

# Performance Analysis of NSUM Estimators in Social-Network Topologies

Sergio Díaz-Aranda<sup>1,2\*</sup>    Jose Aguilar<sup>1,3</sup>    Juan Marcos Ramírez<sup>1</sup>  
David Rabanedo<sup>2</sup>    Antonio Fernández Anta<sup>1</sup>    Rosa E. Lillo<sup>2,4</sup>

<sup>1</sup>IMDEA Networks Institute

<sup>2</sup>uc3m-Santander Big Data Institute, Universidad Carlos III de Madrid

<sup>3</sup> Universidad de Los Andes

<sup>4</sup>Department of Statistics, Universidad Carlos III de Madrid

## Abstract

The Network Scale-up Methods (NSUM) are methods to estimate unknown populations based on indirect surveys in which the participants provide information about aggregated data of their acquaintances. This preserves the privacy and may lead to higher participation. During the last thirty years, new NSUM estimators have emerged. However, conditions related to the design of the experiments and the robustness of the estimators have not been studied in depth, especially in a realistic simulation environment. This study aims to compare nine NSUM estimators under relevant conditions in the literature through simulation experiments. We have analyzed how the NSUM is affected by the network topology, transmission and recall errors, the distribution of the unknown subpopulation, the number and sizes of subpopulations, and sample size. This article shows that some NSUM estimators barely used are better and more robust to some conditions, especially when the network is scale-free or under barrier effects. In addition, some methods are very sensitive to recall errors. In terms of the subpopulations configuration, we observe that the number of known subpopulations usually employed is quite large and that the most common NSUM is robust to the number and sizes of the subpopulations.

*Keywords:* Aggregated relational data, Indirect surveys, Size estimation, Ego-networks

---

\*This work has been supported by *Ministerio de Ciencia e Innovación, Gobierno de España*, grant numbers PDC2022-133359, PID2022-137243OB-100 and TED2021-131264B-100 funded by MCIN/AEI/10.13039/501100011033, *European Union NextGenerationEU/PRTR*. This initiative has also been partially carried out within the framework of *Recovery, Transformation and Resilience Plan funds*, financed by the *European Union (Next Generation)* through the grant *ANTICIPA*. The work of Sergio Díaz-Aranda has been funded by *Comunidad de Madrid* predoctoral grant PIPF-2022/COM-24467.

# 1 Introduction

The Network Scale-up Methods (NSUM) are survey-based methods that use information about the acquaintances of the participants instead of their own. The methods arise from the fact that surveys with sensitive populations face reliability issues since the respondents might not be willing to participate or to be honest, resulting in sampling bias and underestimation. Over the last few decades, indirect reporting has emerged as a strategy to address these issues, which can increase participation and data collection while reducing study costs. The NSUM studies only require aggregate data of the contact network, which is obtained through the reporting of different groups known by the respondent (what is called aggregated relational data (ARD) (Feehan, Salganik, 2016)), using considerably less data and respecting the privacy of the surveyed. For these reasons, the NSUM have been employed to estimate the size of subpopulations in which no sampling frame exists and public acknowledgment of membership in the population is potentially threatening, where the term subpopulation denotes a subset of the general population, and size refers to the number of people in the subpopulation. Based on the work of (Bernard et al., 1989), it has been employed in several contexts such as abortions (Sully et al., 2020), female sex workers (Jing et al., 2018), men who have sex with men (Ezoe et al., 2012), drug users (Salganik et al., 2010), and HIV issues, currently used by the UNAIDS<sup>1</sup> (UNAIDS, 2010).

Throughout these years much research about NSUM has emerged (see for example, the reviews (Laga et al., 2021) and (McCormick, 2020)) focusing mainly on the introduction of new estimators, especially in the presence of bias (see (Laga et al., 2021)). However, the performance and comparison of the methods under various conditions have not been studied in depth. These conditions include the performance under biases, sample size, the unknown subpopulation, and the selection of the known subpopulation. We will refer to unknown population as a subpopulation whose size is not known, and to known subpopulations as subpopulations whose size is known and are used to estimate the network size of the participants. In particular, there is no guideline on the sizes of the known subpopulations chosen, the number of known subpopulations to employ, or the influence of the intersection of the known subpopulations, which we will denote as overlapping of subpopulations. Most

---

<sup>1</sup>The Joint United Nations Programme on HIV/AIDS

of the studies only compare two estimators (Killworth et al., 1998b), (Salganik et al., 2011), (Snidero et al., 2004) (the last two only in one dataset), or three (Kunke et al., 2024). A comparison of many of them was done in the supplementary material of (Laga et al., 2021) but for a single dataset. (Snidero et al., 2004) analyzed two NSUM estimators in Monte Carlo experiments to analyze the robustness under the size, coverage, optimal number, and overlapping of the populations. (Kunke et al., 2024) compared the estimators given in (Killworth et al., 1998a) and (Killworth et al., 1998b) under the prevalence of the unknown subpopulation, dissortative and assortative mixing, and degree ratio. The number of subpopulations required for the first NSUM estimator was studied by (Killworth et al., 1998a) in an elementary Monte Carlo simulation. (Srivastava et al., 2024) proposed a latent network model obtaining analytical conditions in terms of variance and dependency with respect to the network size, in which the NSUM error is lower than classical methods relying solely on participant information. Finally, analytical error bounds for two classical NSUM estimators were provided in (Díaz-Aranda et al., 2024), with refinements for random networks.

The main contribution is the comparison of a large set of NSUM estimators, nine in total, through an extensive simulation study under the presence of the relevant conditions of the NSUM literature. Specifically, we explore the impact of the distribution of the subpopulations (number of subpopulations, sizes, presence of overlapping), the distribution of the unknown subpopulation, sample size, and the presence of the next biases: transmission error and recall error. Actually, to our knowledge, it is the first work that compares the performance of the NSUM under the transmission and recall errors. Additionally, we present a way to model the biases transmission error and recall error. Finally, we measure the NSUM performance in a realistic simulation environment, where we included social network topologies that can be found in real data, such as small world and preferential attachment networks.

This article is organized as follows. Section 2 presents a historical overview of the NSUM and the biases faced by the estimators. Section 3 displays the NSUM estimators that will be used in the subsequent simulations. In Section 4, we present the parameters and domains that will be compared. The results of the simulations are exhibited in Section

5, along with a discussion of estimator performance. Section 6 assesses the results on a real network using both real and simulated subpopulations. Final remarks are in Section 7.

## 2 Overview of the Network Scale-Up Methods

The NSUM are the methods that employed ARD to estimate unknown subpopulations or personal network sizes (Laga et al., 2021). The ARD is the data collected from the ego-networks (networks surrounding the subjects that contain the participants and their contacts (Newman, 2018)) using questions of the type: “How many people do you know who have characteristic X?”. The individual from which the networks are generated is called ego and the contacts are denoted as alters. The general framework of all the methods is that the ego-network contains information that is approximately representative of the entire population, at least for the ARD questions. More concretely, suppose that we have a population of size  $N$  and several ARD questions, the main assumption is that the following equations hold:

$$\mathbb{E} \left[ \frac{y_{ij}}{d_i} \right] = \frac{N_j}{N} \quad (1)$$

where  $y_{ij}$  represents the number of people that participant  $i$  knows in the subpopulations  $j$ ,  $d_i$  is the degree, the size of the ego-network, and  $N_j$  is the total number of people belonging to the subpopulation  $j$ . We will consider  $L$  known subpopulation, with known sizes denoted as  $N_k$ , and an unknown subpopulation. We will refer to  $y_{iu}$  and  $N_u$  as the number of people that participant  $i$  knows in the unknown subpopulation, and the size of the unknown subpopulation, which is often called as “hidden population” or “hard-to-reach population”. The NSUM uses the ARD provided by the respondents to obtain values for the participants’ degrees. This is mainly done using known subpopulations, although other approaches ask directly for the degrees (Garcia-Agundez et al., 2021). Additional samples of the unknown population are employed in some estimators (Feehan, Salganik, 2016) and (Verdery et al., 2019).

Over the years, several biases have been discovered that affect the constant proportionality assumption, and as a consequence, the performance of the NSUM. These biases are the barrier effect, transmission error, and recall error. These errors come from survey

behavior and the way the different subpopulations are distributed. The transmission error is the bias produced by the fact that the participant does not know if the acquaintances are in the subpopulations. The recall error occurs when the participants recall inaccurately the number of alters of some subpopulations. The participant may overestimate or underestimate this number. The barrier effect is the bias produced because the probability of knowing someone in a population depends on the characteristics of the ego. This means that a participant is more likely to know someone in certain subpopulations.

### 3 NSUM estimators

(Laga et al., 2021) reviews the previous NSUM in the literature from which we will take most of them. For simplicity, we will exclude estimators that use demographic variables (McCormick, Zheng, 2012), and venue-based sampling (Verdery et al., 2019). We include the estimator of (Garcia-Agundez et al., 2021), and two unemployed estimators, one of them pointed out by (Kunke et al., 2024). During the development of the paper, a new NSUM has emerged (Laga et al., 2023), that takes into account the correlations between the variables. We do not analyze the correlation of the variables in our work, but the overlapping of the known subpopulation may lead to correlated ARD responses because the participants are more likely to know a person in several subpopulations. Our results do not show a significant variation in the performance of the estimators in this case, as we show in the supplementary material.

#### 3.1 Frequentist estimators

The frequentist estimations usually use a two-step procedure: an estimation of the degrees of the subjects, and a posterior estimation of the unknown subpopulation using that information. In terms of the degrees, in some cases, they use the fundamental Eq. (1) for the unknown subpopulations to obtain  $L$  estimates of the  $d_i$ , and then take the average:

$$\hat{d}_i^A = \frac{1}{L} \sum_{k=1}^L d_i^k, \quad (2)$$

where the  $d_i^k$  is the estimate corresponding to  $y_{ik}N/N_k$ . Another option is to assume that the  $y_{ik}$  are  $L$  independent random variables that follow a binomial distribution,  $y_{ik} \sim \text{Bin}(d_i, N_k/N)$ , and then estimate the degrees via a maximum likelihood estimation:

$$\hat{d}_i^{ML} = N \cdot \frac{\sum_{k=1}^L y_{ik}}{\sum_{k=1}^L N_k}. \quad (3)$$

For the estimation of the size of the unknown population, there are two options. One proposal is to use Eq. (1) to obtain  $n$  estimates of the number of the unknown population,  $N_u^i = Nd_i/y_{iu}$ , and to take averages,

$$\hat{N}_u^A = \frac{1}{n} \sum_{i=1}^n N_u^i. \quad (4)$$

Another proposal for the size estimation is to assume that  $y_{iu} \sim \text{Bin}(d_i, N_u/N)$  for  $i = 1, \dots, n$  are independent random variables, and to use a maximum likelihood estimation

$$\hat{N}_u^{ML} = N \frac{\sum_{i=1}^n y_{iu}}{\sum_{i=1}^n d_i}. \quad (5)$$

The NSUM proposal is to substitute the degrees for the estimates degrees in the equations for the size estimation, Eq. (4) and Eq. (5). An alternative is to employ the approach of (Garcia-Agundez et al., 2021), when participants provide the degree, which we denote as  $\delta_i$ . These six combinations are summarized in the next Table. The MLE estimator is by far the most employed NSUM, see (Ocagli et al., 2021) for a list of NSUM studies.

Degree \ Size	Average	Max. Likelihood
Average	MoS (Means of Sums) (Killworth et al., 1998a)	KLN (Kunke et al., 2024)
Max. Likelihood	Plug-in MLE (PiMLE) (Killworth et al., 1998a)	MLE (Killworth et al., 1998b)
Direct	Average of Ratios (AoR)	Ratio of Averages (RoA) (Garcia-Agundez et al., 2021)

Table 1: Basic NSUM estimators. The rows represent the way to obtain the degree and the columns the size

### 3.2 Bayesian estimators

The first estimator is the Overdispersed estimator (Zheng et al., 2006). It was created to handle the overdispersion of the data produced by the different propensities of the

subpopulation to form ties. This is represented by the parameter  $\omega_k$ . The model is:

$$y_{ik} \sim \text{Negative Binomial}(\text{mean} = e^{\alpha_i + \beta_k}, \text{overdispersion} = \omega_k). \quad (6)$$

where  $k = 1, \dots, L$  or  $u$ ,  $e^{\alpha_i}$  is the expected degree of  $i$ , and  $e^{\beta_k}$  is the proportion of links of group  $k$ . For further details about the prior distribution and the algorithm, see (Zheng et al., 2006).

On the other hand, (Teo et al., 2019) proposed a new estimator that we denote by TPC estimator. They presented the following model:

$$y_{ik} \sim \text{Poisson}(\lambda \alpha_i N_k), \quad (7)$$

with  $\lambda$  be a scale-up parameter such that  $y_{ik} = \lambda N_k$ , for every  $k = 1, \dots, L$  or  $u$ ; and  $\alpha_i$  a parameter that characterizes the variability in the network size. Details about the algorithm can be found in (Teo et al., 2019). This work includes other improvements to the model through questions about perceptions and demographic variables. These last adjustments are not included because they exceed the scope of the study.

(Maltiel et al., 2015) proposed a Bayesian model and 4 modifications to handle the biases. The Maltiel’s estimators include priors whose support is bounded. This feature generates acceptance rate problems in our data. For that reason, we have not included this model in the rest of the article. For more details about our simulations with these estimators, see the supplementary material.

### 3.3 Generalized Network Scale-Up Method

The Generalized Network Scale-up Method (GNSUM) was suggested by (Feehan, Salganik, 2016) to reduce the transmission error. In addition, the GNSUM allows other types of samplings. However, this estimator needs an additional sample from the unknown subpopulation. This sample from the unknown subpopulation allows to collect data about the number of alters in some groups who are aware of the participants belonging to the unknown subpopulation. This data is known as enriched aggregate relational data (EARD). The purpose of EARD is to estimate the relative visibility of the unknown subpopulations. The relative visibility refers to the average of the visibility of each individual in the unknown subpopulation, which is the number of contacts that know the belonging of the

individual to the unknown subpopulation. The EARD questions measure the visibility of the participants of the unknown subpopulations in some groups, denoted as alter groups. Similarly to other NSUM estimators, the EARD is used to estimate the relative visibility and then estimate the unknown subpopulation. The GNSUM estimator is:

$$\hat{N}_u^{GNSUM} = \frac{\sum_{i \in s_F} y_{iu} / \pi_i}{\hat{v}_{u,F}}, \quad (8)$$

where  $\pi_i$  are the inclusion probabilities from a sampling design and  $\hat{v}_{u,F}$  is the estimated relative visibility. In our specific case, we consider as alter groups the known subpopulations, and the sample is taken uniformly. More details about the GNSUM are in Section 1.2 in the supplementary material.

**Adjusted estimators.** (Feehan, Salganik, 2016) proved that the GNSUM estimator is equal to the unknown subpopulation estimate Eq. (5) corrected by some adjustment factors, one of them the inverse of the transmission rate. The transmission rate is the fraction of a respondent’s network that is aware the respondent belongs to the unknown subpopulation. However, the transmission rate estimation also needs the information of a sample of the unknown subpopulation. (Feehan, Salganik, 2016) proposed using the MLE estimator corrected by these factors. It is also known as GNSUM. Apart from this adjustment, we consider the adjustment of the other NSUM estimators by the inverse of an estimate of the transmission rate, which we will denote in the rest of the paper as adjusted NSUM estimators.

## 4 Simulation Environment and Parameters

We define four main scenarios characterized by the type of network and distribution of the unknown population, as shown in Table 2. For each main scenario, we define the belonging to the subpopulations, the ARD, and the EARD. Then, we compare, for several samples, the performance of the different methods as the other relevant parameters vary. These parameters of the other relevant scenarios are transmission error, recall error, number of known subpopulations, sizes of known subpopulations, presence of disjoint subpopulations (subpopulations with no overlapping), and sample size. The NSUM estimators are com-



pared using the next metrics: mean, standard deviation, mean absolute error (MAE), and mean squared error (MSE).

## 4.1 Main scenarios

	Network	unknown subpopulation
W-U	Watts-Strogatz	Uniform
W-S	Watts-Strogatz	SIR
P-U	Pref. attachment	Uniform
P-S	Pref. attachment	SIR

Table 2: Main scenarios to be compared.

### 4.1.1 Networks

We use two types of networks with properties found in real networks. On the one hand, we use a small-world network. Therefore, it has a high clustering coefficient and a short average path length (Newman, 2018). Concretely, we use a Watts-Strogatz network. The network is built as follows: the nodes are placed and are connected to the nearest neighbors, and with a fixed probability an edge is removed, and two nodes are connected. On the other hand, we use a scale-free network, a network generated by a preferential attachment process. The construction of the graph is done by adding  $m$  vertices successively so that the vertices join proportionally to the node with the most connections. The degree distribution follows a power distribution, unlike the Watt-Strogatz networks, whose nodes have similar degrees. More details about the network topologies are in the supplementary material.

### 4.1.2 Unknown subpopulation distribution

The first case is when the individuals of the unknown subpopulation are randomly chosen. In this setting, the probability is the same for every node and independent of the network. On the other hand, we model the dependencies via a SIR process through the network. It is a foundational model for infections, which models some use cases from NSUM such as drug usages (Salganik et al., 2010) and COVID-19 (Garcia-Agundez et al., 2021). In this model, there are three groups: susceptible, infected (unknown subpopulation), and recovered from

the infection. A node infects a neighbor with probability  $\beta$ , and the nodes are recovered with probability  $\gamma$ , which cannot be infected anymore.

## 4.2 Bias modeling

### 4.2.1 Transmission error

We model the transmission error through a parameter  $\nu \in [0, 1]$  that we call the visibility factor. This number represents the probability that a person who knows an acquaintance of the unknown subpopulation is informed of the belonging of this acquaintance in the unknown subpopulation. We assume that the transmission error affects uniformly the nodes in the unknown subpopulation and then the expected value of  $y_{iu}$  is:

$$E[y_{iu}] = \nu d_i \frac{N_u}{N} = d_i \frac{N_u}{N/\nu}. \quad (9)$$

In our setting, we can correct the estimator by multiplying them by  $1/\nu$ .

A measure related to the transmission error is the transmission rate (Laga et al., 2021) (Feehan, Salganik, 2016). This is one of the factors proposed by (Feehan, Salganik, 2016) to correct the MLE. The expected value of the transmission rate is  $\nu$ . We estimate this factor using the Comming-out-rate (Ezoe et al., 2012). This method requires a sample of the unknown subpopulation. It divides the average number of acquaintances who are aware of belonging to the unknown subpopulation by their average personal network size.

### 4.2.2 Recall error

The recall error is the bias due to memory inaccuracies. We model the recall error through a non-negative number  $\rho$  that we call the memory factor. There is no barrier error when  $\rho$  is zero, and the bias is more pronounced as  $\rho$  increases. We create the report of the different variables by rounding (to two decimals) a truncated normal distribution whose mean is the real value, and the standard deviation is the product of the real value times  $\rho$ . The truncation of the normal distributions is because they may produce unrealistic data. Additionally, we assume symmetry for simplicity. We only include the details for the modeling of the unknown population responses. The rest is in the supplementary material.

Recall that  $y_{iu}$  is the response of the participant  $i$  about the ARD question of the unknown subpopulation. There are two natural bounds for the responses of the unknown population:  $y_{iu} \geq 0$  and  $y_{iu} \leq \delta_i$ , where  $\delta_i$  is the response of the size of the ego-network of  $i$ . But, we add some constraints to get a distribution as symmetrical as possible,  $a = \max(-0.5, 2x_{iu} - \delta_i + 0.5)$  and  $b = \min(2x_{iu} + 0.5, \delta_i + 0.5)$ . The truncated normal distribution has mean  $\mu = x_{iu}$ , standard deviation  $\sigma = x_{iu}\rho$ , lower bound  $a$ , and upper bound  $b$ . The differences of the intervals from the mean are both equal to  $\min(x_{iu}, \delta_i - x_{iu})$ . We add and subtract 0.5 so that the boundaries of the normal can be reached. In addition, the mean and bounds are also selected to have the expected value of the real personal size. Finally, Later,  $y_{iu}$  is obtained by taking a value of the distribution and rounding it.

### 4.2.3 Barrier effect

This bias is not compared systematically in this paper. However, the SIR produces barrier effects. Additionally, the degree variation is not considered in the design phase and may produce this bias.

## 4.3 Motivation for the other parameters

### 4.3.1 Subpopulations

More than 20 subpopulations are usually employed because the PiMLE needed more than 20 subpopulations to be essentially unbiased in a simple Monte Carlo simulation. However, it is barely employed, and the simulations of (Snidero et al., 2004) showed in their simulation study that the number did not affect the performance of the MLE and PiMLE. Concerning the sizes of the subpopulations, some estimators are more affected if the subpopulations have unequal sizes, (Habecker et al., 2015). They also suggested using homogeneous subpopulations for the MoS to outperform the estimates, and because the variance is higher when the subpopulations are small. (Snidero et al., 2004) also shows that homogeneous subpopulations improve the estimates, and the bias of the PiMLE depends on the size of the subpopulations. The overlapping of the subpopulations is another source of error. (Snidero et al., 2004) showed that the bias of the PiMLE increased impressively when a significant overlapping arose whereas the effect in the MLE is limited.

### 4.3.2 Sample size

Most of the NSUM estimators are unbiased if the degree estimation is perfect. However, this is not possible in most cases. Moreover, the NSUM estimators assume that the information of the ARD is independent, thus the participants are not connected. In practice, this condition may also not hold, affecting the estimates. Another issue is that there is a lack of guidelines about the sample size that the researchers should choose for their studies.

## 5 Simulation Study

We construct the four basic scenarios considering networks of  $N = 10000$  nodes to make the population as big as possible but computationally feasible. We chose a Watts-Strogatz network with 25 initial nodes connected to each side and the probability of changing the connection equal to 0.1. For the case of the preferential attachment network, we pick a network with a linear preferential attachment (power equal to one) and number of edges adding each time equal to 25 to get a similar average degree (49.935 compared to 50 of the Watts-Strogatz). For the distribution of the unknown subpopulation, we set 100 nodes at random in every case, and we implement a SIR process in each network. The first one is a SIR with  $\beta = 0.046$  and  $\gamma = 0.01$  where we choose the process at iteration 5. The other process, the SIR of the preferential attachment, has  $\beta = 0.039$  and  $\gamma = 0.010$ , and the SIR is chosen at iteration 2. We select parameters of the SIR model such that the size of the unknown subpopulations is 100 in order to compare with the uniform case.

Simulation goal	Visibility factor ( $\nu$ )	Memory factor ( $\rho$ )	Number known subp.	Sizes known subp.	Presence of disjoint subp.	Sample size	Section in the paper
Trans. error	[0,1]	0	20	Baseline	No	50	<a href="#">5.1.1</a>
Barrier error	1	[0,1]	20	Baseline	No	50	<a href="#">5.1.2</a>
Subpop. number	1	0	2-20	Even (big & small)	Yes/No	500	<a href="#">5.1.3</a>
Subpop.sizes	1	0	20	All (see Table 5)	Yes/No	500	<a href="#">5.1.4</a>
Sample size	1	0	20	Baseline	No	[25,3000]	<a href="#">5.1.5</a>

Table 3: Characteristics for each basic scenario.

The types of simulations that we conduct are summarized in Table 3. We vary one parameter as the other parameters remain fixed. The rows represent the family of simula-

tions. The parameter that is measured is shown in the first column. For the fifth column, the Baseline subpopulations refer to the configuration  $S_0$  of Table 5. These are the sizes of the subpopulations that we use except in those simulations where we observe the effect of the subpopulations. In the case of even subpopulations in the third row, it refers to subpopulations of sizes 0.1 and subpopulations of sizes 0.01. The rest of the subpopulation sizes employed can be found in Table 5. The sixth column represents the presence of cases with disjoint known subpopulations.

**Comparison between estimators** Some NSUM estimators require more information than others. In particular, the GNSUM estimator requires a sample from the unknown subpopulation, which is a lot of information compared to other estimators. The same is true for the adjusted estimators. Similarly, the RoA and AoR use the information on the personal network size of the participants. In the cases where the recall error is not present, the RoA and AoR have perfect knowledge of the size of the ego-network, and hence, the comparison with the rest of the estimators is not fair. However, the results of the RoA and AoR help to understand the behavior of other estimators closely related to them (see Table 1).

## 5.1 Simulations results

We show the performance of the NSUM estimators under the scenarios of Table 5.1.5. We present figures with boxplots and radar charts for the mean and MAE for some simulation parameters. In addition, we show tables in the supplementary material whose rows represent the value of the estimators for the mean, standard deviation, MAE, and MSE, and the column represents the parameters that it is being analyzed. The tables with all the parameters are included for a detailed comparison between the estimators. The Table 4 shows the location of the figures and tables. The Tables in the supplementary material appear under the reference S.number.

Simulation	W-U scenario	W-S scenario	P-U scenario	P-S scenario
Trans. error	Figures 1 and 2, and Tables S.2 and S.3	Tables S.4 and S.5	Tables S.6 and S.7	Tables S.8 and S.9.
Barrier error	Figure 3 and Table S.10	Figure 4 and Table S.11	Table S.12	Table S.13
Subpop. number	Figure 5 and Tables S.14 and S.15	Tables S.18 and S.19	Tables S.22 and S.23	Tables S.26 and S.27
Subpop. number small	Tables S.16 and S.17	Tables S.20, S.21	Tables S.24 and S.25	Tables S.28 and S.29
Subpop. sizes	Figure 6 and Tables S.30 and S.31	Tables S.32 and S.33	Tables S.34 and S.35	Figure 7 and Tables S.36 and S.37
Sample size	Figure 8 and Table S.38	Table S.39	Figure 9 and Table S.40	Table S.41

Table 4: Location of the tables in the article and supplementary material.

### 5.1.1 Transmission error simulations

We compare the NSUM estimators while the transmission error varies, which is modeled with the visibility factor  $\nu$ . We provide two figures showing the performance of the nine NSUM estimators (Figure 1) and the adjusted estimators (Figure 2) in the W-U scenario. These figures do not include the Overdispersed estimator because of the bad performance.

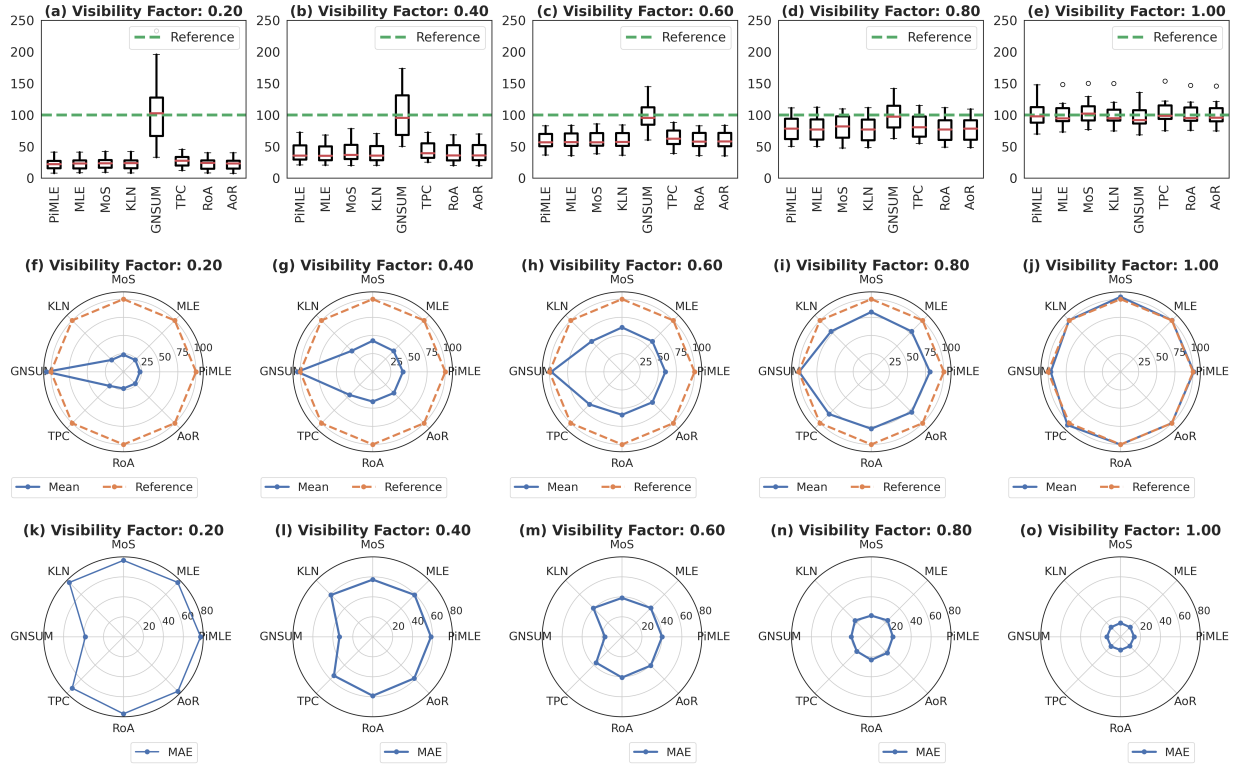


Figure 1: Results obtained by the standard methods varying the visibility factor in the W-U scenario. (a)-(e) Boxplots of the unknown subpopulation estimates across methods. (f)-(j) Mean of unknown subpopulation estimates. (k)-(o) MAE of the unknown subpopulation estimates.

We observe that the GNSUM is an estimator that works well for every scenario and

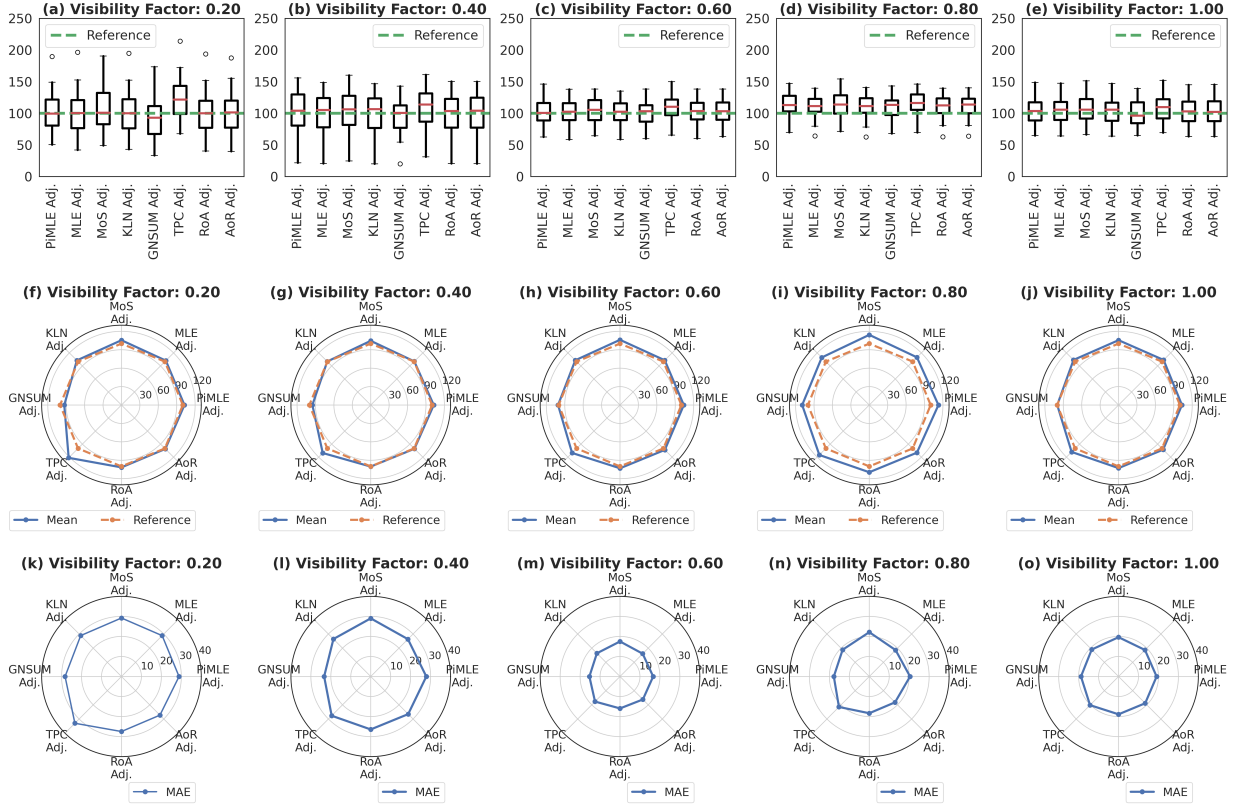


Figure 2: Results obtained by the GNSUM and the adjusted methods varying the visibility factor in the W-U scenario. The adjusted methods have been corrected by the inverse of the estimation of the transmission rate using the same sample of the unknown subpopulation as the GNSUM. (a)-(e) Boxplots of the unknown subpopulation estimates across methods. (f)-(j) Mean of unknown subpopulation estimates. (k)-(o) MAE of the unknown subpopulation estimates.

value of the visibility factor, especially when the visibility factor is greater than 0.2. This is observed in Figure 1 in the W-U scenario. In addition, we observe that the estimates of the GNSUM are good although the standard deviation is high. The standard deviation, MAE, and MSE decrease when the visibility factor increases, and therefore, when the transmission error decreases. The Overdispersed estimator performs badly for every value of  $\nu$  in each scenario. For the rest of the estimators, they have bad performance when the visibility is lower than 1. A special case is the P-S scenario, where the estimates are not good for the mean, MAE, and MSE, even when the visibility factor is 1, and therefore, there is no transmission error.

The adjusted estimators give similar values to the GNSUM, as depicted in Figure 2. As we see in Table 6, some adjusted estimators are better than the GNSUM in some scenarios. Additionally, we notice that the MAE and MSE decrease as the transmission error decreases. However, in the P-S scenario, the adjusted estimators do not produce reliable estimates.

### 5.1.2 Recall error simulations

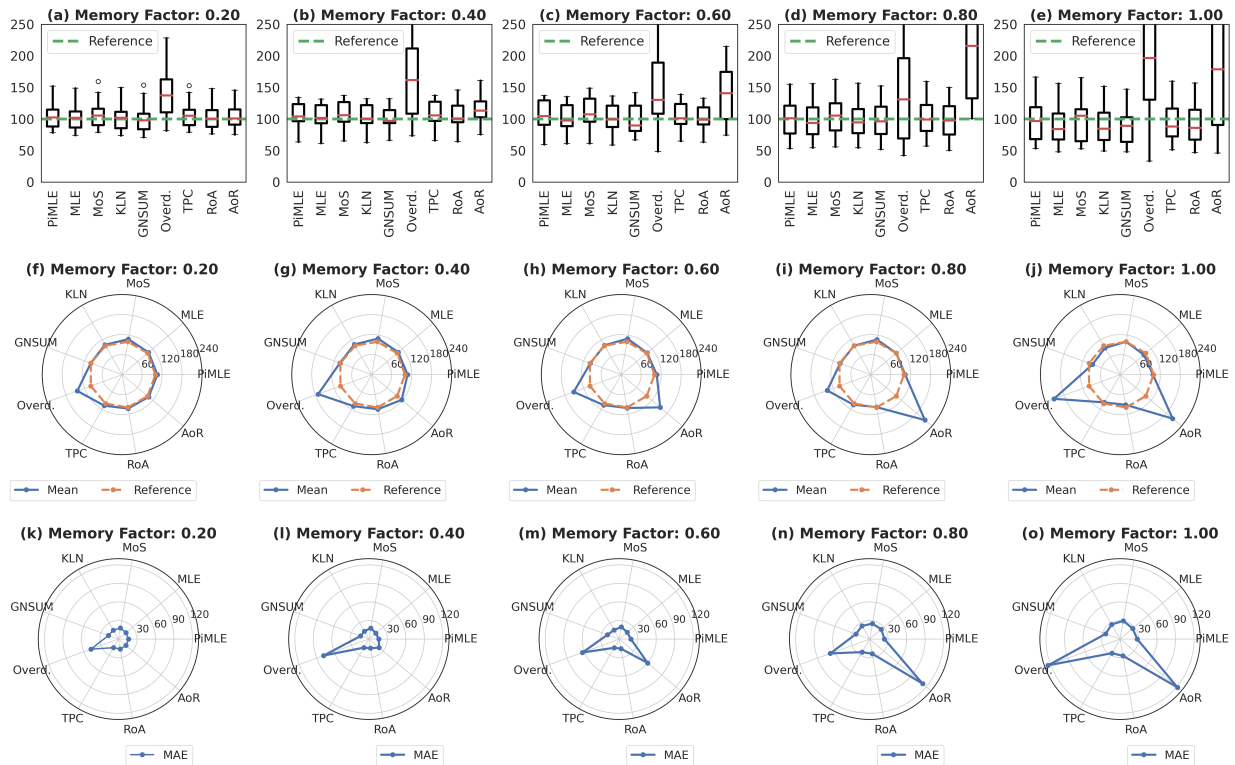


Figure 3: Results obtained varying the memory factor in the W-U scenario. (a)-(e) Box-plots of the unknown subpopulation estimates across methods. (f)-(j) Mean of unknown subpopulation estimates. (k)-(o) MAE of the unknown subpopulation estimates.

We show the results of the NSUM estimators in the W-U scenario while the recall error increases in Figure 3. This figure contains the boxplot and radar charts for the 9 NSUM estimators. We observe that the AoR is the most affected estimator by this bias. This estimator gives considerably higher estimates than the rest of the estimates with a low increment of the bias, around 0.3 for the mean (see Table S.10), and significantly higher



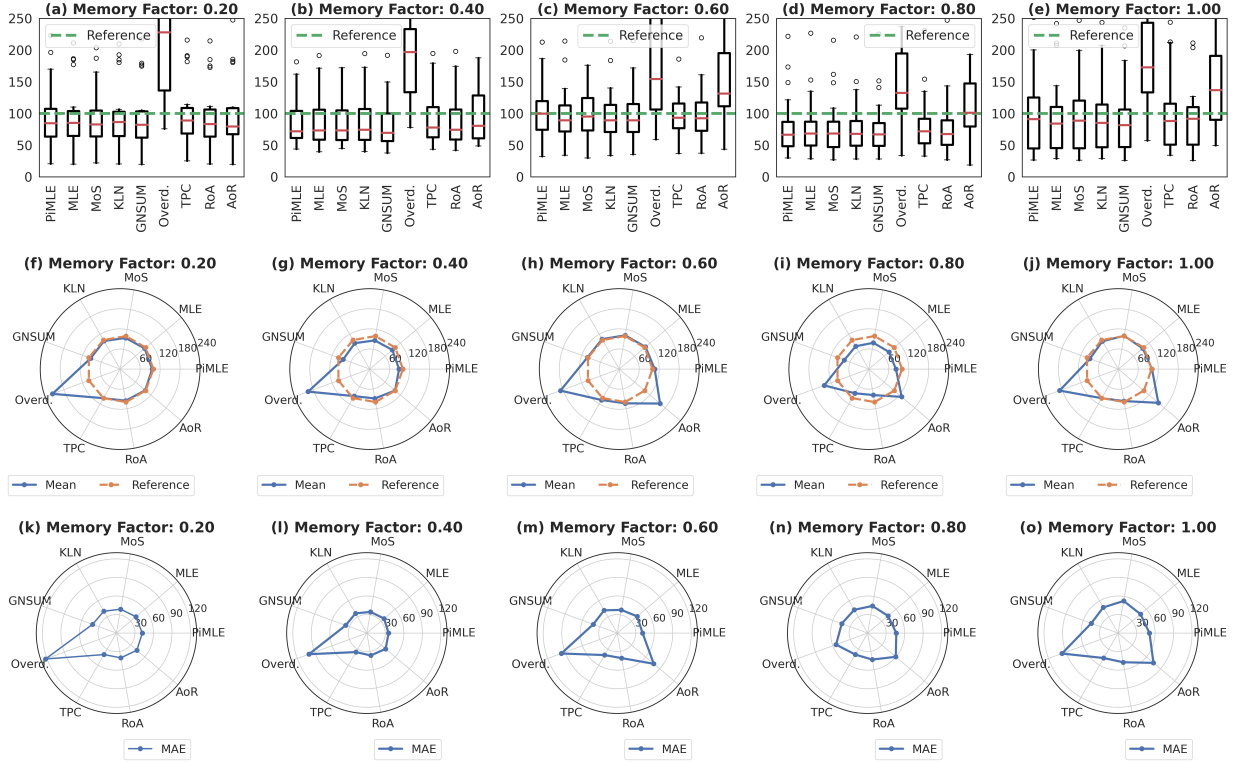


Figure 4: Results obtained varying the memory factor in the W-S scenario. (a)-(e) Box-plots of the unknown subpopulation estimates across methods. (f)-(j) Mean of unknown subpopulation estimates. (k)-(o) MAE of the unknown subpopulation estimates.

MAE and MSE for values of the memory factor above 0.4. For all the estimators, the MAE increases as the memory factor increases, as Figure 3 displays, but the errors are not as significant as the AoR and Overdispersed. Additionally, PiMLE and MoS have a moderate increase of the mean, MAE, and MSE as the memory factor increases, as we observe in Table S.10 for the values 0.6 and 0.8 of the memory factor. Additionally, as in the previous Section, the Overdispersed estimator gives bad estimates.

The case of the W-S scenario is in Figure 4 and Table S.11. We observe that the AoR has a high mean and MAE when the memory factor is greater than 0.5. Unlike the previous case, the other estimators except the Overdispersed are similar as the memory factor increases, being the MoS by far the best in terms of MAE and MSE as we see in Table 6.

In the P-U scenario, the results are similar, although the estimators tend to underestimate. The supplementary material shows the results in Table S.12. The AoR is the worst

as in the other scenarios; the Overdispersed provides worse metrics for the mean, MAE, and MSE. PiMLE and MoS are comparable to the other estimators but are more unstable when the recall error increases. Table S.12 displays that PiMLE and MoS have a high mean for a memory factor equal to 0.8, and high MAE and MSE for a memory factor of 0.9. The GNSUM is slightly worse but this behavior is constant across all the memory factor values.

In the P-S scenario all the estimators, except the GNSUM, have a bad performance, as we see in Table S.13. The AoR is again considerably worse when the memory factor increases, while the other estimators do not present significant differences.

### 5.1.3 Number of subpopulations simulations

We have calculated the NSUM estimates varying the number of known subpopulations used to estimate the degree. To illustrate this, we provide Figure 5 and Tables in the supplementary material (see Table 4). The figure includes a boxplot and radar charts for the mean and MAE in the W-U scenario with known subpopulations with overlapping. This figure does not include the Overdispersed estimator due to its significantly bad performance.

We do not observe significant differences when the number of subpopulations varies, as we can see in Figure 5 and the supplementary material (Tables S.14, S.18, S.22 and S.26). Most of the estimators are very close, except the Overdispersed. The PiMLE, MoS, and TPC slightly differ from the rest of the estimators when the number of known subpopulations is small, as we see for example in the case of two subpopulations in the W-U scenario in Table S.14. Additionally, the values are closer to the RoA and AoR, which do not use subpopulations to estimate the degree, and instead, in these simulations, they use the real personal network size. The variation of the RoA and AoR is then produced by randomness. We notice that the estimators are very similar in terms of all metrics with six or more subpopulations. The MAE and MSE decrease with the number of subpopulations, but the gains are not very significant. In addition, the MLE and KLN are closer to the RoA and AoR with four or more. Regarding the effect of the number of subpopulations when the subpopulations are small, we observe that the results are worse for the standard deviation, MAE, and MSE. The disjoint subpopulations improve the MAE and MSE in the case of medium sizes, but when the subpopulations are small there is no significant effect.

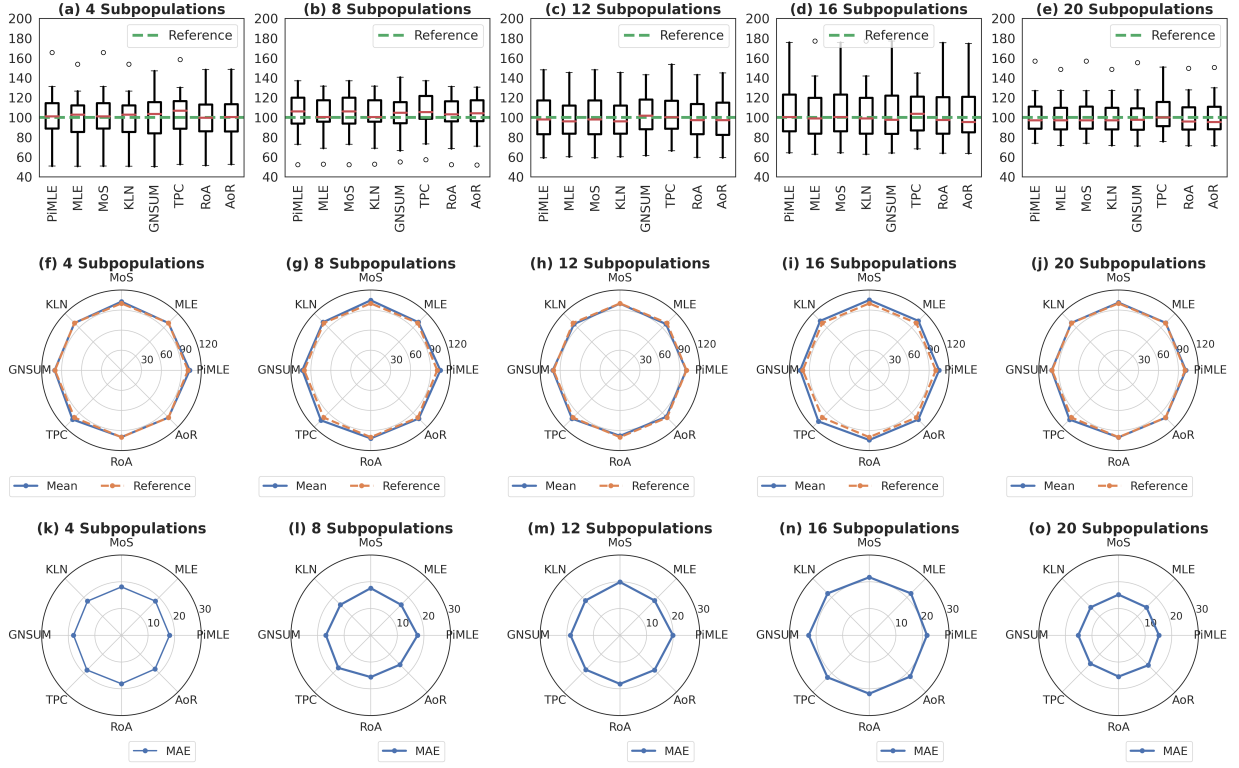


Figure 5: Results obtained varying the number of known subpopulations in the W-U scenario. Overlapping of subpopulations is included. (a)-(e) Boxplots of the unknown subpopulation estimates across methods. (f)-(j) Mean of unknown subpopulation estimates. (k)-(o) MAE of the unknown subpopulation estimates.

For the PiMLE and MoS, we observe that the estimators do not give values when the subpopulation sizes are small (see Tables S.16, S.20, S.24, S.28, S.17, S.21, S.25 and S.29). Indeed, the estimators need at least 12 subpopulations to obtain results (see Table S.16). The values of the metric in the few cases where they work are not good, with high mean MAE and MSE. When the subpopulations have bigger sizes, some cases do not give results with two subpopulations (see Tables S.22, S.23, S.26 S.27). This behavior is produced because some estimates for the degree are zero. Additionally, in the case of Watts-Strogatz network and SIR (see Tables S.18, S.19), with less than eight and four subpopulations (for the not disjoint and disjoint case, respectively) the PiMLE and MoS have high MAE and MSE, and mean above the real value. For the rest of the main scenarios and parameters, the values are similar to the other estimators.

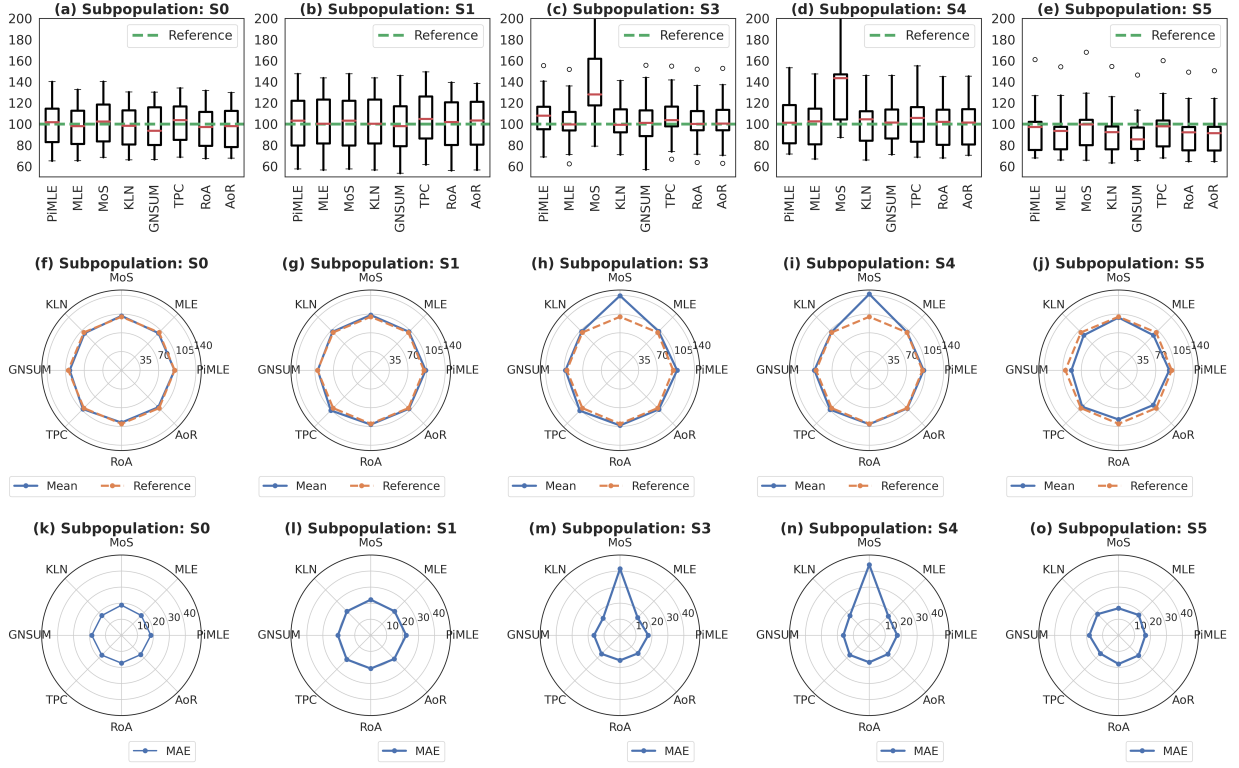


Figure 6: Results obtained for different subpopulations in the W-U scenario. (a)-(e) Boxplots of the unknown subpopulation estimates across methods. (f)-(j) Mean of unknown subpopulation estimates. (k)-(o) MAE of the unknown subpopulation estimates.

The Overdispersed estimator does not provide good results as in other simulations. In this case, increasing the number of subpopulations increases the bias, MAE, and MSE when the subpopulations are big. This effect does not occur in small subpopulations. Also, the values in this case are better than in other scenarios, although they are worse than the other NSUM estimators. Only the number of subpopulations equal to 6, 8, 12, and 16 gives similar values to the rest of the estimators in the P-U scenario.

#### 5.1.4 Sizes of subpopulations simulations

For the simulations of the sizes of the subpopulation, we use the subpopulations of Table 5. This group includes our baseline subpopulations, two even subpopulations, one uneven subpopulations, and other combinations where the size of groups is small, except one population of big size.

Subp.	Scenario	Sizes
$S_0$	Baseline	0.016, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02
		0.02, 0.04, 0.04, 0.04, 0.04, 0.04, 0.08, 0.08, 0.08, 0.16
$S_1$	Even	0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1
$S_2$	Even and small	0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01
$S_3$	One large, the rest small	0.3, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01
$S_4$	An even bigger one, the rest small	0.5, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01
$S_5$	Uneven	0.2, 0.2, 0.1, 0.1, 0.05, 0.05, 0.02, 0.02

Table 5: Relative subpopulation sizes employed in the simulations.

We show the results of the W-U and P-S scenarios in Figures 6 and 7, respectively. We excluded the subpopulations  $S_2$  in Table 5 since the PiMLE and MoS do not give values and the Overdispersed estimator due to its bad performance.

The MoS is the most affected by the different sizes, especially when the subpopulations have unequal sizes and there is a big difference between them, subpopulations  $S_3$ ,  $S_4$ . We observe this behavior in Figures 6 and 7, where the estimates differ considerably in those cases. However, the MoS is less affected if there are intermediate sizes between the bigger and the smaller as in  $S_5$ . As we can observe in Figures 6 and 7.

For the rest of the estimators, the values are close to each other, except in the case of a preferential attachment network and SIR, where only the GNUSM gets reliable estimates for all of the subpopulation configurations as Figure 7 shows. In addition, the values are similar to the estimators RoA and AoR, which use the real degree in this context. However, the PiMLE, TPC, and GNSUM are slightly affected by the known subpopulation configurations. PiMLE is affected by the unequal sizes, represented by subpopulations  $S_3$ ,  $S_5$ , but it is not affected by uneven population with one big subpopulation ( $S_4$ ), unlike the MoS. Additionally, PIMLE presents issues when the known subpopulations are small, represented in case  $S_2$  because of the zero-degree estimates. The TPC is not affected in the P-U scenario, where it is the best method (see Table 6) and is less affected than the PiMLE and GNSUM in the W-S scenario. The MLE and KLN shows to be very robust varying the sizes of the known subpopulations.

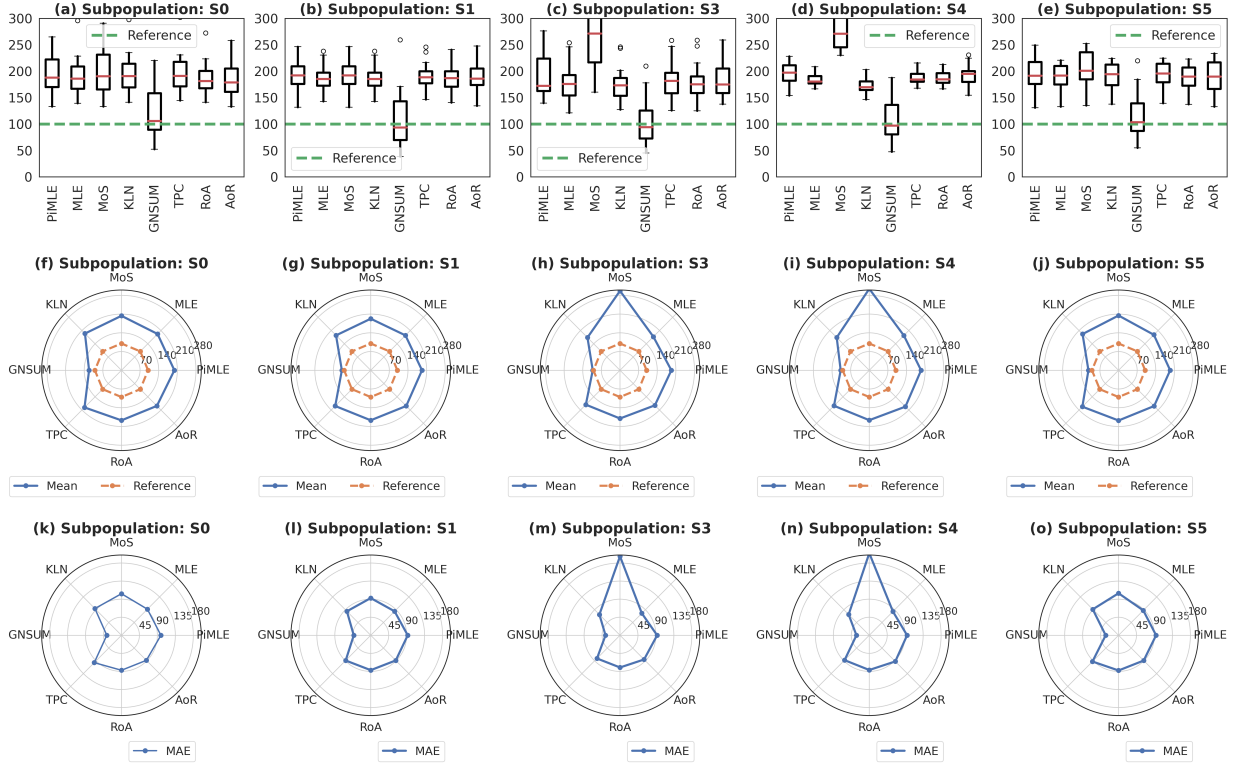


Figure 7: Results obtained for different subpopulations in the P-S scenario. (a)-(e) Boxplots of the unknown subpopulation estimates across methods. (f)-(j) Mean of unknown subpopulation estimates. (k)-(o) MAE of the unknown subpopulation estimates.

### 5.1.5 Sample size simulations

We evaluate the performance of the NSUM estimators as the sample size varies. We provide two figures with the boxplots and radar charts in the scenarios W-U and P-U. We include detailed tables with results in all the scenarios in the supplementary material.

We note in Figures 8 and 9 that the estimators reduce the error when the sample size increases. In the W-U scenario, all the estimators converge to the real value, as Figure 8 shows. The same behavior is observed when the unknown subpopulation is replaced in the W-S scenario in Table S.38. In the P-U scenario, Figure 9 shows that increasing the sample size improves the estimates, although PiMLE and MoS are the only estimators that converge to the real value, as noted in Table S.40. The GNSUM slightly overestimates while MLE, KLN, TPC, RoA, and AoR slightly underestimate. Additionally, Figure 9 and Table S.40 indicate that the GNUSM has high variance and MAE and MSE higher than the

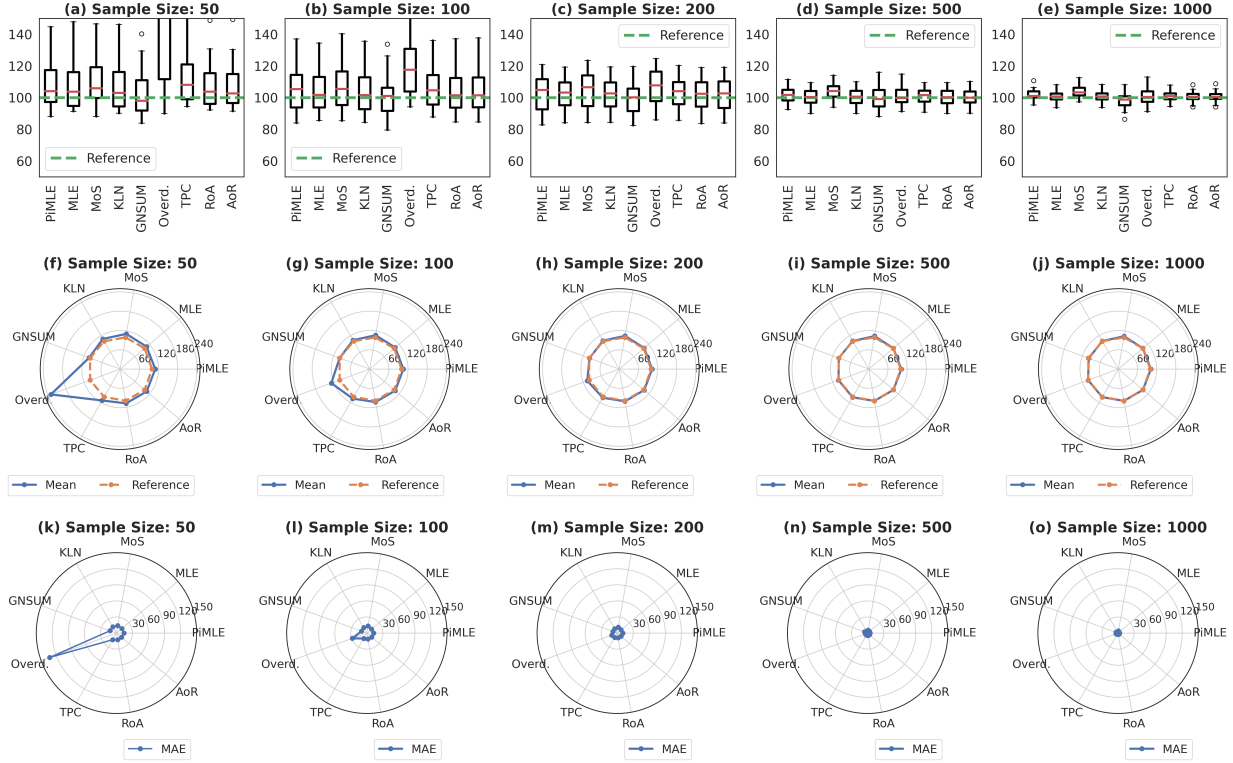


Figure 8: Results obtained varying the sample size in the W-U scenario. (a)-(e) Boxplots of the unknown subpopulation estimates across methods. (f)-(j) Mean of unknown subpopulation estimates. (k)-(o) MAE of the unknown subpopulation estimates.

other methods. In the last case, the P-S scenario, the sample size does not lead to better results, with GNSUM being the only estimator providing a reliable estimate, although with a slightly high mean.

The MAE and MSE are better as the sample increases. The MAE is below ten percent for a sample of 200 – 300 for the W-U scenario; 1000 for the W-S scenario; and 200 for the P-U scenario. In the P-S scenario, the errors are always bigger than ten percent.

In addition, we observe that the Overdispersed estimator needs more samples than the other NSUM estimators to obtain reliable estimates for the mean, and error similar to the other estimators, as is seen in Figures 8 and 9. It needs a sample between 100 in the W-U scenario, 200 in the P-U scenario, and 500 in the W-S scenario. This does not apply to the special case of the P-S scenario, where the estimates are far from the real value for every value of the sample size like the other estimators (except the GNSUM). In this case, the Overdispersed estimator is similar to the other estimators for a sample of 50.

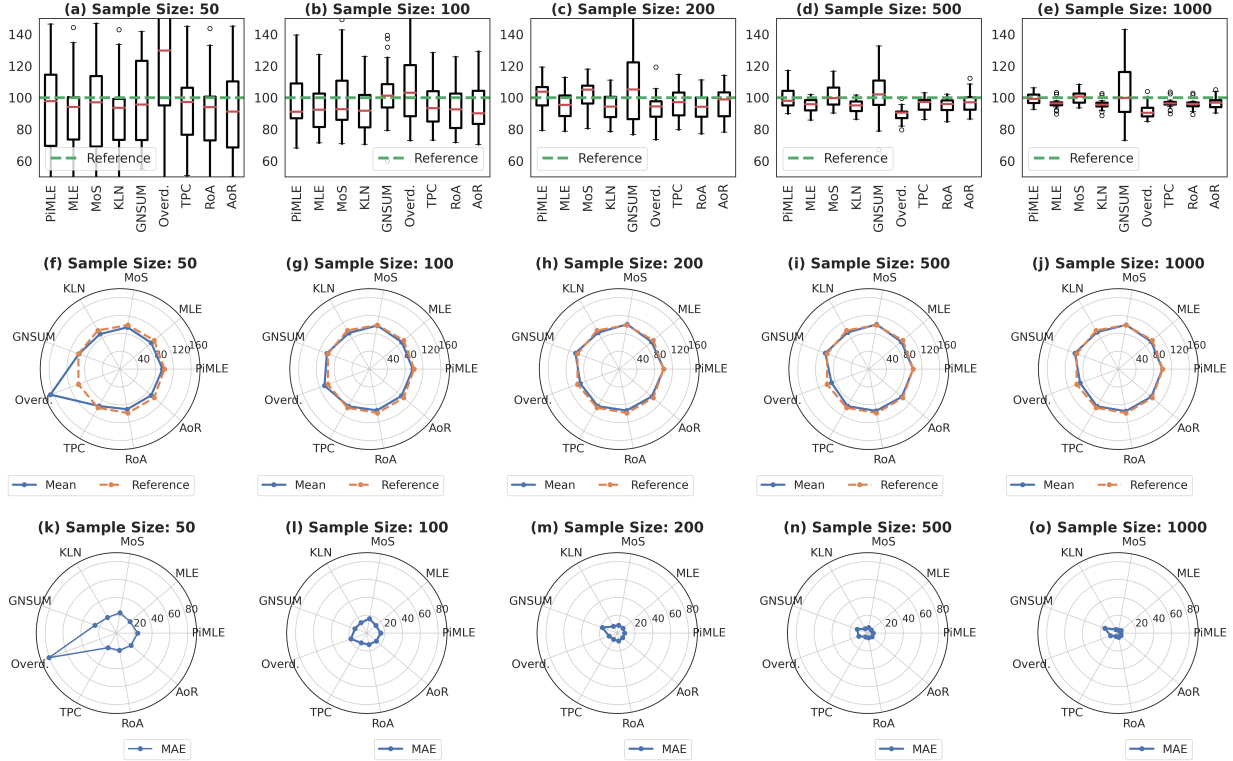


Figure 9: Results obtained varying the sample size in the P-U scenario. (a)-(e) Boxplots of the unknown subpopulation estimates across methods. (f)-(j) Mean of unknown subpopulation estimates. (k)-(o) MAE of the unknown subpopulation estimates.

## 5.2 Discussion

The simulation shows that the network topology, the unknown subpopulation distribution, and the different parameters of the simulation affect the results. Table 6 displays the best method in each simulation and scenario. This table does not include the estimators RoA and AoR in the simulations varying the number and sizes of the known subpopulations because they do not use the known subpopulations. We also include Tables S.45 and S.46 in the supplementary material, which provide the values of the best method for the subpopulation number and subpopulation sizes, both with and without including RoA and AoR. Moreover, the rows of the table corresponding to the number of subpopulations only include the case of subpopulations of medium sizes with overlapping. Additionally, we include in Figure 10 how the expectation of the ratios of the ARD of the unknown subpopulation by the real degree differs from Eq. (1).



Simulation goal	Main scenarios	Mean	MAE	MSE
Trans. error	W-U	GNSUM/100.64	GNSUM/24.78	GNSUM/1183.1
	W-S	MLE(adj)/98.27	MoS(adj)/44.08	MoS(adj)/3198.95
	P-U	PiMLE(adj)/100.74	KLN(adj)/24.98	KLN(adj)/1072.24
	P-S	GNSUM/112.20	GNSUM/39.89	GNSUM/3214.67
Recall error	W-U	MLE/100.35	MLE/18.38	MLE/546.82
	W-S	TPC/96.14	MoS/38.8	MoS/2337.73
	P-U	PiMLE/99.65	TPC/18.95	TPC/564.44
	P-S	GNSUM/118.82	GNSUM/38.55	GNSUM/2884.27
Subpop. number	W-U	MLE-KLN/100.00	MLE-KLN/18.36	MLE-KLN/530.82
	W-S	MLE-KLN/105.38	TPC/42.39	MLE-KLN/3036.94
	P-U	TPC/99.93	TPC/15.71	TPC/382.1
	P-S	GNSUM/115.58	GNSUM/ 37.52	GNSUM/ 2976.36
Subpop. sizes	W-U	MLE-GNSUM/99.80	MLE/18.34	KLN/501.14
	W-S	KLN/105.02	KLN/42.87	KLN/2768.76
	P-U	TPC/99.53	TPC/17.31	TPC/439.21
	P-S	GNSUM/111.43	GNSUM/ 36.52	GNSUM/ 2531.72
Sample size	W-U	GNSUM/99.06	RoA/8.29	GNSUM/181.61
	W-S	TPC/98.12	MoS/20.04	MoS/797.07
	P-U	MoS/100.32	TPC/9.36	TPC/183.81
	P-S	GNSUM/110.28	GNSUM/ 32.07	GNSUM/ 1959.8

Table 6: Best method for each scenario. For each metric, it shows the average of the values corresponding to each parameter. AoR and RoA are not included in the simulations about the number and sizes of known subpopulations because they do not use the known subpopulations.

**Effect of the network topology and unknown subpopulation distribution.** In the Watts-Strogatz case with uniform distribution of the unknown subpopulation, the W-U scenario, the MLE is usually the most robust, as we observe in Table 6. The Watts-Strogatz model is related to the Erdős–Rényi model, as it becomes the Erdős–Rényi model in the limit case when the probability of rewiring each edge is 1. The binomial model, in which MLE, KLN, and RoA are based, approximates to Erdős–Rényi model (Kunke et al., 2024). These models are especially good in this scenario.

In the W-S scenario, the estimates are worse due to the barrier effect since the probability of knowing someone in the unknown subpopulation depends on the position in the network. In this case, the unknown distribution is not distributed uniformly, and hence, it does not hold the conditions of the methods involving the binomial model for the size estimation. Despite this fact, these estimators, MLE, KLN, and RoA, have a good performance, compared to other methods. However, the main assumption of Eq 1 holds in this scenario, as seen in Figure 10, and the methods relying only on this equation, MoS, and PiMLE, have better performance.

In the P-U scenario the estimates are also worse than in the W-U scenario. Some higher-degree nodes are counted multiple times, slightly affecting the estimates in terms of MAE and leading to some underestimation. However, PiMLE and MoS use the ratios between the ARD responses for the unknown subpopulation and the degree estimated with the ARD of the subpopulation, which reduces the influence of higher-connected nodes. The TPC model assumes that  $y_{ij} = \lambda N_k$ , and introduces an additional parameter  $\alpha_i$  to account for the variability of the network. This seems to be particularly effective in this scenario. It performs best, especially in terms of MAE and MSE, as shown in Table 6.

In the P-S scenario, only the GNSUM produces reliable estimates. The SIR process produces that the high-degree nodes are more likely to belong to the unknown subpopulation. This means the higher degree nodes belonging to the unknown subpopulation nodes are counted multiple times. It produces higher ARD responses for the unknown subpopulations leading to overestimation. The GNSUM corrects this issue with additional information on the links of the unknown subpopulation through the sample of the unknown subpopulation.

The GNSUM estimator does not seem to depend on the network structure and performs well in every scenario. However, it requires significantly more information, including a sample of the unknown subpopulation. The Overdispersed estimator gives a bad performance in the simulations for every scenario. The overdispersion parameters for the propensity of subpopulation link may cause some overfitting as in our case the propensity of the known someone in the subpopulations is similar for each group.

**Effect of the transmission error.** The GNSUM estimator is the best when the transmission error is present, which is expected since it utilizes significantly more information. The GNSUM corrects bias using how visible the members of the unknown subpopulation are. For this, the GNSUM needs a sample of the unknown subpopulations and a battery of questions about the contacts of the sample, which diverges from the philosophy of preserving the privacy of the participants of the NSUM. The rest of the estimators underestimate when the transmission error is present. The values are proportional to the transmission error because the expected value of  $y_{iu}$  is  $\nu d_i N_u / N$  (Eq 9) when the transmission error is the same across all individuals in the network, as we observed in Section 4.2.1. Consequently,

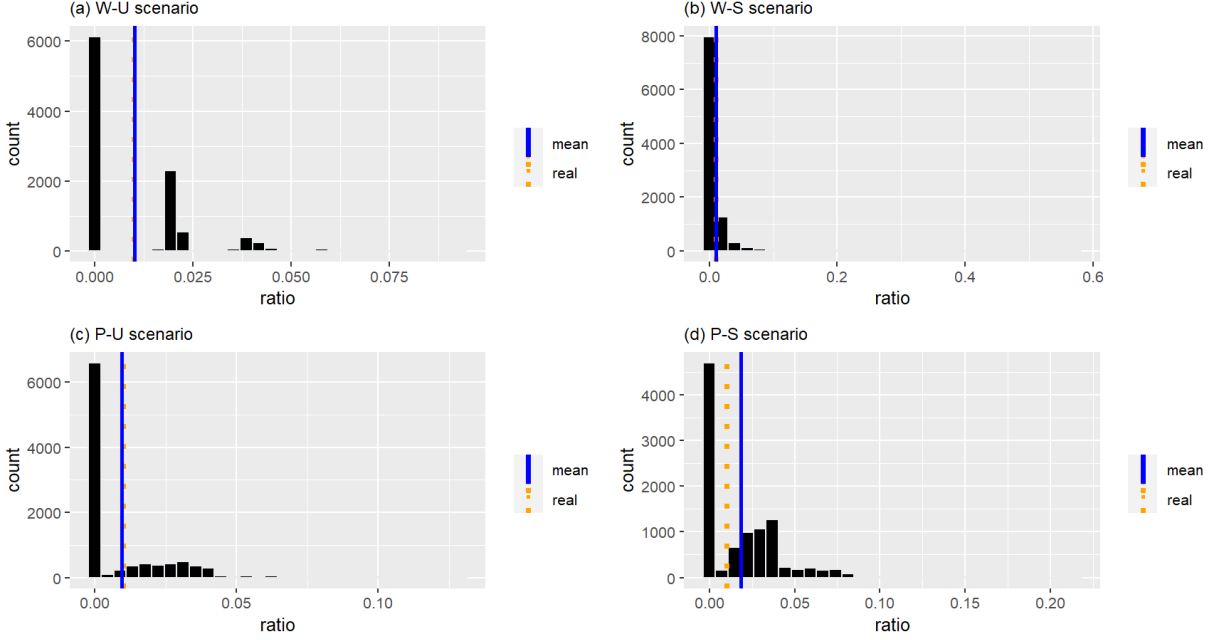


Figure 10: Histogram of the ratio of responses  $y_{iu}/d_i$  for the four scenarios.

the estimators adjusted by the inverse of the estimated  $\nu$  are close to the ground truth in most scenarios. The correction proves ineffective in the P-S scenario, unlike the GNSUM, which makes sense since the estimators without transmission error are inadequate. (Feehan, Salganik, 2016) prove that the GNSUM is equivalent to the correction by some factors of the MLE and RoA, provided that the estimated average degree in the denominators is accurate. However, this adjustment includes another factor, the degree ratio, the fraction of links of the unknown subpopulation relative to the number of links in the total population.

**Effect of the recall error.** The AoR is the most affected. Since the recall error applies both ARD of  $y_{iu}$  and the reported degree  $\delta_i$ , the ratios  $y_{iu}/\delta_i$  may be altered substantially if the recall error is high. PiMLE and MoS, which also use the average of the ratios to estimate the size (Eq. (4)), are affected to a lesser extent, being closer to the real value. Unlike the AoR, PiMLE and MoS use estimated degrees, which are obtained through linear combinations of the ARD responses of the known populations  $y_{ik}$ . As a result, the errors tend to cancel out, making the estimated degrees more accurate and improving the size estimation. MLE, KLN, and RoA use ratios of averages (see Table 1), and the errors of the two averages tend to cancel out. The GNUSM has a similar performance to MLE, KLN,

and RoA, since it is also expressed as the ratio of a linear combination of the ARD in the numerator, and a linear combination of the EARD in the denominator (see Eq. (??) and the supplementary material).

**Effect of the number of known subpopulations** The main conclusion of the variation of the number of known subpopulations is that the common number of subpopulations, 20, can be reduced. The estimators MLE and KLN are particularly stables. The randomness of choosing a person with contacts in a known subpopulation is compensated in the sample because these estimators only need to estimate the average degree (see Eq. (5) and Table 1). The GNSUM is the same replacing the estimated degree with the estimated relative visibility (see Section 3.3 and supplementary material), which is obtained by a linear combination of the EARD questions of the known subpopulations. MLE and KLN are very close with more than 4 subpopulations in every scenario. GNSUM and TPC are also stable, they need around 6 subpopulations to stabilize the MAE and have a very tight estimation, as we see in the supplementary material. When the subpopulations have small sizes, PiMLE and MoS present issues since they use the average of the ratios of the ARD response by the estimated degree. When the subpopulations have small sizes, the sample nodes may not be connected with any member of the subpopulations, and then, the estimated personal size is 0. Additionally, PiMLE and Mos only have good performance with small sizes in the W-U scenario with 20 subpopulations. Finally, the presence of disjoint known subpopulations improves the estimates, although it is not very significant.

**Effect of the sizes of the known subpopulations** The MoS is affected if the sizes of the known subpopulations are small since some degree estimates are zero (see Eq. (2) and 4). In the case of a combination of small subpopulations with a bigger one, MoS gives bad estimates. MoS estimates the participant’s degree by a linear combination of  $y_{ik}$  for the  $L$  subpopulations, where the small subpopulations are weighted more (Eq.(2)). Since it is less likely that a participant is linked with someone in the subpopulations, some personal sizes are underestimates, which causes the final overestimations in the subpopulations  $S_3$ ,  $S_4$  of Table 5. Subpopulations of intermediate sizes may compensate for this issue, as is the case of subpopulation  $S_5$ . PiMLE is also critically affected when the subpopulations are

all small for the same reason as the MoS. The cases of small subpopulations with one big subpopulations are considerably better, although they affect the estimates. This is because the degree of each participant is estimated by Eq. (3), which is the ratio of  $N \sum_{k=1}^L y_{ik}$  by  $\sum_{k=1}^L N_k$ , and therefore, the effect of the small subpopulation is more limited. The MLE and KLN use the estimated average degree in which the errors cancel out. The same occurs for the GNSUM because it uses the relative visibility of the unknown subpopulation. The TPC is affected although the estimates are closer to the real value. The TPC estimator assumes that the scaling is the same, with  $y_{ij} = \lambda N_j$ . Therefore, it is affected by unequal populations.

**Effect of the sample size.** We observe that increasing the sample size improves the estimation with the estimators yielding good results, especially in terms of MAE and MSE errors. This is particularly noticeable in the Overdispersed case, which did not yield good results with small samples. The overdispersion parameters for the propensity to form links with the subpopulations may cause some overfitting, as the propensity of the known someone in the subpopulations is similar for each group. Increasing the sample can reduce this overfitting.

Additionally, we observe that the estimators require more samples in the scenarios with more barrier error, W-S and P-U scenarios, than in the W-U scenario. The values of the mean become closer when the sample increases. However, while all the estimators converge in the W-U scenario, in the W-S and P-U scenarios most methods slightly underestimate. We conjecture that is because some nodes are counted multiple times and most of them do not belong to the unknown subpopulation. In the P-U scenario, the nodes with more links are considered multiple times, and most of them do not belong to the unknown subpopulation. There are three special estimators in this case: PiMLE, MoS, and GNSUM. PiMLE and MoS use the ratios between the ARD of the unknown subpopulation and the estimated degree. The estimated degrees use the ARD of the known subpopulation, so the higher connected nodes are less taken into account, see Eq. (2) and (3). We hypothesize that GNSUM is overestimated because the relative visibility is underestimated, as the unknown subpopulation nodes are mainly in the tail of the degree distribution.

## 6 Case Study

This section evaluates the performance of the NSUM estimators in a real network. Specifically, we use a friendship network extracted from Deezer, a popular music streaming platform, focusing on Croatia users. This dataset contains mutual friendship connections and users’ preferred music genres. This dataset was collected in November 2017 by (Rozemberczki et al., 2019) as a part of their study on graph embedding techniques and it is publicly available through the Stanford Large Network Dataset Collection: <https://snap.stanford.edu/data/gemsec-Deezer.html>. This dataset consists of 85 musical preferences for each user, with 54573 nodes and average degree of 18.26. We take the 21 musical preferences with more presence in the network. The genres and their relative size are in Table 7. We also use synthetic subpopulations with the same sizes generated with the procedures of Section 4.

Musical Genres		
Asian Music (0.91)	Country (4.89)	Folk (18.57)
Blues (5.37)	Dancefloor (9.27)	Hard Rock (9.12)
Chill Out (1.57)	Dancehall Ragga (1.29)	Indie Rock (5.95)
Classical (3.19)	Disco (5.55)	Jazz (4.91)
Comedy (2.22)	East Coast (2.92)	Kids (3.85)
Contemporary R B (13.21)	Film Scores (16.44)	Latin Music (12.69)
Contemporary Soul (2.50)	Films Games (17.36)	Metal (9.67)

Table 7: Popular musical genres and the percentage of users for the Deezer dataset.

This section is organized as follows. First, we evaluate the NSUM estimators simulating subpopulations uniformly across the real network. Next, we present a scenario where an unknown subpopulation is distributed through a SIR process over the network. Subsequently, we estimate some musical preferences of the Deezer dataset using other musical preferences as known subpopulations. Finally, we discuss the results for all the cases.

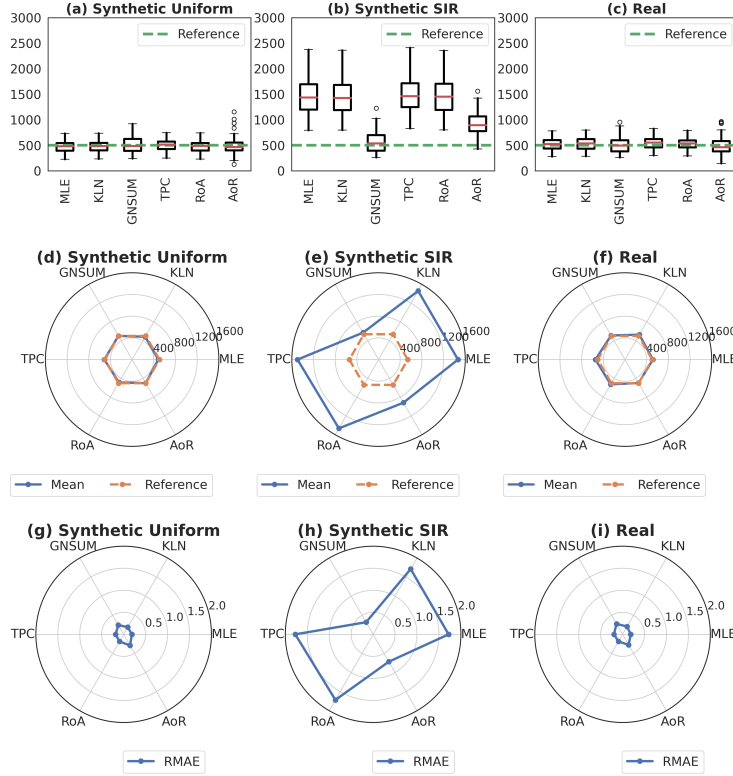


Figure 11: Results obtained from a real network with a real unknown subpopulation (Real), a synthetic unknown subpopulation uniformly distributed (Synthetic Uniform), and a synthetic unknown subpopulation using a SIR model (Synthetic SIR). (a)-(c) Boxplots of the estimates. (d)-(f) Mean of estimates. (g)-(i) RMAE estimates.

## 6.1 Real network with subpopulations distributed uniformly

We use the Deezer dataset as the underlying network and simulate the subpopulations uniformly using the sizes of Table 7. We use synthetic subpopulations with sizes corresponding to Asian Music, Blues, and Chillout as the unknown subpopulations. We evaluate the results of 50 estimates, taking samples of 100 nodes and other samples of 30 nodes for the unknown subpopulation. We provide a boxplot, mean, and relative mean absolute error (RMAE) of the estimates using Asian Music as the unknown subpopulation in Figure 11. In Table S.42 of the supplementary material, we include the values for the next metrics: mean, standard deviation, RMAE, and mean squared relative error (MSRE).

We observe in Figure 11 that MLE, KLN, GNSUM, TPC, AoR, and RoA are close to the real values. However, we observe an underestimation in these estimators except

GNSUM and AoR. Additionally, we observe more variance for the GNSUM and AoR. PiMLE and MoS do not give values because some individuals do not have contacts in any subpopulation, producing degree estimates with zero values. The Overdispersed has poor performance in the cases of Asian Music and Chillout, but its estimates are good in the case of Blues, as seen in Table [S.42](#).

## 6.2 Real network with SIR for the unknown subpopulation

We generate the known subpopulations as in the previous case, but we also simulate an unknown subpopulation using a SIR process. The synthetic subpopulations have sizes relatively similar to the real data. We evaluate the performance of the estimators for 50 iterations. In each iteration, we use samples of 100 from the entire network and 30 from the subset of nodes in the unknown subpopulation. We include a boxplot and two radar charts for the mean and RMAE in Figure [11](#) for the case of synthetic Asian Music. Values of the mean, standard deviation, RMAE, and MSRE are included in Tables [S.43](#).

In this case, the GNSUM is the only estimator that provides reliable estimates, as observed in Figure [11](#). The other estimators present a very significant overestimation. AoR is less affected but has a high error and is far from the real value. Since the subpopulations have small sizes, PiMLE and MoS do not produce results, since they estimate some degrees as zero. The other two cases are similar, as Table [S.43](#) shows.

## 6.3 Real network with real subpopulations

We use the 21 most popular Deezer music genres, displayed in Table [7](#). We select one of the musical genres as the unknown subpopulation and the rest as the known subpopulations. We apply this three times, using as known subpopulation the musical genres Asian Music, Blues, and Chill Out. We provide the results for 50 estimates of the subpopulations using samples of 30 from the unknown subpopulation and 100 from the entire network. We include in Figure [12](#) the boxplot for the three cases and radar charts for the mean and RMAE. In the supplementary material, we provide the mean, standard deviation, RMAE, and MSE values in Table [S.44](#).

We observe in Figure [12](#) that the distribution of the subpopulations affects the estima-



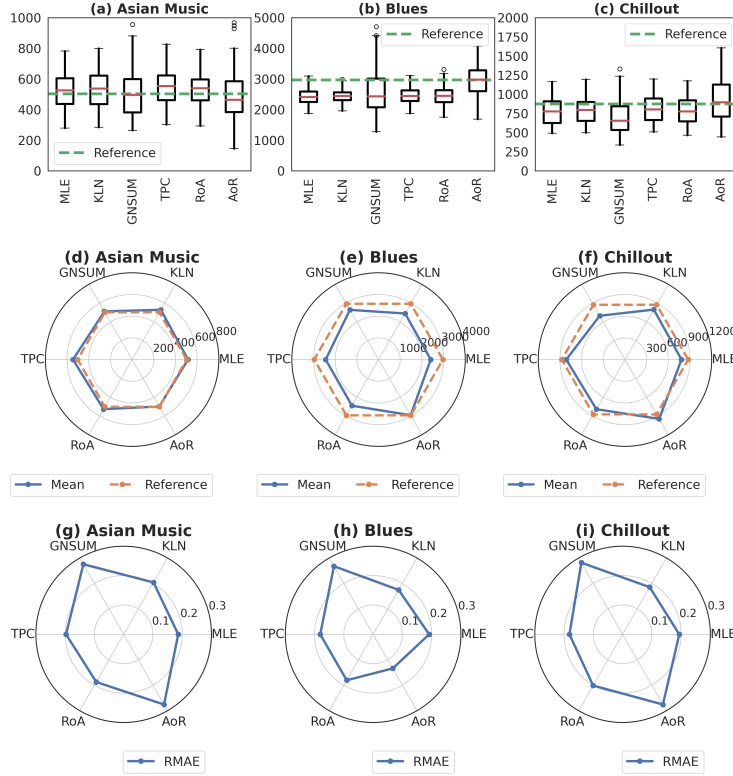


Figure 12: Results obtained from a real network with real subpopulations. (a)-(c) Boxplots of the unknown subpopulation estimates. (d)-(f) Mean of unknown subpopulation estimates. (g)-(i) RMAE of the unknown subpopulation estimates.

tors. While the estimates are close in the case of Asian Music, MLE, KLN, RoA, and TPC are affected in terms of mean estimating Blues. Additionally, GNSUM underestimates the Chillout subpopulations. We note that AoR is more stable in terms of the mean, but the RMAE is high. The other estimates are stable for this metric, although the GNUSM also has a high error. PiMLE and MoS present issues since some degrees estimates are zero. The Overdispersed estimator provides reliable estimates in the case of Blues, but significantly overestimates in the other two cases, as seen in Table S.44.

## 6.4 Discussion

We observe that the results of the real network and synthetic subpopulations validate our conclusions of Section 5.2. The real network is similar to the preferential attachment network since its degree distribution is exponential. The case of subpopulations simulated

with subpopulations distributed uniformly, and the case of SIR for the unknown subpopulation are similar to the P-U and P-S scenarios, and the performance of the estimators is comparable with these scenarios. One difference with our simulations is the bad performance of the PiMLE and MoS due to the small relative size of the subpopulations, which is consistent with our analysis of the sizes of the known subpopulations. The other difference is that the AoR is slightly better. We guess that one factor is the lower average degree, decreasing the frequency with which multiple nodes are reported.

The real network with real subpopulations of the Deezer dataset shows that the distribution of the subpopulations affects the performance. Nodes with similar preferences tend to be connected with each other, increasing the barrier effect. The difference with the P-S scenario is that concentration is not produced in the higher-degree nodes for the selected real subpopulations. This is similar to the P-U scenario, underestimating whether the subpopulations are connected to the higher-degree nodes. For the preference Blues, the degree of the individuals in the subpopulation is lower than the average degree. This condition affects the NSUM as (Feehan, Salganik, 2016) shows. However, this does not occur for the preference Chillout, in which there are deviations of the mean. In the P-U scenario, PiMLE and MoS are the best because they use the average of the ratios, Eq. (4), using the estimated degrees through subpopulations concentrated mainly in the tail of the distribution. In the real data, PiMLE and MoS do not give results because the subpopulations are small. The AoR, which uses Eq. (4) with the real degrees, is better than the other methods compared to the P-U scenario. We conjecture that it is produced because the network is bigger and the average degree lower, reducing the impact of the higher degree nodes in the ratios. In addition, AoR and GNSUM are the most reliable in the real case although the variance, RMAE, and MSRE are high in some cases (see Figure 12 and Table S.44). For the AoR, the behavior is caused by the variation in the proportion of contacts in some groups, and the effect on the GNSUM is likely to be caused by the small sample.

## 7 Conclusion

This work compares via simulation most of the NSUM estimators in networks with 10000 nodes and some properties found in real networks: a small world network (a Watts-Strogatz estimator) and a scale-free network obtained by a linear preferential attachment process. In each of them, we conduct two examples of the unknown subpopulation. In the first case, the unknown subpopulation is obtained uniformly and independent of the network. The second case is an unknown subpopulation distributed by a SIR process through the network, which is an example of the presence of the barrier effect. In these cases, the paper studies how the NSUM estimators vary as two biases (transmission and recall error) increase, how the number and sizes of the subpopulation affect the estimates, and the effect of the sample size. Additionally, we evaluate the difference in the performance of the estimators with real data. Concretely, we use a real network with real subpopulations and synthetic subpopulations simulated with the same procedures as the simulations.

The main findings comparing the NSUM estimators across the scenarios are as follows: (1) the number of subpopulations does not affect the estimators, except for the Overdispersed, PiMLE, and MoS; and for the last two, only six can be employed if the subpopulations are large, (2) the network topology has an impact on the results, the best method depends on the network, and in the preferential attachment networks, some estimators are slightly biased or give unreliable results, (3) dependent responses obtained by reporting the same node multiple times do not have an impact on the estimates, (4) in the presence of the recall error, the AoR is very unstable, and the PiMLE and MoS are slightly unstable in some cases when the recall error is high, (5) the Overdispersed estimator needs more sample to obtain more reliable estimates than the other estimators, and the errors increase with more subpopulations, (6) the adjustment by an estimation of the transmission error is similar than using the GNSUM in some cases, and some corrected estimators are better; but in the the case of a SIR process and preferential attachment the network, the results, with a certain transmission rate estimator, are considerably worse respect to the GNSUM, indicating the necessity for additional adjustment.

This work has some limitations. The first limitation is that the analysis is based on a simulated environment, and therefore, there may be some differences in real cases. More

concretely, the known subpopulations are assumed to be independent and uniformly distributed along the network, when in most real cases this does not happen. Also, we assume that the underlying network is an undirected graph, which means that the links between the individuals are reciprocal. Another limitation is that we use large dimensional matrices to calculate the reported values after the transmission and barrier error. This leads to high computational cost. For that reason, we use a few examples of networks and unknown subpopulation distribution, and one instance for each parameter in every simulation.

The NSUM methods have not been tested in many datasets. This is the first study that compares many of the estimators for the first time. Additionally, due to the computational cost, few examples were added. The first further work is to compare the estimators in more examples of network topologies and distribution of the subpopulations and the unknown subpopulation. Since the epidemiological SIR processes usually do not occur in the contact network, further work is to analyze the case when the unknown subpopulation is propagated through another network. Another future work is the study of the effect of the demographic characteristic to be able to compare new estimators such as (McCormick, Zheng, 2012), that take these variables into account. Because the sample size depends on the network topology and the distribution of the unknown subpopulation, the study of the effective sample size needs further development. Another future work is the correction of the NSUM estimators under other biases since we observe that the transmission error correction is insufficient when the barrier effect is present. It is also worth analyzing another bias arising from malicious responses, which is known as response bias, analyzing the performance of the estimators under different types of contamination. Another subsequent line of research is the analysis of the effect of the connection between the subpopulations produced in real cases. Finally, some estimators that are not frequently employed are better in some scenarios, further research may help to select the best method.

## Code and Data Availability Statement

The data and code are publicly available at <https://github.com/SDiazAranda/NSUM-Simulations>

## Declaration of interest statement

We have no conflicts of interest to disclose.

## SUPPLEMENTARY MATERIAL

**Supplementary Material to “Performance Analysis of NSUM Estimators in Social-Network Topologies”:** This supplementary material contains more exhaustive details on the estimators and network topologies. Moreover, it includes the tables with all the simulations and metrics.

## References

- Bernard H Russell, Johnsen Eugene C, Killworth Peter D, Robinson Scott. Estimating the Size of an Average Personal Network and of an Event Subpopulation // In *The Small World*. 1989. 159–175.
- Díaz-Aranda Sergio, Ramírez Juan Marcos, Daga Mohit, Champati Jaya Prakash, Aguilar José, Lillo Rosa Elvira, Anta Antonio Fernández. Error Bounds for the Network Scale-Up Method // arXiv preprint arXiv:2407.10640. 2024.
- Ezoe Satoshi, Morooka Takeo, Noda Tatsuya, Sabin Miriam Lewis, Koike Soichi. Population size estimation of men who have sex with men through the network scale-up method in Japan // *PloS one*. 2012. 7, 1. e31184.
- Feehan Dennis M, Salganik Matthew J. Generalizing the network scale-up method: a new estimator for the size of hidden populations // *Sociological methodology*. 2016. 46, 1. 153–186.
- Garcia-Agundez Augusto, Ojo Oluwasegun, Hernández-Roig Harold A, Baquero Carlos, Frey Davide, Georgiou Chryssis, Goessens Mathieu, Lillo Rosa E, Menezes Raquel, Nicolaou Nicolas, others. Estimating the COVID-19 prevalence in Spain with indirect reporting via open surveys // *Frontiers in Public Health*. 2021. 9. 658544.
- Habecker Patrick, Dombrowski Kirk, Khan Bilal. Improving the network scale-up estimator: Incorporating means of sums, recursive back estimation, and sampling weights // *PloS one*. 2015. 10, 12. e0143406.

- Jing L, Lu Q, Cui Y, Yu H, Wang T. Combining the randomized response technique and the network scale-up method to estimate the female sex worker population size: an exploratory study // *Public Health*. 2018. 160. 81–86.
- Killworth Peter D, Johnsen Eugene C, McCarty Christopher, Shelley Gene Ann, Bernard H Russell. A social network approach to estimating seroprevalence in the United States // *Social networks*. 1998a. 20, 1. 23–50.
- Killworth Peter D, McCarty Christopher, Bernard H Russell, Shelley Gene Ann, Johnsen Eugene C. Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach // *Evaluation review*. 1998b. 22, 2. 289–308.
- Kunke Jessica P, Laga Ian, Niu Xiaoyue, McCormick Tyler H. Comparing the Robustness of Simple Network Scale-Up Method Estimators // *Sociological Methodology*. 2024. 00811750241242791.
- Laga Ian, Bao Le, Niu Xiaoyue. Thirty years of the network scale-up method // *Journal of the American Statistical Association*. 2021. 116, 535. 1548–1559.
- Laga Ian, Bao Le, Niu Xiaoyue. A Correlated Network Scale-up Model: Finding the Connection Between Subpopulations // *Journal of the American Statistical Association*. 2023. 1–18.
- Maltiel Rachael, Raftery Adrian E, McCormick Tyler H, Baraff Aaron J. Estimating population size using the network scale up method // *The annals of applied statistics*. 2015. 9, 3. 1247.
- McCormick Tyler H. The network scale-up method // *The Oxford Handbook of Social Networks*. 2020. 153.
- McCormick Tyler H, Zheng Tian. Latent demographic profile estimation in hard-to-reach groups // *The annals of applied statistics*. 2012. 6, 4. 1795.
- Newman Mark. *Networks*. 2018.
- Ocagli Honoria, Azzolina Danila, Lorenzoni Giulia, Gallipoli Silvia, Martinato Matteo, Acar Aslihan S, Berchiolla Paola, Gregori Dario, Group INCIDENT Study, others . Using Social Networks to Estimate the Number of COVID-19 Cases: The Incident (Hidden COVID-19 Cases Network Estimation) Study Protocol // *International journal of environmental research and public health*. 2021. 18, 11. 5713.

- Rozemberczki Benedek, Davies Ryan, Sarkar Rik, Sutton Charles. GEMSEC: Graph Embedding with Self Clustering // Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2019. 2019. 65–72.
- Salganik MJ, Fazito D, Mello MB, others . Estimating the number of heavy drug users in Curitiba, Brazil using multiple methods. 2010.
- Salganik Matthew J, Fazito Dimitri, Bertoni Neilane, Abdo Alexandre H, Mello Maeve B, Bastos Francisco I. 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. 2011. 174, 10. 1190–1196.
- Snidero Silvia, Corradetti Roberto, Gregori Dario. The network scale-up method: A simulation study in case of overlapping sub-populations // Advances in Methodology and Statistics. 2004. 1, 2. 395–405.
- Srivastava Ajitesh, Ramirez Juan Marcos, Díaz-Aranda Sergio, Aguilar Jose, Anta Antonio Fernández, Ortega Antonio, Lillo Rosa Elvira. Nowcasting Temporal Trends Using Indirect Surveys // Proceedings of the AAAI Conference on Artificial Intelligence. 38, 20. 2024. 22359–22367.
- Sully Elizabeth, Giorgio Margaret, Anjur-Dietrich Selena. Estimating abortion incidence using the network scale-up method // Demographic Research. 2020. 43. 1651–1684.
- Teo Alvin Kuo Jing, Prem Kiesha, Chen Mark IC, Roellin Adrian, Wong Mee Lian, La Hanh Hao, Cook Alex R. Estimating the size of key populations for HIV in Singapore using the network scale-up method // Sexually transmitted infections. 2019. 95, 8. 602–607.
- UNAIDS WHO. Guidelines on estimating the size of populations most at risk to HIV // Geneva, Switzerland: World Health Organization. 2010. 51.
- Verdery Ashton M, Weir Sharon, Reynolds Zahra, Mulholland Grace, Edwards Jessie K. Estimating hidden population sizes with venue-based sampling: extensions of the generalized network scale-up estimator // Epidemiology (Cambridge, Mass.). 2019. 30, 6. 901.
- Zheng Tian, Salganik Matthew J, Gelman Andrew. 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. 2006. 101, 474. 409–423.