Thirty Years of The Network Scale-up Method

Ian Laga, Le Bao & Xiaoyue Niu

To cite this article: Ian Laga, Le Bao & Xiaoyue Niu (2021) Thirty Years of The Network Scale-up Method, Journal of the American Statistical Association, 116:535, 1548-1559, DOI: 10.1080/01621459.2021.1935267

To link to this article: https://doi.org/10.1080/01621459.2021.1935267

Published online: 21 Jul 2021.
Thirty Years of The Network Scale-up Method

Ian Laga, Le Bao, and Xiaoyue Niu
Department of Statistics, Eberly College of Science, Pennsylvania State University, University Park, PA

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 dataset 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.

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 subpopulations with the highest HIV prevalence. 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 Xs 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 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 (Brezas 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 article and provide a brief introduction to relevant network and NSUM terminology and notation to streamline reading of later sections. Aggregated relational data (ARD) refer 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 $N_k$, 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, and Kimberly), Native Americans, commercial pilots, and homicide victims. The three unknown subpopulations were HIV positive individuals, women raped in the past 12 months, and homicide victims. These unknown subpopulations formed 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, that is,

$$\frac{y_{ik}}{d_i} = \frac{N_k}{N} \quad (1)$$

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, Kumar, and Sarlos 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 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 introduced 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) walked 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 while McCormick (2020) spent more time discussing designing and analyzing an ARD survey. In addition, we apply many of the models discussed in this review to one canonical dataset 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 article 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 NSUM 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 ARD

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 (i) transmission errors, (ii) barrier effects, (iii) recall errors, and (iv) 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 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, Salganik, and Gelman 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, \( \tau \), of a subpopulation. Also frequently called the visibility factor, \( \tau \) 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 and Zheng 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 undercount or overestimate the true number of alters. Little if any research has been done on 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 progresses, 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 underestimated the number of alters in large subpopulations (Killworth et al. 2003; Zheng, Salganik, and Gelman 2006; McCormick and Zheng 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 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 respon-
dent 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 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 (i) frequentist models which provide subpopulation size estimates, (ii) Bayesian models which handle the biases through the prior distributions, and (iii) 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

\[
P(W_k) = \sum_{m=d_{\text{min}}}^{d_{\text{max}}} P(W_k | y_{ik} = m) P(y_{ik} = m),
\]

where the degrees can vary over the integers from \(d_{\text{min}}\) to \(d_{\text{max}}\).

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 \(g\), \(1 \leq g \leq d_{\text{max}}\), such that

\[
P(W_k) = \sum_{m=d_{\text{min}}}^{d_{\text{max}}} (1 - N_k/(N - g))^m P(y_{ik} = m)
\]

\[
\approx \sum_{m=d_{\text{min}}}^{d_{\text{max}}} (1 - N_k/N)^m P(y_{ik} = m),
\]

where the approximation holds since \(g\) is small with respect to \(N\). Now consider (3) for subpopulations 1 to \(L\), with corresponding values \(\epsilon_1 = 1 - N_{i1}/N\) to \(\epsilon_L = 1 - N_{iL}/N\). 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

\[
1 > P(W_{i1}) > P(W_{i2}) > \cdots > P(W_{iL}).
\]

Therefore, if \(P(W_{iu})\) for the unknown subpopulation \(u\) is such that \(P(W_{i-1}) > P(W_{iu}) > P(W_{i})\), then the subpopulation size is also bounded, where

\[
1 - \frac{N_{j-1}}{N} > 1 - \frac{N_k}{N} > 1 - \frac{N_j}{N} \implies \frac{N_{j-1}}{N} < \frac{N_k}{N} < \frac{N_j}{N}.
\]

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 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 \(d_i\):

\[
L(d; y) = \prod_{k=1}^{L} \left( \frac{d_i}{y_{ik}} \right) \left( \frac{N_k}{N} \right)^{y_{ik}} \left( 1 - \frac{N_k}{N} \right)^{d_i - y_{ik}}.
\]

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

\[
\hat{d}_i = N \sum_{k=1}^{L} \frac{y_{ik}}{N}.
\]

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

\[
\hat{N}_u^{\text{PIMLE}} = \frac{1}{n} \sum_{i=1}^{n} N_i \cdot y_{iu}/\hat{d}_i,
\]

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 \(\hat{N}_u^{\text{PIMLE}}\) 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 Equation (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

\[
L(N_k; y, \{d_i\}) = \prod_{i=1}^{n} \left( d_i \left( \frac{N_u}{N} \right)^{y_iu} \left( \frac{N - N_u}{N} \right)^{d_i-y_iu} \right)
\]

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

\[
\hat{N}_{u, MLE} = N \cdot \frac{\sum_{i=1}^{n} y_{iu}}{\sum_{i=1}^{n} d_i} = \frac{\sum_{i=1}^{n} y_{iu}}{\sum_{i=1}^{n} \frac{\sum_{k=1}^{L} N_k}{N}} y_{iu} \cdot \frac{\sum_{k=1}^{L} N_k}{N} \cdot \frac{1}{d_i}.
\]

(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

\[
SE(N_{u, MLE}) = \sqrt{\frac{N \cdot N_u}{\sum_{i=1}^{n} d_i}}.
\]

(10)

which decreases as the network sizes \( d_i \) increases. When the degrees are small or prevalence is relatively large, the standard error in Equation (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. To weight the subpopulations equally, the mean of sums (MoS) estimator first estimates the \( \hat{d}_i \) based on each subpopulation, averages those \( \hat{d}_i \)'s, and then back-estimates \( N_u \) based on the \( \hat{d}_i \)'s (Habecker, Dombrowski, and Khan 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 \( \hat{N}_u \),

\[
\hat{N}_{u, MoS} = N \cdot \frac{\sum_{i=1}^{n} y_{iu} w_i}{\sum_{i=1}^{n} d_i}.
\]

(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, Dombrowski, and Khan (2015) proposed controlling this variance by choosing subpopulations of similar size.

Habecker, Dombrowski, and Khan (2015) also proposed improving 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

\[
\hat{N}_{u, WML} = N \cdot \frac{\sum_{i=1}^{n} y_{iu} w_i}{\sum_{i=1}^{n} d_i}
\]

and

\[
\hat{N}_{u, WMoS} = \frac{\sum_{i=1}^{n} (y_{iu} w_i)/\hat{d}_i}{\sum_{i=1}^{n} d_i},
\]

(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 \( k \) 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 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 \( G_u \), that is, 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 \( y_{G,i} = \sum_{i \in G} y_{iu} \) to be the total number of in-reports and \( v_{G,G} = \sum_{i \in G} v_{i,G} \) to be the total number of out-reports, then \( v_{G,G} = v_{G,F} \). Multiplying both sides by \( N_u \), we can write \( N_u = y_{G,F}/v_{G,F} \). This must also be the case for the frame population, that is, \( N_u = y_{F,u}/v_{F,u} \). Then, the estimator \( \hat{N}_{u, GNSUM} \) is found by estimating the numerator and denominator separately, that is,

\[
\hat{N}_{u, GNSUM} = \frac{\hat{y}_{G,F}}{\hat{v}_{F,u}}.
\]

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

Feehan and Salganik (2016) made several important connections between their model 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 article. This decomposition also leads to 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 \( \hat{v}_{i,F} \) 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) used 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 Overdispersed Model

Zheng, Salganik, and Gelman (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, Salganik, and Gelman (2006) proposed the overdispersed model, given by

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

The parameter \( \alpha_i = e^{\delta_i} \) represents the expected degree of respondent \( i \), \( \beta_k = e^{\delta_k} \) represents the proportion of total links that involve subpopulation \( k \), and \( \gamma_{ik} \) allows for extra variability in the model not accounted for by \( \alpha_i \) and \( \beta_k \). Thus, if the \( \gamma_{ik} \) are constant, then 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, \( \gamma_{ik} \) 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 \( 1/(\log k - 1) \), 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 \( \text{var}(y_{ik}) = \omega_k e^{\alpha_i + \beta_k} \).

Note that the \( \alpha_i \)'s and \( \beta_k \)'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, Salganik, and Gelman (2006).

The main motivation and utility of this article are the relationship between the subpopulations and the overdispersion, \( \omega_k \). Figure 4 of the original article provides a nice visual summary of the overdispersion estimates. For the considered dataset, 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, Salganik, and Zheng (2010) noted that the normalization procedure of Zheng, Salganik, and Gelman (2006) did not ensure the degrees are estimated accurately since the transmission errors and the barrier effects can still bias the degree estimates. Thus, the authors proposed introducing latent nonrandom mixing to account for the biases. Unlike previous models, the McCormick, Salganik, and Zheng (2010) model estimated 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

\[
y_{ik} \sim \text{negative-binomial} \left( \text{mean} = \mu_{ike}, \text{overdispersion} = \omega_k \right), \tag{17}
\]

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

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

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 and Zheng (2007) to the mean of their negative binomial model, that is, replace the previous mean with

\[
\mu_{ike} = df \left( \sum_{a=1}^{A} m(e,a) \frac{N_{ak}}{N_a} \right), \tag{19}
\]
where the full details of the calibration curve $f(x)$ can be found in the original article.

The authors also provided 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 (i) there is random mixing or (ii) the known subpopulations represent a scaled-down population, that is,

$$\sum_{k=1}^{L} N_{lk} \sum_{k=1}^{K} N_k = N_N.$$  \hspace{1cm} (20)

In words, McCormick, Salganik, and Zheng (2010) explained 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 article.

Note, however, that the McCormick, Salganik, and Zheng (2010) model did not estimate the size of unknown subpopulations, focusing instead on estimating the individual degree and the distribution of network sizes. In a follow-up article, McCormick and Zheng (2012) extended the McCormick, Salganik, and Zheng (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 Equations (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) contained 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 article 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

$$y_{ik} \sim \text{Binom} \left( \frac{d_i e^{\tau_k q_{ik}}}{} \right),$$  \hspace{1cm} (21)

where $d_i \sim \text{log-normal}(\mu, \sigma^2)$, $r_k \sim N \left( a + b \log(N_k), \sigma^2 \right)$, $q_{ik} \sim \text{Beta} \left( \text{mean} = m_k, \text{dispersion} = \rho_k \right)$, and $\tau_k \sim \text{Beta} \left( \text{mean} = \eta_k, \text{dispersion} = \nu_k \right)$. The parameters $r_k$ handle the recall error, $q_{ik}$ handle the barrier effects, $\tau_k$ (fixed at 1 for known subpopulations) handle the transmission effects, and $d_i$ represents the varying degree between respondents. Furthermore, the 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 $\tau_k$ 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, Salganik, and Gelman 2006; McCormick and Zheng 2007; McCormick, Salganik, and Zheng 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

$$y_{ik} \sim \text{Poisson} \left( \lambda a_i \exp \{ \beta_k [x_{i,k} - U_k] \} N_k \right),$$  \hspace{1cm} (22)

while transmission error and barrier effect model is then given by

$$y_{ik} \sim \text{Poisson} \left( \lambda a_i \exp \{ \beta_k [x_{i,k} - U_k] \} \exp \left\{ \sum_{j=1}^{p} y_{ijk} z_i \right\} N_k \right),$$  \hspace{1cm} (23)

where $U$ represents the upper bound of the Likert scale, $x_{i,k}$ is respondent $i$’s response to the question, and $z = (z_1, \ldots, z_n)$ represents the column-centered covariate matrix. Like McCormick, Salganik, and Zheng (2010) and 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 are 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, Raftery, and Handcock (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 $\delta_{ij}$ 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 $\lambda_{ik} = \sum_{j \in G_k} P(\delta_{ik} = 1 | Z_i, Z_j \in G_k)$. In the complete network case, $j \in G_k$ are observed and known, while for ARD they are unobserved, making it impossible to calculate $\lambda_{ik}$ directly. Instead, the rate is approximated by

$$
\lambda_{ik} \approx N_k \int_{Z \in G_k} P(\delta_{ik} = 1 | Z_i, Z_j \in G_k)P(Z_j \in G_k)dZ_j \in G_k,
$$

(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

$$
y_{ik} | d_i, \beta_k, \xi, \eta_k, \theta(z_i, v_k) \sim \text{Poisson}
$$

$$
\left( d_i \beta_i \left( \frac{C_{p+1}(\xi)C_{p+1}(\eta_k)}{C_{p+1}(0)C_{p+1}(\sqrt{\xi^2 + \eta_k^2 + 2 \xi \eta_k \cos(\theta(z_i, v_k))})} \right) \right),
$$

(25)

where $C_{p+1}(\cdot)$ is the normalizing constant of the von-Mises Fisher distribution and $\theta(z_i, v_k)$ 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, Salganik, and Gelman (2006), and McCormick and Zheng (2015) offered 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 dataset 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
Brez 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 article is remarkably different from our previous discussions, we refer the reader to the original article for model details.

Brez 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 article 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 population parameters of networks, network sampling methods (Bhattacharyya et al. 2015; Green and Shalizi 2017; Levin and Levin 2019; Lin, Lunde, and Sarkar 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 datasets 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, Salganik, and Zheng 2010). It is unlikely that a respondent reports knowing someone without also knowing their name, thereby removing 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, Dombrowski, and Khan (2015) proposed trimming one subpopulation at a time by removing only the worst performing subpopulation.
removing one other subpopulation. 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 dataset. 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 an 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, Salganik, and Zheng (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 datasets. McCormick and Zheng (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 and Zheng (2007), let $e^{\beta_{k}}$ 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

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

where $a$ controls how fast the derivative of this curve approaches 1/2 and $b$ controls at what value of $\beta_k$ 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 error-in-variables model $\log(\hat{N}_k) = a + b \log(N_k) + \delta_k + \epsilon_k$ is fit via maximum likelihood to estimate $a$, $b$, and the variances of $\delta_k$ and $\epsilon_k$, where $\delta_k \sim N(0, \delta^2_k)$ and $\epsilon_k \sim N(0, \epsilon^2_k)$. Estimates of $\log(N_k)$ from the posterior are then transformed using the estimated $a$, $b$, and an additional random noise. The transformation is given by

$$\frac{Y^{[t]}_k - a}{b} + Z, \quad (27)$$

where $Y^{[t]}_k$ is the $t$th MCMC sample from the posterior distribution of $\log(N_k)$ and $Z \sim N(0, \sigma^2_Z / b^2)$.

## 5. Discussion

In this article, 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 article. 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 deserves 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/registration, 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 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, Salganik, and Gelman (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 dataset 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 Xs 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 estimator. Finally, Breza et al. (2017) suggested that ARD should move beyond social networks. For example, the authors considered 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 articles 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.

Acknowledgments
The authors thank to Tian Zheng for the helpful discussions, and Tyler McCormick and Alex Cook for sharing their code.

Funding
This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI136664.

References

Bhattacharyya, S., Bickel, P. J. (2015), “Subsampling Bootstrap of Count Features of Networks,” *Annals of Statistics*, 43, 2384–2411. [1555]

Breza, E., and Chandrasekhar, A. G. (2019), “Social Networks, Reputation, and Commitment: Evidence From a Savings Monitors Experiment,” *Econometrica*, 87, 175–216. [1554]

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. [1548,1554,1555,1556] [1555]

——— (2019), “Consistently Estimating Graph Statistics Using Aggregated Relational Data,” arXiv: 1908.09881. [1555,1556]

Dasgupta, A., Kumar, R., and Sarlos, T. (2014), “On Estimating the Average Degree,” in *Proceedings of the 23rd International Conference on World Wide Web*, pp. 795–806. [1549]

Feehan, D., and Salganik, M. (2014), *The Network Reporting Package*. [1549]

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, 153–186. [1548,1552,1556]

Green, A., and Shalizi, C. R. (2017), “Bootstrapping Exchangeable Random Graphs,” arXiv: 1711.00813. [1555]

Guo, W., Bao, S., Lin, W., Wu, G., Zhang, W., Hladik, W., Abdul-Quader, A., Bultery, 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, e17196. [1555]

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, e0143406. [1552,1555,1556]

Haghdost, 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, e2018041. [1550]

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, 1090–1098. [1555]

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. [1550,1551]

Johnsen, E. C., Bernard, H. R., Killworth, P. D., Shelley, G. A., and McCartney, C. (1995), “A Social Network Approach to Corroborating the Number of AIDS/HIV+ Victims in the US,” *Social Networks*, 17, 167–187. [1549]

Killworth, P. D., Johnsen, E. C., McCartney, C., Shelley, G. A., and Bernard, H. R. (1998a), “A Social Network Approach to Estimating Seroprevalence in the United States,” *Social Networks*, 20, 23–50. [1551,1552,1554,1556]

Killworth, P. D., McCartney, 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, 141–160. [1550,1554,1556]

Killworth, P. D., McCartney, 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, 289–308. [1549,1552,1556]

Killworth, P. D., McCartney, 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, 84–112. [1550]

Levin, K., and Levine, E. (2019), “Bootstrapping Networks With Latent Space Structure,” arXiv: 1907.10821. [1555]

Lin, Q., Lunde, R., and Sarkar, P. (2020), “On the Theoretical Properties of the Network Jackknife,” *Proceeding of the International Conference on Machine Learning*, 119, 6105–6115. [1555]

Maltiel, R., and Baraff, A. J. (2015), NSUM: Network Scale Up Method (R package version 1.0). [1549,1554]

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, 1247–1277. [1548,1550,1554,1556]

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, 28–39. [1549,1550]

McCormick, T. H. (2020), “The Network Scale-up Method,” in *The Oxford Handbook of Social Networks*, eds. R. Light and J. Moody, Oxford University Press, p. 153. [1549]

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, 59–70. [1553,1554,1555,1556]

McCormick, T. H., and Zheng, T. (2012), “Latent Demographic Profile Estimation in Hard-to-Reach Groups,” *The Annals of Applied Statistics*, 6, 1795–1813. [1554,1556]

——— (2015), “Latent Surface Models for Networks Using Aggregated Relational Data,” *Journal of the American Statistical Association*, 110, 1684–1695. [1556,1555,1556]

McCormick, T. H., and Zheng, T. (2007), “Adjusting for Recall Bias in ‘How Many X’s Do You Know?’ Surveys,” in *Proceedings of the Joint Statistical Meetings*, [https://www.stat.columbia.edu/~tzheng/files/McCormick_jsm07.pdf](https://www.stat.columbia.edu/~tzheng/files/McCormick_jsm07.pdf). [1550,1553,1554,1556]

Moody, J. (2005), “Fighting a Hydra: A Note on the Network Embeddedness of the War on Terror,” *Structure and Dynamics*, 1. [https://doi.org/10.5070/D912003275](https://doi.org/10.5070/D912003275) [1557]

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [1549,1554]

Rogerson, P. A. (1997), “Estimating the Size of Social Networks,” *Geographical Analysis*, 29, 50–63. [1549]

Rohe, K. (2019), “A Critical Threshold for Design Effects in Network Sampling,” *Annals of Statistics*, 47, 556–582. [1557]

Sabin, K., Zhao, J., Callera, 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, e0155150. [1548]

Salganik, M. J., Fazio, 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, 1190–1196. [1550,1557]

Salganik, M. J., Mello, M. B., Abdo, A. H., Bertoni, N., Fazio, D., and Bastos, F. I. (2011b), “The Game of Contacts: Estimating the Social Visibility of Groups,” *Social Networks*, 33, 70–78. [1549]

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, 189–217. [1550,1556]

Shelton, J. F. (2015), “Proposed Utilization of the Network Scale-up Method to Estimate the Prevalence of Trafficked Persons,” in *Forum on Crime & Society* (Vol. 8) ed. United Nations, United Nations Publications. [1557]

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, 287–305. [1550]

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, 602–607. [1554,1556]

UNAIDS and WHO (2010), *Guidelines on Estimating the Size of Populations Most at Risk to HIV*, Geneva, Switzerland: World Health Organization, p. 51. [1548]

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, 901–910. [1553,1556]

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, 409–423. [1550,1553,1554,1555,1556,1557]