\gamma_{age,l} \left[z_i^{age} - \bar{z}^{age}\right] + \gamma_{born,l} \left[z_i^{born} - \bar{z}^{born}\right]$$
$$+ \gamma_{employed,l} \left[z_i^{employed} - \bar{z}^{employed}\right] + \gamma_{gender,l} \left[z_i^{gender} - \bar{z}^{gender}\right] S_l^H\})),$$

where z_i^{age} is the age of respondent i, and z_i^{born}, $z_i^{employed}$, and z_i^{gender} are indicator variable with values 1 if respondent i is born in Curitiba, employed, or male.

Below, we compare the distribution of the estimated respondent degrees and the estimated size of the unknown subpopulation. Additionally, we briefly discuss the distinguishing features of models, where available. We note that this is not an exhaustive study of the models nor is it a complete analysis of the Curitiba data set. However, this serves as an introduction to applying the various models to a canonical ARD data set and a comparison of how each model relates to one another.

A.1   Degree Estimates

All models we consider estimate the respondent degrees, either as a first step to estimate population sizes or simultaneously alongside other model parameters. Figure (1) plots smoothed densities of the estimated degrees, upper-truncated at 1000 for visibility. The degrees corresponding to Bayesian models are calculated using the posterior mean of the samples. For the most part, the models produce very similar degree estimates (e.g. Killworth MLE, Maltiel Random Degree, Teo Basic, and Teo Barrier Effects), although there are some differences between the distributions of estimated degrees. The Zheng Overdispersed and Maltiel Barrier Effects models both have peaks corresponding to higher degrees than the other models. The most noticeable difference is the large variability of the Habecker MoS degree estimates, with one degree estimate even larger than 11000. The next largest estimate from another model is less than 3500, produced from the Maltiel Barrier Effects model. The variability from the Habecker MoS model is expected, as Killworth et al. (1998a) advised against the model because the degree estimates have "unacceptably high variance" due to the influence that small subpopulations have on the estimates. This finding motivates the use of the recursive trimming, as suggested in Habecker et al. (2015), which would remove the small subpopulations, leading to more trustworthy results.
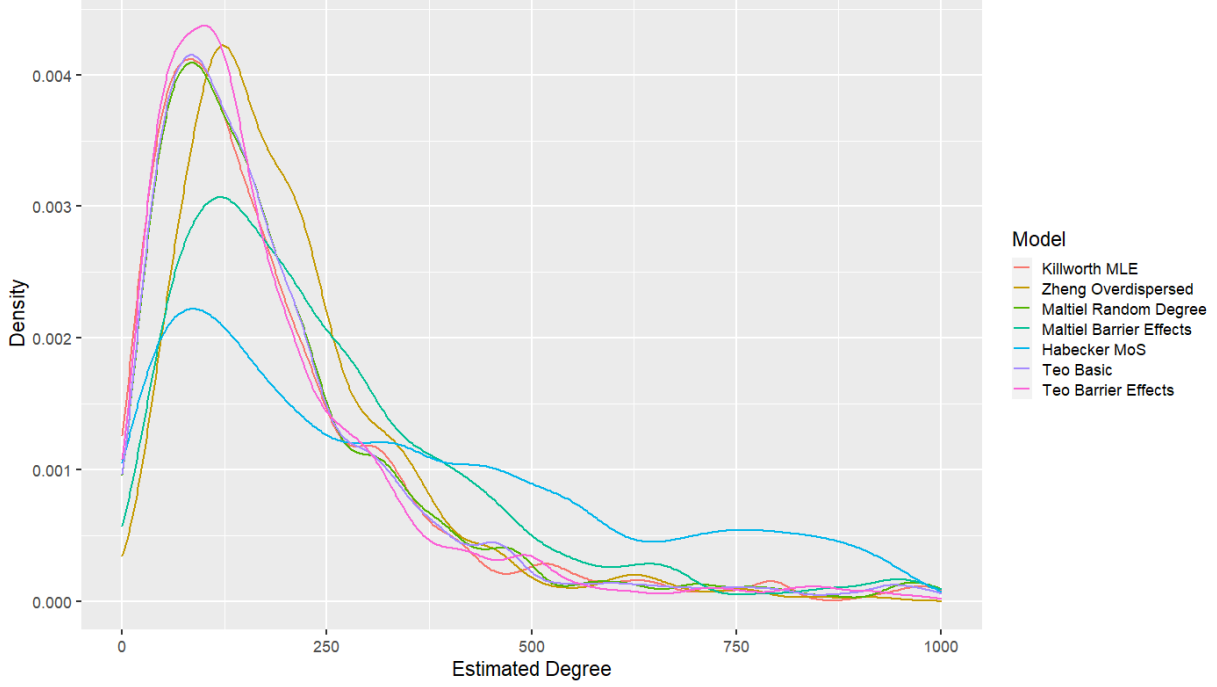
Figure 1: Smoothed density curves of estimated degrees of Curitiba, Brazil ARD respondents. The x-axis was upper-truncated at 1000 for visibility.

A.2 Unknown Population Size Estimates

We also present the heavy drug user prevalence estimates and their corresponding 95% confidence/credible intervals in Figure 2. Both the point estimate and the uncertainty intervals vary significantly. Many methods estimate very similar point estimates to the Killworth MLE model (Zheng Overdispered, Maltiel Random Degree, Teo Basic, and Teo Barrier Effects). This result is expected, since the overall structure of the models are very similar (Maltiel Random Degree can be thought of as the Bayesian version of the Killworth MLE model). The Zheng Overdispersed model simply accounts for extra variability in the

37

model through the overdispersion parameter. It is thus surprising that the Maltiel Barrier Effects model does not also have a similar point estimate, since the model also includes an extra error term to account for the extra variability. It's possible that the slight differences in the assumed distributions of the error terms leads to significantly different results. Lastly, we note that the Teo Barrier Effects model estimates a slightly larger prevalence than the Teo Basic model, but produces a similarly sized uncertainty interval. Our analysis here does not shed light on whether the demographic information offers more reliable estimates in any aspect, but does show that estimates can be influenced by information about the respondents.

Because of the inclusion of a transmission error term, the Feehan Generalized, Maltiel Transmission, and Maltiel Combined estimates are all significantly higher than any of the models we applied. From the game of contacts, Salganik et al. (2011a) estimated the transmission rate of heavy drug users is 0.77, i.e. "there is about a 75% chance that an alter connected to a heavy drug user in Curitiba will be aware that the given ego is heavy drug user". Combined with an estimated popularity factor of 0.69, the generalized NSUM multiplies the Killworth MLE estimate by a factor of 1.88, almost doubling the estimated prevalence. The transmission error has a similar effect on the Maltiel transmission and combined models. Note that the Maltiel Transmission and Maltiel Combined point estimate and uncertainty intervals are approximated from Fig. 5. of Maltiel et al. (2015) and may differ from the true values by a small amount.

Regarding the uncertainty intervals, it's clear that models without any overdispersion estimate (Killworth MLE, Maltiel Random Degree, Teo Basic, and Teo Barrier Effects) have much smaller uncertainty intervals since these models do not account for any extra variability. It's unclear from this study whether the larger uncertainty intervals are desirable,
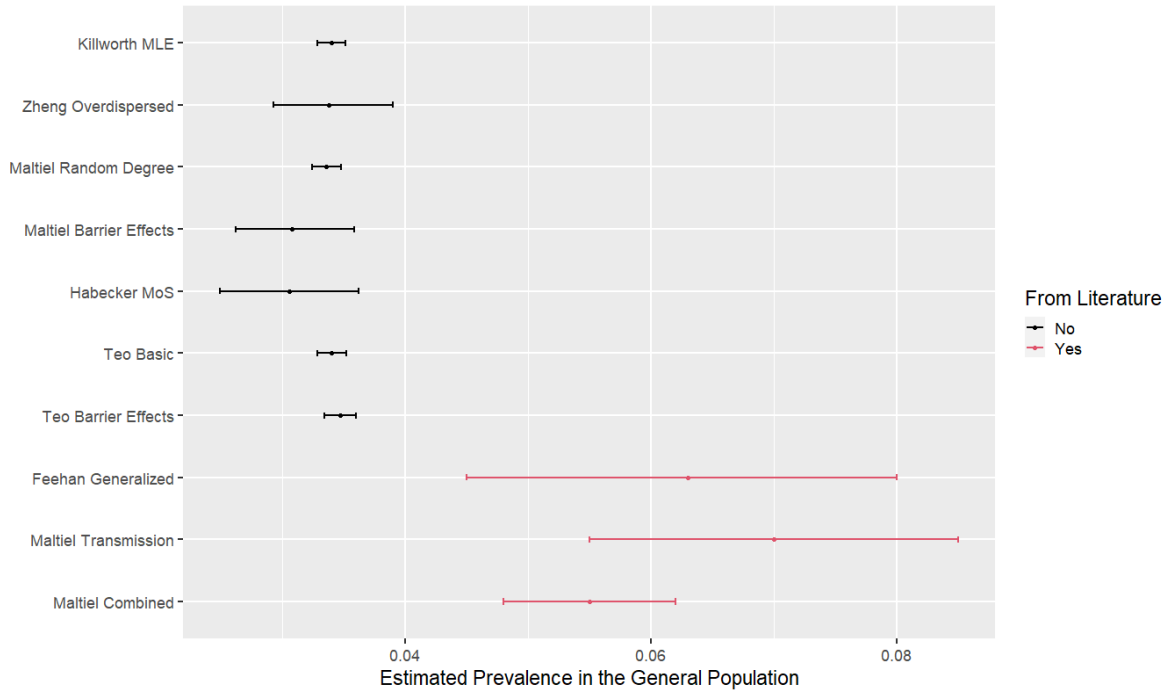
Figure 2: Estimated prevalences of heavy drug users in Curitiba, Brazil from the applied methods. The black error bars correspond to estimates we generated while the remaining error bars are reported in their respective manuscripts.

although the inclusion of extra variability makes sense given the biases present in ARD. It's also potentially noteworthy that the only model considered which includes respondent covariate information (Teo Barrier Effects) estimates significantly higher prevalence. As discussed earlier, the structure of the Habecker MoS estimator produces significantly larger credible intervals than the Killworth MLE model, although this may be more desirable with respect to the prevalence estimates than for the degree estimates.

A.3   Unique Inferences

While all methods considered estimate respondent degrees and unknown subpopulation sizes, many models also provide insight into unique aspects of the ARD. Zheng et al. (2006) produces overdispersion estimates for each subpopulation, measuring the variation in propensities for the respondents to form ties in each subpopulation. In the Curitiba data set, the subpopulations with the largest variation are heavy drug users and public university students while the lowest is women recently married. These results are logical since while most people are roughly equally likely to know a women who was recently married, certain populations are far more likely to know heavy drug users and public university students than others. Like our findings for heavy drug users, Zheng et al. (2006) found that other marginalized populations like including HIV/AIDS patients and homeless persons also have larger overdispersions.

Similarly, the Teo et al. (2019) models provide unique inference into how respondent information relates to how many people they report knowing from each subpopulation. While the Curitiba data set does not contain any information respondent perception of the subpopulations, the barrier effects model we fit provides information about how respondent age, employment status, gender, and birth location affects the likelihood of knowing someone from each subpopulation. For example, in our analysis, model estimates suggest that older respondents report knowing fewer boys and girls under five but more men and women over seventy. Furthermore, employed respondents are more likely to know bank tellers than those who are unemployed.

A.4 Discussion

While this is not meant to be an exhaustive comparison of all existing NSUM models, we apply a variety of NSUM models to a canonical ARD data set to demonstrate how each model compares in practice. We have shown that small differences in modeling assumptions can produce substantial differences with respect to both inference on shared parameters (degree and unknown subpopulation sizes) as well as to new insights through unique parameters. While it is difficult to compare the accuracy of estimates, each model can provide information about important quantities of inference, including overdispersion (Zheng et al., 2006), transmission effects (Salganik et al., 2011a; Maltiel et al., 2015; Teo et al., 2019), and respondent demographics (Teo et al., 2019). All R code created for this study along with a simulated Curitiba data set can be found at `github.com/ilaga/Thirty_Years_NSUM`.

Table 2: Curitiba, Brazil subpopulations and sizes

Subpopulation	Known Size
New mother under 20	4,088,438
New mother 20 and over	935,153
Man recently married	1,328,606
Woman recently married	3,966,956
Hospitalized for recent traffic accident	899,443
Died	2,993,846
Public middle school student	1,869,098
Private middle school student	258,619
Private high school student	150,989
Public university student	144,130
Taxi driver	104,186
Bus driver	47,587
Bank teller	278,195
Construction worker	222,884
On disability	762,877
City employee	323,863
Girl under five	108,511
Boy under five	178,364
Woman over seventy	472,657
Man over seventy	69,471

B Table of Network Scale-up Method Studies

Table 3: A survey of NSUM applied studies. Country and year of study are recorded when provided. When the year was not provided, the publication year is denoted by *.

Subpopulation	Country, year	References
FSW	Rwanda, 2011	Rwanda Biomedical Center (2012)
	China, 2011	Guo et al. (2013)
	China, 2012	Jing et al. (2018)
	Iran, 2014*	Maghsoudi et al. (2014)
	Iran, 2014	Sharifi et al. (2017)
	Iran, 2014*	JafariKhounigh et al. (2014)
	Singapore, 2017	Teo et al. (2019)
Drug Users	United States, 1997	Kadushin et al. (2006)
	Brazil, 2009	Salganik et al. (2011b)
	Brazil, 2009-2010	Salganik et al. (2011a)
	Rwanda, 2011	Rwanda Biomedical Center (2012)
	China, 2011	Guo et al. (2013)
	Iran, 2012*	Shokoohi et al. (2012)
	Iran, 2014*	Maghsoudi et al. (2014)
	Iran, 2016	Nasiri et al. (2019)
	Iran, 2013	Nikfarjam et al. (2016)
	Iran, 2012-2013	Kazemzadeh et al. (2016)

Table 3: (continued)

	Iran, 2016*	Sheikhzadeh et al. (2016)
	Iran, 2016	Zahedi et al. (2018)
	Iran, 2014*	JafariKhounigh et al. (2014)
	Iran, 2015	Sajjadi et al. (2018)
	Iran, 2016-2017	Narouee et al. (2019)
	Thailand, 2014	Kanato (2015)
	Georgia, 2014	BPU and CIF (2015)
	Singapore, 2017	Teo et al. (2019)
MCFSW	Rwanda, 2011	Rwanda Biomedical Center (2012)
	China, 2011	Guo et al. (2013)
	Iran, 2014*	JafariKhounigh et al. (2014)
	Iran, 2015	Sajjadi et al. (2018)
	Singapore, 2017	Teo et al. (2019)
MSM	Japan, 2009	Ezoe et al. (2012)
	Rwanda, 2011	Rwanda Biomedical Center (2012)
	China, 2011	Guo et al. (2013)
	China, 2012	Jing et al. (2014)
	China, 2012	Wang et al. (2015)
	Georgia, 2014	Sulaberidze et al. (2016)
	Iran, 2014*	JafariKhounigh et al. (2014)
	Iran, 2015	Sajjadi et al. (2018)
	Singapore, 2017	Teo et al. (2019)

44

Abortion	Iran, 2012	Rastegari et al. (2014)
	Iran, 2015	Zamanian et al. (2016)
	Iran, 2016	Zamanian et al. (2019)
Alcohol	Iran, 2012	Nikfarjam et al. (2017)
	Iran, 2012-2013	Kazemzadeh et al. (2016)
	Iran, 2014*	JafariKhounigh et al. (2014)
	Iran, 2015	Sajjadi et al. (2018)
	Iran, 2016*	Sheikhzadeh et al. (2016)
Cancer	Iran, 2012-2013	Haghdoost et al. (2015)
	Iran, 2014	Vardanjani et al. (2015)
Choking	Italy, 2004	Snidero et al. (2007)
	Italy, 2004	Snidero et al. (2012)
Social Network Size	Iran, 2010*	Shokoohi et al. (2010)
	Iran, 2012	Shati et al. (2014)
	Iran, 2013*	Rastegari et al. (2013)
Disabilities	Iran, 2012	Mohebbi et al. (2014)
Religious Groups	United States, 2016	Yang and Yang (2017)
Seroprevalence	United States, 1994	Killworth et al. (1998b)
Rape	United States, 1994	Killworth et al. (1998b)
Homelessness	United States, 1994	Killworth et al. (1998b)
RWOS	Iran, 2012-2013	Kazemzadeh et al. (2016)
EPMS	Iran, 2012-2013	Kazemzadeh et al. (2016)

	Iran, 2014*	JafariKhounigh et al. (2014)
	Iran, 2015	Sajjadi et al. (2018)
	Iran, 2016*	Sheikhzadeh et al. (2016)
PED	UK and Southern Ireland, 2013*	James et al. (2013)
Sex Trafficked	United States	Shelton (2015)
Suicide	Iran, 2014	Daneshi et al. (2014)

References

Bernard, H. R., Hallett, T., Iovita, A., Johnsen, E. C., Lyerla, R., McCarty, C., Mahy, M., Salganik, M. J., Saliuk, T., Scutelniciuc, O., et al. (2010). Counting hard-to-count populations: the network scale-up method for public health. *Sexually transmitted infections*, 86(Suppl 2):ii11–ii15.

Bernard, H. R., Johnsen, E. C., Killworth, P. D., and Robinson, S. (1989). Estimating the size of an average personal network and of an event subpopulation. In *The small world*, pages 159–175. Ablex Press.

Bernard, H. R., Johnsen, E. C., Killworth, P. D., and Robinson, S. (1991). Estimating the size of an average personal network and of an event subpopulation: Some empirical results. *Social science research*, 20(2):109–121.

Bhattacharyya, S., Bickel, P. J., et al. (2015). Subsampling bootstrap of count features of networks. *Annals of Statistics*, 43(6):2384–2411.

BPU, B. and CIF, C. (2015). Population size estimation of people who inject drugs in georgia 2016. *Tbilisi: CIF*, pages 1–51.

Breza, E. and Chandrasekhar, A. G. (2019). Social networks, reputation, and commitment: evidence from a savings monitors experiment. *Econometrica*, 87(1):175–216.

Breza, E., Chandrasekhar, A. G., McCormick, T. H., and Pan, M. (2017). Using aggregated relational data to feasibly identify network structure without network data. Technical report, National Bureau of Economic Research.

Breza, E., Chandrasekhar, A. G., McCormick, T. H., and Pan, M. (2019). Consistently estimating graph statistics using aggregated relational data. *arXiv preprint arXiv:1908.09881*.

Daneshi, S., Haghdoost, A., Baneshi, M., and Zolala, F. (2014). The estimated frequency of spinal cord injury, amputation (hands and feet) and death in the bam earthquake using the network scale up method. *Iranian Journal of Epidemiology*, 10(3):9–14.

Dasgupta, A., Kumar, R., and Sarlos, T. (2014). On estimating the average degree. In *Proceedings of the 23rd international conference on World wide web*, pages 795–806.

Ezoe, S., Morooka, T., Noda, T., Sabin, M. L., and Koike, S. (2012). Population size estimation of men who have sex with men through the network scale-up method in japan. *PloS one*, 7(1):e31184.

Feehan, D. and Salganik, M. (2014). *The networkreporting package.*

Feehan, D. M. and Salganik, M. J. (2016). Generalizing the network scale-up method: a new estimator for the size of hidden populations. *Sociological methodology*, 46(1):153–186.

Green, A. and Shalizi, C. R. (2017). Bootstrapping exchangeable random graphs. *arXiv preprint arXiv:1711.00813*.

Guo, W., Bao, S., Lin, W., Wu, G., Zhang, W., Hladik, W., Abdul-Quader, A., Bulterys, M., Fuller, S., and Wang, L. (2013). Estimating the size of hiv key affected populations in chongqing, china, using the network scale-up method. *PloS one*, 8(8):e71796.

Habecker, P., Dombrowski, K., and Khan, B. (2015). Improving the network scale-up estimator: Incorporating means of sums, recursive back estimation, and sampling weights. *PloS one*, 10(12).

Haghdoost, A., Gohari, M. A., Mirzazadeh, A., Zolala, F., and Baneshi, M. R. (2018). A review of methods to estimate the visibility factor for bias correction in network scale-up studies. *Epidemiology and health*, 40.

Haghdoost, A. A., Baneshi, M. R., Haji-Maghsoodi, S., Molavi-Vardanjani, H., and Mohebbi, E. (2015). Application of a network scale-up method to estimate the size of population of breast, ovarian/cervical, prostate and bladder cancers. *Asian Pac J Cancer Prev*, 16(8):3273–7.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098.

JafariKhounigh, A., Haghdoost, A. A., SalariLak, S., Zeinalzadeh, A. H., Yousefi-Farkhad, R., Mohammadzadeh, M., and Holakouie-Naieni, K. (2014). Size estimation of most-at-risk groups of hiv/aids using network scale-up in tabriz, iran. *Journal of Clinical Research & Governance*, 3(1):21–26.

James, R. A., Nepusz, T., Naughton, D. P., and Petróczi, A. (2013). A potential inflating effect in estimation models: Cautionary evidence from comparing performance enhancing drug and herbal hormonal supplement use estimates. *Psychology of Sport and Exercise*, 14(1):84–96.

Jing, L., Lu, Q., Cui, Y., Yu, H., and Wang, T. (2018). Combining the randomized response technique and the network scale-up method to estimate the female sex worker population size: an exploratory study. *Public health*, 160:81–86.

Jing, L., Qu, C., Yu, H., Wang, T., and Cui, Y. (2014). Estimating the sizes of populations at high risk for hiv: a comparison study. *PloS one*, 9(4).

Johnsen, E. C., Bernard, H. R., Killworth, P. D., Shelley, G. A., and McCarty, C. (1995). A social network approach to corroborating the number of aids/hiv+ victims in the us. *Social networks*, 17(3-4):167–187.

Kadushin, C., Killworth, P. D., Bernard, H. R., and Beveridge, A. A. (2006). Scale-up methods as applied to estimates of heroin use. *Journal of Drug Issues*, 36(2):417–440.

Kanato, M. (2015). Size estimation of injecting drug users through the network scale-up method in thailand. *Journal of the Medical Association of Thailand= Chotmaihet thangphaet*, 98:S17–24.

Kazemzadeh, Y., Shokoohi, M., Baneshi, M. R., and Haghdoost, A. A. (2016). The frequency of high-risk behaviors among iranian college students using indirect methods: network scale-up and crosswise model. *International journal of high risk behaviors & addiction*, 5(3).

Killworth, P. D., Johnsen, E. C., McCarty, C., Shelley, G. A., and Bernard, H. R. (1998a). A social network approach to estimating seroprevalence in the united states. *Social networks*, 20(1):23–50.

Killworth, P. D., McCarty, C., Bernard, H. R., Johnsen, E. C., Domini, J., and Shelley, G. A. (2003). Two interpretations of reports of knowledge of subpopulation sizes. *Social networks*, 25(2):141–160.

Killworth, P. D., McCarty, C., Bernard, H. R., Shelley, G. A., and Johnsen, E. C. (1998b). Estimation of seroprevalence, rape, and homelessness in the united states using a social network approach. *Evaluation review*, 22(2):289–308.

Killworth, P. D., McCarty, C., Johnsen, E. C., Bernard, H. R., and Shelley, G. A. (2006). Investigating the variation of personal network size under unknown error conditions. *Sociological Methods & Research*, 35(1):84–112.

Levin, K. and Levina, E. (2019). Bootstrapping networks with latent space structure. *arXiv preprint arXiv:1907.10821*.

Lin, Q., Lunde, R., and Sarkar, P. (2020). On the theoretical properties of the network jackknife. In *International Conference on Machine Learning*, pages 6105–6115. PMLR.

Maghsoudi, A., Baneshi, M. R., Neydavoodi, M., and Haghdoost, A. (2014). Network scale-up correction factors for population size estimation of people who inject drugs and female sex workers in iran. *PloS one*, 9(11):e110917.

Maltiel, R. and Baraff, A. J. (2015). *NSUM: Network Scale Up Method*. R package version 1.0.

Maltiel, R., Raftery, A. E., McCormick, T. H., and Baraff, A. J. (2015). Estimating population size using the network scale up method. *The annals of applied statistics*, 9(3):1247.

McCarty, C., Killworth, P. D., Bernard, H. R., Johnsen, E. C., and Shelley, G. A. (2001). Comparing two methods for estimating network size. *Human organization*, 60(1):28–39.

McCormick, T. H. (2020). The network scale-up method. *The Oxford Handbook of Social Networks*, page 153.

McCormick, T. H., Salganik, M. J., and Zheng, T. (2010). How many people do you know?: Efficiently estimating personal network size. *Journal of the American Statistical Association*, 105(489):59–70.

McCormick, T. H. and Zheng, T. (2012). Latent demographic profile estimation in hard-to-reach groups. *The annals of applied statistics*, 6(4):1795.

McCormick, T. H. and Zheng, T. (2015). Latent surface models for networks using aggregated relational data. *Journal of the American Statistical Association*, 110(512):1684–1695.

McCormick, T. H., Zheng, T., et al. (2007). Adjusting for recall bias in "how many x's do you know?" surveys. In *Proceedings of the joint statistical meetings*.

Mohebbi, E., Baneshi, M. R., Haji-Maghsoodi, S., and Haghdoost, A. A. (2014). The application of network scale up method on estimating the prevalence of some disabilities in the southeast of iran. *Journal of research in health sciences*, 14(4):272–275.

Moody, J. (2005). Fighting a hydra: A note on the network embeddedness of the war on terror. *Structure and Dynamics*, 1(2).

Narouee, S., Shati, M., Nasehi, M., and Dadgar, F. (2019). The size estimation of injection drug users (idus) using the network scale-up method (nsum) in iranshahr, iran. *Medical Journal of The Islamic Republic of Iran (MJIRI)*, 33(1):972–976.

Nasiri, N., Abedi, L., Hajebi, A., Noroozi, A., Khalili, M., Chegeni, M., Nili, S., Taheri-Soodejani, M., Noroozi, M., Shahesmaeili, A., et al. (2019). Population size estimation of tramadol misusers in urban population in iran: Synthesis of methods and results. *Addiction & Health*, 11(3):173.

Nikfarjam, A., Hajimaghsoudi, S., Rastegari, A., Haghdoost, A. A., Nasehi, A. A., Memaryan, N., Tarjoman, T., and Baneshi, M. R. (2017). The frequency of alcohol use in iranian urban population: the results of a national network scale up survey. *International journal of health policy and management*, 6(2):97.

Nikfarjam, A., Shokoohi, M., Shahesmaeili, A., Haghdoost, A. A., Baneshi, M. R., Haji-Maghsoudi, S., Rastegari, A., Nasehi, A. A., Memaryan, N., and Tarjoman, T. (2016). National population size estimation of illicit drug users through the network scale-up method in 2013 in iran. *International Journal of Drug Policy*, 31:147–152.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rastegari, A., Baneshi, M. R., Haji-Maghsoudi, S., Nakhaee, N., Eslami, M., Malekafzali, H., and Haghdoost, A. A. (2014). Estimating the annual incidence of abortions in iran applying a network scale-up approach. *Iranian Red Crescent Medical Journal*, 16(10).

Rastegari, A., Haji-Maghsoudi, S., Haghdoost, A., Shatti, M., Tarjoman, T., and Baneshi, M. R. (2013). The estimation of active social network size of the iranian population. *Global journal of health science*, 5(4):217.

Rogerson, P. A. (1997). Estimating the size of social networks. *Geographical Analysis*, 29(1):50–63.

Rohe, K. et al. (2019). A critical threshold for design effects in network sampling. *Annals of Statistics*, 47(1):556–582.

Rwanda Biomedical Center (2012). Estimating the size of key populations at higher risk of hiv through a household survey (esphs) rwanda 2011. Technical report, Technical Report. Calverton, MD: RBC/IHDPC, SPF, UNAIDS, and ICF International.

Sabin, K., Zhao, J., Calleja, J. M. G., Sheng, Y., Garcia, S. A., Reinisch, A., and Komatsu, R. (2016). Availability and quality of size estimations of female sex workers, men who have sex with men, people who inject drugs and transgender women in low-and middle-income countries. *PLoS One*, 11(5).

Sajjadi, H., Shushtari, Z. J., Shati, M., Salimi, Y., Dejman, M., Vameghi, M., Karimi, S., and Mahmoodi, Z. (2018). An indirect estimation of the population size of students with high-risk behaviors in select universities of medical sciences: A network scale-up study. *PloS one*, 13(5).

Salganik, M. J., Fazito, D., Bertoni, N., Abdo, A. H., Mello, M. B., and Bastos, F. I. (2011a). Assessing network scale-up estimates for groups most at risk of hiv/aids: evidence from a multiple-method study of heavy drug users in curitiba, brazil. *American journal of epidemiology*, 174(10):1190–1196.

Salganik, M. J., Mello, M. B., Abdo, A. H., Bertoni, N., Fazito, D., and Bastos, F. I. (2011b). The game of contacts: estimating the social visibility of groups. *Social networks*, 33(1):70–78.

Sharifi, H., Karamouzian, M., Baneshi, M. R., Shokoohi, M., Haghdoost, A., McFarland, W., and Mirzazadeh, A. (2017). Population size estimation of female sex workers in iran: Synthesis of methods and results. *PloS one*, 12(8).

Shati, M., Haghdoost, A., Majdzadeh, R., Mohammad, K., and Mortazavi, S. (2014). Social network size estimation and determinants in tehran province residents. *Iranian journal of public health*, 43(8):1079.

Sheikhzadeh, K., Baneshi, M. R., Afshari, M., and Haghdoost, A. A. (2016). Comparing direct, network scale-up, and proxy respondent methods in estimating risky behaviors among collegians. *Journal of Substance Use*, 21(1):9–13.

Shelley, G. A., Bernard, H. R., Killworth, P., Johnsen, E., and McCarty, C. (1995). Who knows your hiv status? what hiv+ patients and their network members know about each other. *Social networks*, 17(3-4):189–217.

Shelley, G. A., Killworth, P. D., Bernard, H. R., McCarty, C., Johnsen, E. C., and Rice, R. E. (2006). Who knows your hiv status ii?: Information propagation within social networks of seropositive people. *Human Organization*, 65(4):430.

Shelton, J. F. (2015). Proposed utilization of the network scale-up method to estimate the prevalence of trafficked persons. In *Forum on Crime & Society*, volume 8.

Shokoohi, M., Baneshi, M. R., and Haghdoost, A. A. (2010). Estimation of the active network size of kermanian males. *Addiction & health*, 2(3-4):81.

Shokoohi, M., Baneshi, M. R., and Haghdoost, A.-A. (2012). Size estimation of groups at high risk of hiv/aids using network scale up in kerman, iran. *International Journal of Preventive Medicine*, 3(7):471.

Snidero, S., Morra, B., Corradetti, R., and Gregori, D. (2007). Use of the scale-up methods in injury prevention research: An empirical assessment to the case of choking in children. *Social Networks*, 29(4):527–538.

Snidero, S., Soriani, N., Baldi, I., Zobec, F., Berchialla, P., and Gregori, D. (2012). Scale-up approach in cati surveys for estimating the number of foreign body injuries in the aerodigestive tract in children. *International journal of environmental research and public health*, 9(11):4056–4067.

Snidero, S., Zobec, F., Berchialla, P., Corradetti, R., and Gregori, D. (2009). Question order and interviewer effects in cati scale-up surveys. *Sociological Methods & Research*, 38(2):287–305.

Sulaberidze, L., Mirzazadeh, A., Chikovani, I., Shengelia, N., Tsereteli, N., and Gotsadze, G. (2016). Population size estimation of men who have sex with men in tbilisi, georgia; multiple methods and triangulation of findings. *PloS one*, 11(2).

Teo, A. K. J., Prem, K., Chen, M. I., Roellin, A., Wong, M. L., La, H. H., and Cook, A. R. (2019). Estimating the size of key populations for hiv in singapore using the network scale-up method. *Sexually transmitted infections*, 95(8):602–607.

UNAIDS and WHO (2010). Guidelines on estimating the size of populations most at risk to hiv. *Geneva, Switzerland: World Health Organization*, page 51.

Vardanjani, H. M., Baneshi, M. R., and Haghdoost, A. (2015). Cancer visibility among iranian familial networks: to what extent can we rely on family history reports? *PloS one*, 10(8).

Verdery, A. M., Weir, S., Reynolds, Z., Mulholland, G., and Edwards, J. K. (2019). Estimating hidden population sizes with venue-based sampling: Extensions of the generalized network scale-up estimator. *Epidemiology*, 30(6):901–910.

Wang, J., Yang, Y., Zhao, W., Su, H., Zhao, Y., Chen, Y., Zhang, T., and Zhang, T. (2015). Application of network scale up method in the estimation of population size for men who have sex with men in shanghai, china. *PloS one*, 10(11).

Yang, X. Y. and Yang, F. (2017). Estimating religious populations with the network scale-up method: A practical alternative to self-report. *Journal for the Scientific Study of Religion*, 56(4):703–719.

Zahedi, R., Noroozi, A., Hajebi, A., Haghdoost, A. A., Baneshi, M. R., Sharifi, H., and Mirzazadeh, A. (2018). Self-reported and network scale-up estimates of substance use prevalence among university students in kerman, iran. *Journal of research in health sciences*, 18(2).

Zamanian, M., Baneshi, M. R., Haghdoost, A., and Zolala, F. (2016). Estimating the visibility rate of abortion: a case study of kerman, iran. *BMJ open*, 6(10):e012761.

Zamanian, M., Zolala, F., Haghdoost, A. A., Haji-Maghsoudi, S., Heydari, Z., and Baneshi, M. R. (2019). Methodological considerations in using the network scale up (nsu) for the estimation of risky behaviors of particular age-gender groups: An example in the case of intentional abortion. *PloS one*, 14(6).

Zheng, T., Salganik, M. J., and Gelman, A. (2006). How many people do you know in prison? using overdispersion in count data to estimate social structure in networks. *Journal of the American Statistical Association*, 101(474):409–423.