

Robust Network Scale-up Method Estimators

ARTICLE INFO

Keywords:

Network Scale-up Method
Aggregated relational data
Robust estimators
Adaptive procedures
M-estimators

ABSTRACT

The Network Scale-up Method (NSUM) is a relatively recent statistical approach for estimating the prevalence of unknown populations through indirect surveys utilizing information about the respondents' social circles. The popularity of NSUM has increased in recent years due to its ability to uphold privacy and cost-effectiveness. However, the NSUM is not exempt from biases resulting from participants' behavior. In addition, the simpler and most popular NSUM estimators are based on averages, making them sensitive to deviations in the samples, which may cause significant errors. This work aims to study how robust procedures can overcome misreporting, contamination, and deviation due to conditions such as barrier effects, prevalence, skewness, and tail length. Specifically, the central objective of the article is to analyze the statistical robustness of NSUM methods, studying whether these methods are affected by outliers or unusual data. We employ eight robust proposals for each of the two classical NSUM estimators. We analyze robust estimators through simulation experiments using synthetic random networks such as Erdős-Rényi, Scale Free, and Stochastic Block Model structures to model different degree distributions and community structures with different prevalence levels in contaminated and uncontaminated scenarios. We compare the results of the simulations with real data on COVID-19 indicators in the United Kingdom and voting intention in the Spanish General Elections of 2023. This article shows that the classical NSUM estimators perform poorly in contaminated scenarios, while most of the robust proposals are not considerably affected. However, the performance of some robust NSUM estimators decreases under barrier effects. In addition, we observe that distortions created by small prevalence play an important role in selecting the most suitable robust NSUM estimator. Particularly, the robustification of the Mean of Ratios (MoR) estimator based on the Myriad operator typically exhibits the best performance (for MoR methods) across the various social network structures for different prevalence levels, reducing the estimation error regarding the non-robust methods by up to three orders of magnitude in contaminated scenarios.

1. Introduction

The *Network Scale-up Methods* (NSUM) are methods to estimate population prevalences based on information collected from individual social networks (Laga et al., 2021). Specifically, NSUM uses the number of connections that a respondent reports within specific groups during a survey, known as *aggregated relational data* (ARD) (McCormick, 2020), to estimate the prevalence or size of these groups. Since NSUM do not require personal information from participants, they have been widely applied to estimate the size of sensitive groups, where disclosing the membership to a group may pose a potential threat, such as sex workers (Jing et al., 2018), drug users (Laga et al., 2023b), child traffickers (Nyarko-Agyei et al., 2024), people who had abortions (Sully et al., 2020), people with COVID-19 (Garcia-Agundez et al., 2021), or people with HIV (Laga et al., 2023a).

During recent years, there have been many improvements in the classical NSUM (Killworth et al., 1998a,b), as noted in the review of Laga et al. (2021). However, no method solves the lack of robustness of classical methods, which use sample means and may be drastically affected by the inclusion of outliers (Hampel et al., 2011). Some studies, such as Garcia-Agundez et al. (2021) and Ramirez et al. (2023), use data preprocessing to mitigate this issue. However, Garcia-Agundez et al. (2021) employ non-robust filtering techniques, while the filtering based on Medcouple used by Ramirez et al. (2023) is overly aggressive, removing up to half of the data. In addition, estimators addressing

robustness concerns have only been proposed in two articles. However, these works only focus on unbiased estimates under specific biases that may affect the estimators. Feehan and Salganik (2016) try to improve the NSUM estimates when the respondents are unaware of their acquaintances' membership and the propensities of forming ties differ depending on the participants. However, this method is also based on sample means and requires significantly more information—specifically, access to a sample from the unknown group. Maltiel et al. (2015) also consider these biases, but Díaz-Aranda et al. (2024) show that it is very sensitive to deviations. Beyond biases, survey data may have issues due to outlying responses, frequently due to malicious participants.

The main contribution of this work is the proposal of several robust NSUM estimators, eight estimators for each classical method, to mitigate these problems. Each of these estimators has been evaluated for its robustness to different types of data deviations, including contamination and distributions with skew and tail length. To our knowledge, this is the first work that proposes robust NSUM procedures within the framework of robust statistics.

A second contribution is the empirical evaluation of the robustness of the estimators in a simulated environment. The simulated social networks are similar to those found in real populations. In this case, the objective is to study how NSUM techniques resist outliers or unusual data, as defined in robust statistics—an aspect that has not been studied in the literature for NSUM techniques. The aim is to study the quality of the results when the data are imperfect or unusual. To this end, we have used the notion of contamination,

ORCID(s):

which allows us to analyze their behavior for various types of deviating data.

Additionally, we validate the performance in real datasets on COVID-19 and voting intention, representing cases of outlying values due to malicious responses and different propensities of forming links, respectively. In addition, we also assess the performance of the NSUM under the classic bias of the NSUM literature: barrier effect, response bias, recall error, and transmission error, as some are equivalent to a few of the cases studied in our contamination (see the Supplementary Material). In particular, to the best of our knowledge, this is the first study analyzing the effect of the response bias and proposing alternatives to mitigate it.

This article is organized as follows. Section 2 presents the NSUM estimators and their robustness properties. In Section 3, we present the robust proposal for each estimator. Section 4 displays the simulations conducted and a discussion about their performance. Section 5 assesses the results on real datasets. The final remarks are in Section 6.

2. Robustness and NSUM

The NSUM use the personal network size to scale the ARD of the social circles of the participants. Traditionally, NSUM utilize groups with known prevalences to first estimate the personal network size of the participants, usually denoted as the *participants' degree*. In our approach, we assume that participants directly report their degrees and that these reports are accurate. In addition, we include another ARD question about a group whose prevalence is unknown. Let d_i represent the reported degree of participant i and y_i the ARD for the unknown group, with the prevalence of this group denoted by ρ . Additionally, we denote by I_n the set of responses obtained by a survey of size n .

The classical NSUM methods are the frequentist estimators, firstly introduced in Killworth et al. (1998a,b); Habecker et al. (2015) (see Laga et al. (2021) for a review of the NSUM estimators). The main idea of the NSUM estimators is that the ratios y_i/d_i possess representative information of the prevalence about the estimators. Under our assumptions there are two classic estimators:

- The *Mean of Ratios* (MoR) estimator is the average of the individual ratios of the participants, defined as

$$\text{MoR}(I_n) = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{d_i}. \quad (1)$$

- The *Ratio of Sums* (RoS) estimator is the ratio of the sum of the number of reported neighbors in the unknown group over the sum of degrees of the participants, defined as

$$\text{RoS}(I_n) = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n d_i}. \quad (2)$$

Killworth et al. (1998a) proposed the MoR assuming that the ratios are representative on average. Subsequently,

Killworth et al. (1998b) proposed the RoS, the most widely used NSUM estimator in the literature (see, for example Ocagli et al. (2021)), complementing the previous idea with a Binomial model for the ARD responses for the unknown group conditioned to the degree

$$y_i \sim \text{Bin}(d_i, \rho). \quad (3)$$

Killworth et al. (1998b) demonstrated that the maximum likelihood estimator corresponds to the ratio of the sums described in Eq. 2.

The NSUM are prone to biases due to the response behavior (Laga et al., 2021). *Response bias* corresponds to malicious responses. *Transmission error* stems from the participants' lack of knowledge of whether their acquaintances belong to the specified groups. *Recall error* arises from inaccuracies in recalling the number of contacts within these groups. Finally, another potential source of bias is the *barrier effect*. This bias arises from differences in the propensity to know someone based on the participant's characteristics, which contradicts the assumption that the ratios are representative.

2.1. The NSUM in the Robust Statistics framework

Robust Statistics is the field of Statistics that examines how the deviations from the main assumptions and the addition of outlying values affect the estimates. A fundamental concept in robust statistics is the *contamination*. Contamination refers to the alteration of a distribution. Concretely, the contamination of a distribution F with a distribution H is defined by the mixture distribution of F with probability $1 - \epsilon$ and H with probability ϵ , denoted as $(1 - \epsilon)F + \epsilon H$. A common example is the contamination with a single value, where the variable takes a fixed value x with some probability. This is modeled mathematically as $F_{x,\epsilon} = (1 - \epsilon)F + \epsilon \delta_x$, where δ_x is the point-mass distribution with support in x , assigning probability 1 to x and 0 elsewhere.

The most common estimator is the sample mean. However, the sample mean is not a robust estimator since the breakdown point is zero and the influence function is not bounded (Hampel et al., 2011). The *breakdown point* is the minimum fraction of contamination of a bad observation that can produce an arbitrary value (see the Supplementary Material). The *influence function* assesses the impact of a single outlier, concretely, the variation of the estimates when the fraction of contamination tends to zero (see Section SMA of the Supplementary Material for a formal definition).

As we observe in Eqs. 1 and 2, the NSUM estimators are based on the sample mean and therefore are not robust. Additionally, the NSUM estimators have more problems because of the nature of the ARD. For instance, the variables y_i , d_i , and y_i/d_i are generally not normal. This is especially true for the ARD of the unknown group, y_i , when the group size is small, as y_i is a discrete variable and most values tend to be zero. The distribution of the y_i has been observed to exhibit heavy tails and skewness in real data (Killworth et al., 1998a) (Zheng et al., 2006). Similarly,

real networks are often modeled by degree distributions that deviate from normality, such as power law distributions (Newman, 2018). Heavy-tailed distributions may lead to inflated standard errors, and poor convergence of the sample mean to a Normal distribution, thereby negatively impacting inferences (Wilcox, 2011). Skewness is another source of inaccurate inferences and can be difficult to distinguish from asymmetric contamination (Wilcox, 2011). Moreover, the assumption that the mean is a reliable measure for the typical participant under study in skewed distributions is questionable (Wilcox, 2011).

The biases of the NSUM can be modeled within this framework for some cases. The response biases caused by malicious responses at the extreme values can be modeled by contamination at the extreme of the distributions, which is a common approach in robust statistics. Contamination naturally applies to the transmission error when a fraction of the respondents do not know the membership of their acquaintances, and therefore follow a different distribution. Additionally, if the unknown group is small, as in many NSUM studies, the transmission error is approximately the contamination with a point-mass distribution with support at 0. The recall error is difficult to model and, to the best of our knowledge, there is only one proposal in Díaz-Aranda et al. (2024). Some instances of undercounting and overcounting can be explained by contamination caused by distributions shifted to the sides. Finally, while the barrier effect is less natural in this framework, it can still be modeled as contamination if the participants having different propensities to know members of the unknown group are limited to a small number of individuals.

3. Robust NSUM estimators

In this section, we propose several robust alternatives for each NSUM estimator. A robust procedure should achieve a good performance in the assumed conditions, stability under small deviations, and avoid catastrophic outcomes under high deviations (Huber and Ronchetti, 2009). Among the various approaches and procedures in robust statistics (see Staudte and Sheather (2011)), we focus on estimators based on descriptive measures of distributions that are not necessarily parametric. A descriptive measure is a parameter associated with each distribution. Below, we present the key concepts for our work.

Definition 1. A *descriptive measure* is a map $T : \mathcal{M} \rightarrow \mathbb{R}$, where \mathcal{M} is the space of all probability distributions on the sample space (Staudte and Sheather, 2011).

An example is the population mean, which can be expressed as $T(F) = E[X]$, where X is a random variable with distribution F . There is a natural estimator of the descriptive measure $T(F)$, $T(x_1, \dots, x_n)$, defined as $T(F_n)$, where x_1, \dots, x_n are samples of the distribution F and $F_n = \frac{1}{n} \sum_i \delta_{x_i}$ is the empirical measure. We are particularly interested in a type of descriptive measure: *location measures*.

Definition 2. *Location measures* are descriptive measures that satisfy the following properties for a random variable X with distribution F_X (Wilcox, 2011):

- (1) $T(F_{X+b}) = T(F_X) + b$.
- (2) $T(F_{-X}) = -T(F_X)$.
- (3) if $X \geq 0$, then $T(F_X) \geq 0$.
- (4) $T(F_{aX}) = aT(F_X)$ for $a > 0$.

An estimator of a location measure is referred to as a *location estimator*.

This section is organized as follows. First, we define and present the location measures that we will employ throughout the study. Next, we propose robustifications of MoR and RoS based on these measures.

3.1. Location estimators

We present the estimators based on location measures that we propose to robustify the NSUM estimators. They are based on the idea of giving less weight to extreme observations. Although the robustness of the sample mean has been widely studied, there is no extensive study regarding improvements when the distribution has skewness, as seen in Wilcox (2011), since asymmetric contamination may be confused with contamination (Huber and Ronchetti, 2009). Additionally, the frequent approach of the estimator in robust statistics assumes continuous distributions (see Huber and Ronchetti (2009)). In our case, we propose estimators derived from the adaptive estimation literature and foundational techniques in robust statistics. Concretely, we consider four types of estimators: the median, the trimmed means, the adaptive trimmed means, and the M-estimators.

Median. The median is an estimator that is very resistant to big deviations since its breakdown point is 0.5 (Maronna et al., 2019). Additionally, the median has good stability with bounded influence function (Huber and Ronchetti, 2009). The median is often selected as a location measure for skewed distributions, where the mean does not represent the typical participant under study (Wilcox, 2011). However, the median can have high bias and low efficiency in discrete distributions, due to the limited number of possible outcomes.

Trimmed Mean (TM). The trimmed mean is a type of L-estimator, an estimator that is a linear combination of order statistics $\sum_i a_i X_{(i)}$ (Huber and Ronchetti, 2009), where $X_{(i)}$ is the i th order statistic. The TM is a particular case in which the order statistics of the extremes are given a zero value. The α -trimmed mean (TM_α) is the estimator defined as

$$TM_\alpha(x_1, \dots, x_n) = \frac{1}{n-2g} \sum_{i=g+1}^{n-g} x_{(i)}, \quad (4)$$

where $g = \lfloor n\alpha \rfloor$ is the largest integer not exceeding $n\alpha$. TM_α has breakdown point of α and a bounded influence function (Huber and Ronchetti, 2009). More information can be found in Section SMA of the Supplementary Material.

Adaptive trimmed means. The TM estimator assigns a zero weight to extreme observations from both sides, which works well for symmetric distributions. However, in the presence of skewness, it is unclear whether the TM is an appropriate location measure. The asymmetric trimming arises as a solution for location measures in skewed data (Hogg, 1974). The (α_1, α_2) -trimmed mean is the estimator

$$TM_{(\alpha_1, \alpha_2)}(x_1, \dots, x_n) = \frac{1}{n - g_1 - g_2} \sum_{i=g_1+1}^{n-g_2} x_{(i)}. \quad (5)$$

where $g_1 = \lfloor n\alpha_1 \rfloor$ and $g_2 = \lfloor n\alpha_2 \rfloor$. Note that $TM_\alpha = TM(\alpha, \alpha)$. We prove in Section SMB in the Supplementary Material that the (α_1, α_2) -trimmed mean has a bounded influence curve, and a breakdown point of $\min(\alpha_1, \alpha_2)$. We also show in Section SMB the location measure for the (α_1, α_2) -trimmed mean.

The primary challenge with asymmetric trimmed means is selecting the appropriate trimming levels. Some trimmings may perform poorly in symmetric cases, and the best trimming values are not known in advance. One proposal is the *adaptive trimmed means*, in which the trimming values are based on the data obtained by the sample (Hogg, 1974). A way to adapt asymmetric trimming rates is the hinge location estimators proposed in Reed III and Stark (1996). They defined some measures of skewness and tail length

$$\begin{aligned} Q_1 &= (U_{(.2)} - L_{(.2)}) / (U_{(.5)} - L_{(.5)}), \\ Q_2 &= (U_{(.05)} - T_{.25}) / (T_{.25} - L_{(.5)}), \\ SK_2 &= (X_{(1)} - XMD) / (XMD - X_{(n)}), \\ SK_5 &= (X_{(1)} - XM) / (XM - X_{(n)}), \end{aligned}$$

where $L_{(\alpha)}$ is the mean of the smallest αn observations, $U_{(\alpha)}$ is the mean of the largest αn observations, $T_{.25}$ is the 0.25-trimmed mean $m_n(0.25)$, XM is the mean, and XMD the median. Calculating the adaptive trimmed means has the following steps:

- (1) Set the total amount of trimming γ ;
- (2) Determine the corresponding trimming proportion from the left side of the distribution

$$\alpha_l = \gamma [UW_X / (UW_X + LW_X)], \quad (6)$$

where UW_X and LW_X are the numerator and denominator portions of the statistics above;

- (3) Set the upper trimming proportion $\alpha_u = \gamma - \alpha_l$;
- (4) Calculate the sample (α_l, α_u) -trimmed mean.

Note that $\alpha_l = \gamma [UW_X / (UW_X + LW_X)] = \gamma [1 / (1 + 1/X)]$, where X represents one of the skewness and tail length measures presented above. Thus, a higher metric value implies more trimming to the lower size.

We denoted the adaptive hinge estimators as HQ_1 , HQ_2 , HSK_2 , and HSK_5 . Reed III and Stark (1996) presented more methods. However, we have selected these four because simulations revealed that HQ_1 , HQ_2 , HSK_2 , and HSK_5 were among the top 4 estimators, and HQ_1 and

HQ_2 among the top 3 (Reed III and Stark, 1996). Additionally, HSK_2 and HSK_5 were also good with symmetric distributions. In addition, Keselman et al. (2007) conducted some experiments to measure the type I error and power of the estimators testing the equality of means, showing that HQ_1 is the best estimator in terms of power and protection against type I error, and HSK_2 and HSK_5 are better with a distribution not far from a normal distribution. To the best of our knowledge, there is no research regarding the classical measures of robustness. One possible reason is the early criticism of other adaptive procedures, like the asymptotically efficient adaptive procedures (Huber and Ronchetti, 2009). A criticism to the adaptive trimmed means is the difficulty of differentiating skewness from contamination (Huber and Ronchetti, 2009). However, adaptive procedures have been shown to be empirically adequate in some families of distributions (Agostinelli et al., 2016) and for hypothesis testing of equality of means (Keselman et al., 2007).

M-estimators. M-estimators are estimators associated with the minimization problem

$$\arg \min_{\theta} \sum_{i=1}^n \xi(x_i; \theta), \quad (7)$$

or, with equation

$$\sum \psi(x_i; \theta) = 0, \quad (8)$$

being ψ the derivative of ξ with respect to θ . The associated measure is the solution of

$$E_F[\psi(X; T(F))] = \int \psi(x; T(F))F(dx). \quad (9)$$

An M-estimator of location is the solution to Eq. 7 in the location problem, i.e., when $\xi(x; \theta)$ is expressed as $\xi(x - \theta)$. One example of an M-estimator is the sample mean, which is the M-estimator with $\xi(x) = x^2/2$. Well-known M-estimators are based on the MoR-Huber functions

$$\psi(x) = \max\{-k, \min(x, k)\} \quad (10)$$

and

$$\xi(x) = \begin{cases} 1/2x^2 & \text{if } x \leq K, \\ |x|K - 1/2K^2 & \text{if } x > K. \end{cases} \quad (11)$$

The idea of the Huber M-estimators is to assign less weight to the extreme values that can distort the sample mean. In particular, we will use the Huber Proposal 2 (Huber and Ronchetti, 2009). This is a Huber M-estimator that uses a scale parameter τ . Concretely, the M-estimator is a solution of

$$\sum_i \psi\left(\frac{x_i - \theta}{\tau}\right) = 0. \quad (12)$$

In this proposal, τ and θ are estimated jointly, using an M-estimator of scale for the scale parameter τ (Huber and

Ronchetti, 2009). Specifically, we use the MoR-Huber M-estimator with $K = 1.5$. The breakdown point for Proposal 2 is 0.27 and it has a bounded influence function (Huber and Ronchetti, 2009).

An M-estimator well suited for heavy-tail distributions is the myriad (Arce, 2004). It gives less weight to the extreme values than the Huber type. Particularly, it corresponds to the maximum likelihood estimator for the Cauchy distribution (Arce, 2004). It is defined as an M-estimator with ξ function

$$\xi(x) = \log(K^2 + x^2), \quad (13)$$

where K is a parameter that measures the influence of the extreme values, with a larger K representing a greater effect on these values. The myriad is employed as a filter for signal processing (Ramirez and Paredes, 2017), where K represents the rejection capability to impulsive noise (Ramirez and Paredes (2016)). To calculate the solution of the myriad, we will use the fast weighted myriad computation of Arce (2004).

3.2. Robustification of MoR and RoS

In the previous section, various location measures resistant to contamination by outlying values have been introduced. They behave similarly to the mean under normal or symmetric short-tailed distributions but offer better performance in the presence of asymmetry and heavy tails. This flexibility allows us to change the underlying assumptions of the classical NSUM estimators.

Robustification of MoR. MoR relies on the assumption that the average of y_i/d_i is representative. This can be expressed in our context as $E[y_i/d_i] = \rho$. Our proposals to robustify the MoR consist of substituting the population mean with one of the robust location measures T , i.e., assuming that $T[y_i/d_i] = \rho$. This leads to an estimator of the prevalence using the location estimator applied to the sample ratios

$$T(I_n) = T(y_1/d_1, \dots, y_n/d_n). \quad (14)$$

Under distributions closer to a Normal distribution, the estimates are similar to the sample means, but with estimates more resistant to deviations.

Robustification of RoS. For the RoS, we first observe that for the Binomial model

$$E[y_i] = E[E[y_i|d_i]] = \sum_k k\rho P[d_i = k] = \rho E[d_i]. \quad (15)$$

Similarly to MoR, we assume that there are some location measures such that $T[y_i] = \rho T[d_i]$. We can construct an estimator for the prevalence by taking the ratios of the estimators of y_i and d_i

$$\text{RoT}(I_n) = T(y_1, \dots, y_n)/T(d_1, \dots, d_n). \quad (16)$$

Table 1 displays the robust estimates that we use in the rest of the study. Each column corresponds to the robust proposals for each non-robust estimator.

Table 1

Robust NSUM estimators for each type of classical NSUM.

Robust NSUM	
MoR-type	RoS-type
MoR-Median	RoS-Median
MoR-TM10	RoS-TM10
MoR-HQ1	RoS-HQ1
MoR-HQ2	RoS-HQ2
MoR-HSK2	RoS-HSK2
MoR-HSK5	RoS-HSK5
MoR-Huber	RoS-Huber
MoR-Myriad	RoS-Myriad

The first column represents the robustification of MoR. We use the location estimator, which applies the ratios of y_i/d_i , and it is denoted using the name of the location estimator. MoR-TM10 denotes the 0.10-trimmed mean $TM_{0.10}$ and MoR-Huber is the name of the Huber M-estimator. The right column represents the robustification of RoS, which uses the ratios of the location estimators.

4. Simulation Study

This section evaluates the performance of the NSUM estimators through an extensive simulation study designed to capture the properties of real-world networks. This approach allows us to gather evidence on how these estimators perform under various conditions that reflect practical scenarios. First, we define the environment and its relation to conditions impacting the stability of the classical NSUM. Subsequently, we present the simulation results. Finally, we include a discussion and remarks on the performance of the estimators. Additional analysis of the barrier effect and other biases—transmission error and recall error—is presented in the supplementary material

4.1. Simulation environments

This section outlines the simulation environment used to assess robust NSUM estimation methods. Specifically, we generate synthetic graphs based on diverse network models, unknown groups with different prevalence levels, and sampling schemes with varying contamination intensities. This simulation framework provides insights into the scenarios where robust estimators must be most effective. In this work, we consider three network models: Erdős–Rényi (ER), Scale-Free (SF), and Stochastic Block Model (SBM). ER networks represent models with good conditions for the MoR and RoS when the prevalence is high. The ER models with low prevalences reflect deviations due to skewness and tail length. An extreme case of these conditions is reflected in SF networks. Finally, SBM networks assess the influence of the barrier effect. For a comprehensive analysis, we describe the characteristics of each model, the type of contamination applied, and the metrics used to evaluate the quality of the estimators.

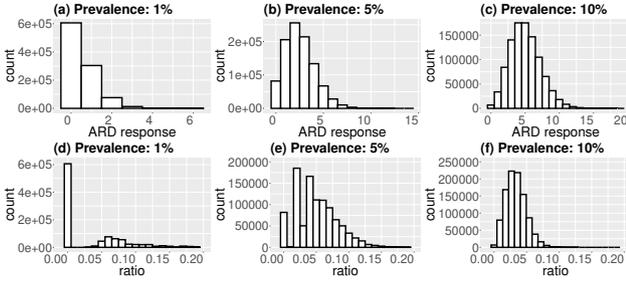


Figure 1: Erdős–Rényi ($\mu = 50$). Histogram of ARD responses y_i for (a) $\rho = 1\%$, (b) $\rho = 5\%$, and (c) $\rho = 10\%$. Histogram of ratios y_i/d_i for (d) $\rho = 1\%$, (e) $\rho = 5\%$, and (f) $\rho = 10\%$.

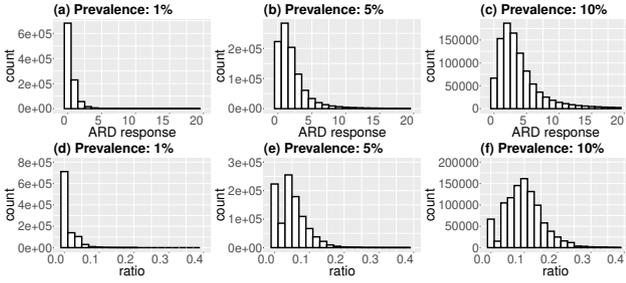


Figure 2: Scale-Free network model ($\gamma = 2.50$). Histogram of ARD responses y_i for (a) $\rho = 1\%$, (b) $\rho = 5\%$, and (c) $\rho = 10\%$. Histogram of ratios y_i/d_i for (d) $\rho = 1\%$, (e) $\rho = 5\%$, and (f) $\rho = 10\%$.

4.1.1. Erdős–Rényi

The Erdős–Rényi (ER) model is a classic random graph model widely used to study the properties of networks. This model involves randomly connecting nodes with a fixed probability $p = \mu/(N - 1)$, where μ represents the mean degree and N is the network size (Cheng et al., 2019). We have chosen the ER model to represent favorable conditions for the classical methods when the prevalence is moderate. Figure 1(f) illustrates that, with relatively high prevalence, the distributions of the ratios are very close to a symmetric distribution with short tails, which benefits the performance of MoR. Additionally, the ER model closely matches the underlying assumptions of RoS, as the conditional distribution of y_i given d_i under an ER model follows a Hypergeometric distribution, which approximates a Binomial distribution when both the network and the unknown group are large (Laga et al., 2023a). The Binomial model corresponds to a similar network model in which nodes connect with others at a fixed probability, with replacement (Laga et al., 2023a). Thus, the ER model provides an ideal balance between the theoretical foundations of RoS and real-world applicability.

We use low prevalences to assess the stability of the NSUM estimators. We observe in Figure 1 that the distribution of the ARD of the unknown group and the ratios present skewness and heavy tails as the prevalence decreases. In addition, low prevalences compromise the approximation to normal distributions of the ARD and the ratios.

4.1.2. Scale-Free

This network model belongs to a class of random graphs with non-uniform degree distributions. The degree of SF networks follows a power law distribution, i.e., the probability $p(d)$ that a node in the network has d connections to other nodes is expressed as $d^{-\gamma}$ (usually with $\gamma > 2$). In particular, this implies a heavy tail in the degree distribution, in which most nodes in these networks have relatively low degrees (few connections to other nodes), while a few nodes, dubbed hubs, have very high degrees (many connections). The SF network model is widely adopted to simulate many real-world networks, including social and biological systems (Newman, 2018).

Particularly, we use the model of Goh et al. (2001). This model indexes the nodes with integer numbers, and gives the i th node a weight proportional to $i^{-\alpha}$, for $\alpha \in [0, 1)$. Then, two nodes are selected to form a link with a probability proportional to their respective normalized weights. Goh et al. (2001) show that the degree distribution follows a power law with $\gamma = (\alpha + 1)/\alpha$.

This model represents a significant departure from conditions of symmetry and short tails under which the classical methods perform well. Figure 2 shows that the distributions exhibit increased skewness and longer tails, even at higher prevalence levels.

4.1.3. Stochastic Block Model

The barrier effect involves different propensities to be connected to different groups depending on the characteristics of the participants. One type of barrier effect arises when the characteristics depend on membership in some groups. Kunke et al. (2024) proposed an SBM to model the propensity of different groups to form links. Specifically, the SBM model is a widely used network model that aims at capturing community patterns in which nodes do not have an equal likelihood of forming links within any group. These networks are valuable for social network analysis, where the interactions within and between distinct groups are different. In general, nodes within the same group are more (or less) likely to connect with each other than with nodes in different groups. To be more precise, consider a network with N nodes partitioned into K disjoint groups, i.e., each node belongs to exactly one group. The likelihood of an edge (connection) between two nodes depends on the groups to which the nodes belong. This behavior is typically represented by a symmetric matrix denoted as P , where p_{ij} represents the probability of an edge between a node in group i and a node in group j .

In our case, we replicate the scenario of (Kunke et al., 2024). They consider two blocks, the unknown group and its complement. The probability of connection between the two groups is p , while the intra-group link probability is ap , where a is a non-negative real representing the network assortativity. When $a < 1$ (dissortativity), nodes in the unknown subpopulation are less likely to connect with other nodes in the same group, which is often unrealistic in practical applications. When $a = 1$, it corresponds to

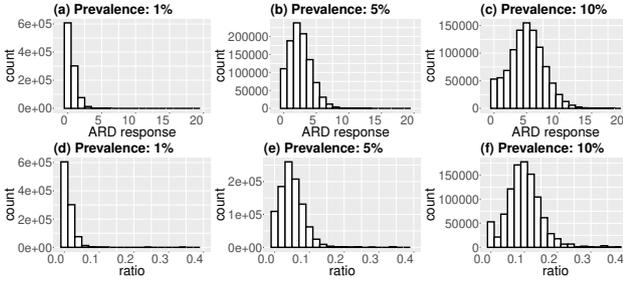


Figure 3: Stochastic Block model ($a = e^4$). Histogram of ARD responses y_i for (a) $\rho = 1\%$, (b) $\rho = 5\%$, and (c) $\rho = 10\%$. Histogram of ratios y_i/d_i for (d) $\rho = 1\%$, (e) $\rho = 5\%$, and (f) $\rho = 10\%$.

the Erdős–Rényi network model. In contrast, when $a > 1$ (assortativity), the nodes in the unknown subpopulation are more likely to connect with other nodes within the same population (Kunke et al., 2024).

Figure 3 shows the impact of prevalence on the skewness and tails of the ARD distribution in a highly assortative scenario.

4.1.4. Contamination model

In our simulation studies, we address the introduction of contamination to the ARD response vector to evaluate the resilience of robust estimation methods under adverse conditions. Specifically, for ER and SF network models, a predetermined fraction ϵ of the ARD response vector entries y_i is deliberately altered to introduce noise. This alteration is conducted as follows. Half of the specified proportion $\epsilon/2$ of the ARD response vector entries is set to zero. This perturbation simulates scenarios where response data might be missing or fail to be reported, a common issue in indirect survey data collection processes. The remaining half $\epsilon/2$ of the affected entries is substituted with the respective node degree value d_i . This replacement simulates a different type of data corruption, potentially due to errors in data handling or misalignment in database entries.

The contamination scenarios defined (missing data and data corruption) constitute extreme forms of under- and over-reporting, something highly relevant to NSUM techniques. Furthermore, some of these contaminations can be classified as biases typical of ARD and NSUM techniques.

4.1.5. Metrics

In this study, we assess the performance of robust NSUM estimators through two statistical metrics: the root mean squared relative error (RMSRE) and normalized bias (NB). These metrics allow us to evaluate the accuracy and reliability of robust estimators under varied simulation conditions. On the one hand, the RMSRE measures the relative discrepancies between the estimated values and the actual prevalence. More precisely, the RMSRE is defined as follows

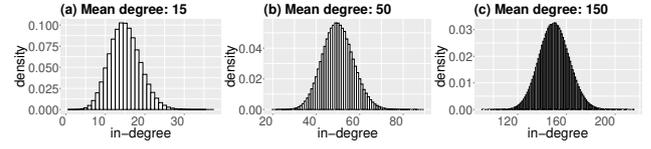


Figure 4: Erdős–Rényi. Histogram of the degree for (a) $\mu = 15$, (b) $\mu = 50$, and (c) $\mu = 150$.

$$\text{RMSRE} = \sqrt{\frac{1}{R} \sum_{i=1}^R \left(\frac{\rho_i - \rho}{\rho} \right)^2} \quad (17)$$

where R is the number of realizations or simulation runs, ρ_i denotes the estimated prevalence obtained from the i th realization, and ρ is the actual prevalence value. This metric effectively captures the average magnitude of errors relative to the true value, providing a measure of estimator precision. To complement the analysis provided by RMSRE, we also calculate the normalized bias (NB), which offers insights into the systematic deviation of the estimators from the actual value. The NB is expressed as

$$\text{NB} = \frac{\left(\frac{1}{R} \sum_{i=1}^R \rho_i \right) - \rho}{\rho} = \frac{\left(\frac{1}{R} \sum_{i=1}^R \rho_i \right)}{\rho} - 1. \quad (18)$$

Hence, the NB quantifies how closely the average estimate matches the actual parameter, normalized by the actual prevalence value. NB helps compare estimators across different prevalence levels as it scales the bias to the actual prevalence value.

4.2. Simulations Results

For each network model, we generated ten synthetic graphs with 10^6 nodes to simulate large-scale network structures. It is essential to mention that our analysis focuses on assessing the accuracy of robust NSUM estimators for prevalence levels ranging from 1% to 10%. In this context, we created ten unknown subpopulations for each network model and prevalence value ρ . Notice that prevalence refers to the proportion of the network nodes belonging to the unknown subpopulation. Every value in tables and figures shown in the section on simulation results is obtained by averaging 1000 realizations. Specifically, the 1000 realizations consider ten networks of a particular model, ten unknown subpopulations with a given ρ , and ten trials. At each trial, a set of node degree samples and a set of ARD responses with a sample size of $n = 500$ are extracted. In each trial samples are selected uniformly at random.

Regarding the parameter of the robust NSUM, we fix the trimming proportion to 0.20, i.e., the trimmed means give zero weight to 20% of the data. In particular, we use $TM_{0.1}$, which we will refer to as MoR-TM10 and RoS-TM10 in the rest of the study. Furthermore, the linearity parameter K assigned to the Myriad estimator is 0.2. We aim to assess the performance of the Myriad-based estimators across

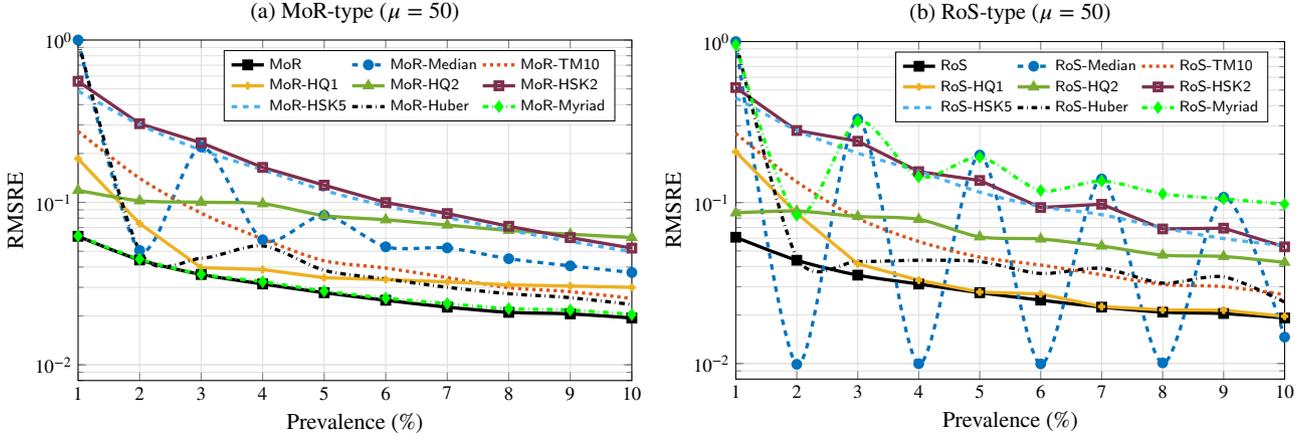


Figure 5: Erdős–Rényi ($\mu = 50$). RMSRE versus the prevalence without contamination for the groups of robust estimators: (a) MoR-type and (b) RoS-type.

various forms of contaminations related to NSUM while maintaining constant linearity parameters. The parameter for the MoR-Huber and RoS-Huber M-estimators is fixed to 1.5.

4.2.1. Erdős–Rényi

This section evaluates the performance of robust NSUM estimation methods using synthetic graphs generated according to the ER model, specifically focusing on networks with mean degrees of 15, 50, and 150. These values were selected considering the social group sizes for good friends, friends, and casual acquaintances according to Dunbar’s theory of social circles (Dunbar, 2010). Figure 4 shows histograms of the degree distribution for ER networks with (a) $\mu = 15$, (b) $\mu = 50$, and (c) $\mu = 150$. As expected, these histograms exhibit a bell-shaped profile centered on the average degree, similar to a binomial distribution.

First, we examine the performance of robust estimators across varying prevalence levels under conditions without contamination. This setup enables the examination of their usability, accuracy, and potential biases without external disruptions, providing insight into their reliability in ideal scenarios without contamination. Our primary focus is on networks with a mean degree of 50, as this provides a midrange complexity suitable for detailed analysis. Figures 1(a)–(c) display the histograms of ARD responses for unknown groups at prevalence levels of 1%, 5%, and 10%, respectively. At the lowest prevalence (1%), more than 60% of respondents report zero known members in the hidden group. As the prevalence increases to 5% and 10%, zero-valued ARD responses show a noticeable decrease. Figures 1(d)–(f) complement this analysis by showing histograms of the ratios for the same prevalence levels. These histograms exhibit a concentration at zero when prevalence is low, gradually diminishing as the hidden population becomes more prevalent.

Figure 5 illustrates the RMSRE obtained by robust estimation methods as a function of the prevalence of the hidden population without contamination. Figure 5(a) displays the RMSRE for the MoR-type estimators across the prevalence interval. Without contamination, the non-robust

MoR estimator consistently achieves the lowest RMSRE values, outperforming all robust variants across the prevalence range. Among the robust approaches, MoR-Myriad exhibits the closest performance to the standard MoR estimator, indicating that this estimator is effective in clean environments. Other robust estimators, such as MoR-TM10, MoR-HQ1, and MoR-Huber, show progressively improved accuracy as prevalence increases, whereas MoR-Median exhibits unstable performance, particularly at low prevalence levels. Notably, MoR-HQ2, MoR-HSK2, and MoR-HSK5 perform comparatively poorly, yielding the highest RMSRE values and demonstrating limited utility in uncontaminated settings. For additional insights, Figure SM1 in the Supplementary Material provides boxplots of the estimated hidden population proportions obtained by the MoR-type methods for prevalence levels ranging from 1% to 10%. These visualizations offer additional insight into estimator variability and bias under uncontaminated conditions.

Figure 5(b) displays the RMSRE obtained by RoS-type methods versus the prevalence in the absence of contamination. Similar to the MoR family, the standard (non-robust) RoS estimator consistently yields the lowest RMSRE values across the prevalence range, indicating superior efficiency under uncontaminated conditions. RoS-Median occasionally outperforms the standard RoS estimator at specific prevalence levels. It is worth noting that the RoS-Median estimator exhibits oscillatory behavior as prevalence increases. This phenomenon stems from the use of medians, which are not smoothly responsive to small changes in the data. Specifically, RoS-Median is defined as the ratio between the median of the ARD responses and the median of the reported degrees. Because the median is a discrete, order-based statistic, the ratio can remain constant over a range of prevalence values and then suddenly shift when a small change in prevalence causes the median of the ARD response to increase by one. For example, in a network with a median degree of 50, the estimator can shift from $1/50$ to $2/50$ as soon as the median ARD count increases by one, resulting in a stepwise behavior in estimation and an oscillatory error pattern as prevalence changes smoothly.

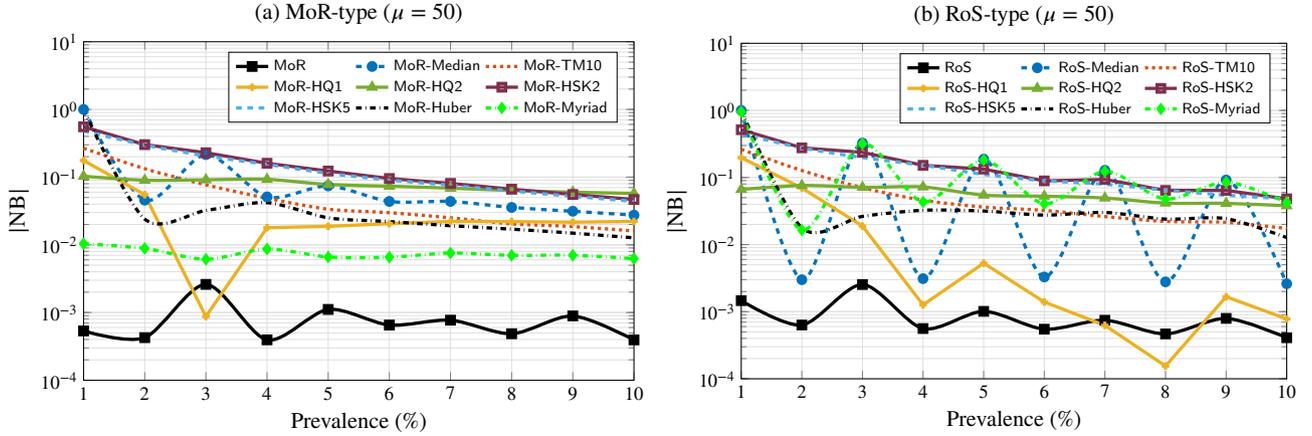


Figure 6: Erdős-Rényi ($\mu = 50$). [NB] versus the prevalence without contamination for the groups of robust estimators: (a) MoR-type and (b) RoS-type.

RoS-Myriad also shows an unstable behavior. These trends are further illustrated in the boxplots of Figure SM2 (Supplementary Material). The RoS-HQ1, RoS-Huber, and RoS-TM10 show improved accuracy as prevalence increases, indicating that the performance of these methods becomes more reliable with a larger unknown population. In contrast, RoS-HSK2 and RoS-HSK5 consistently underperform across all prevalence levels, producing the highest RMSRE values.

Figure 6(a) displays the absolute value of the normalized bias [NB] obtained by MoR-type estimators for different prevalence levels (1% to 10%) under contamination-free conditions. As expected, the standard (non-robust) MoR estimator yields the lowest bias throughout the prevalence interval. This behavior is consistent with the non-robust MoR being an unbiased maximum likelihood estimator (MLE) when ratios follow a Gaussian distribution (Laga et al., 2021). On the other hand, all MoR-type robust methods exhibit non-zero bias when there is no contamination, indicating deviations of estimations from actual prevalence values. We also observe a reduction in bias as the prevalence increases for all robust techniques, leading to more accurate estimations as the proportion of the unknown group increases. Among the robust approaches, MoR-Myriad consistently achieves the lowest bias across all prevalence levels. Moreover, MoR-HQ1 and MoR-Huber estimators also exhibit relatively low bias for prevalence levels above 1%, reinforcing their utility in moderate or high prevalence scenarios.

Figure 6(b) illustrates the absolute values of the normalized bias [NB] yielded by the RoS-type estimators across a range of prevalence levels under uncontaminated conditions. The non-robust RoS method exhibits the lowest bias across prevalence levels. Among the robust methods, RoS-HQ1 shows very low bias magnitudes as the prevalence increases, suggesting minimal deviation from the true prevalence. In contrast, RoS-Median shows high variability in its bias, especially at lower prevalence levels, reducing its reliability as an estimator in scenarios involving small hidden populations. The RoS-HSK2 and RoS-HSK5 estimators yield

the largest bias values across the entire prevalence range, indicating poor estimation accuracy in environments without contamination.

To further examine the performance of the estimation methods in scenarios without contamination, Figure SM3 in the Supplementary Material depicts the RMSRE produced by the various estimation techniques against the prevalence for ER networks with a mean degree of $\mu = 15$. Moreover, Figure SM4 presents the absolute values of the NB yielded by the estimators for different prevalence values. In addition, Figures SM5 and SM6 display the boxplots of the estimates of the unknown population rate yielded by the various methods for different prevalence levels in the absence of contamination and mean degree of $\mu = 15$.

Similarly, Figure SM7 in the Supplementary Material shows the RMSRE generated by the NSUM estimators as a function of the prevalence for ER networks with a mean degree of $\mu = 150$. Figure SM8 displays the absolute values of the NB for different prevalence levels without contamination. These figures indicate that the non-robust methods outperform the robust estimators when the network follows an ER model in scenarios without contamination. Additionally, robust estimation methods exhibit lower error and bias values as the mean degree increases, indicating that their responses are more accurate as link probabilities between network nodes increase. Figures SM9 and SM10 show the boxplots of the estimates of the unknown population rate produced by the various methods for different prevalence levels in the absence of contamination and mean degree $\mu = 150$.

We also evaluate the performance of NSUM estimation methods for the ER model under varying contamination levels. Figures 7(a)-(c) display the RMSRE produced by the MoR-type methods at contamination levels ranging from 0% (no noise) to 10%, across three prevalence values of 1%, 5%, and 10%, respectively. As illustrated in the figures, the performance of the non-robust MoR estimator significantly deteriorates as the level of contamination increases across all three prevalence scenarios. It is worth noting that the MoR-HQ2 method exhibits a marked decline in performance with

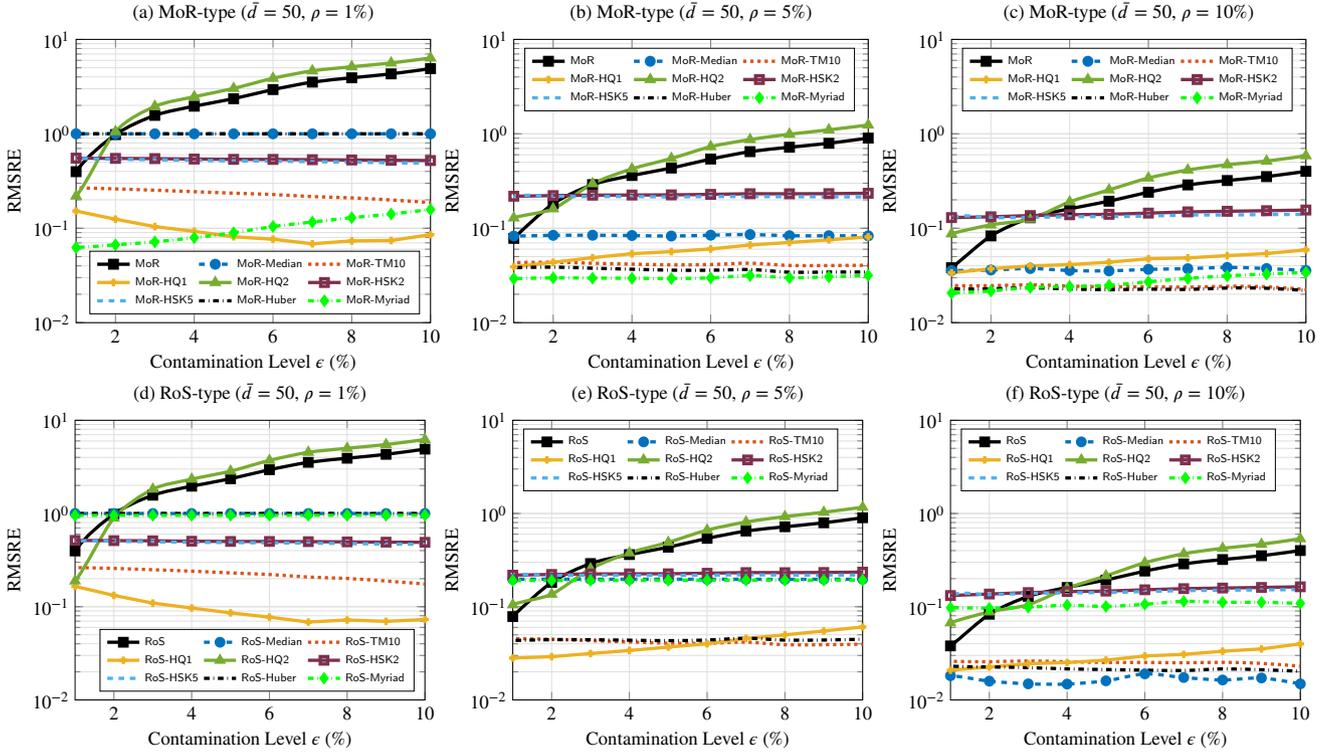


Figure 7: Erdős–Rényi ($\mu = 50$). RMSRE versus the contamination level for MoR-type methods and prevalences: (a) 1%, (b) 5%, (c) 10%; RoS-type methods and prevalences: (d) 1%, (e) 5%, (f) 10%.

increasing contamination. In particular, once the contamination level exceeds 2%, MoR-HQ2 produces estimation errors larger than those of the non-robust MoR estimator, indicating inferior performance in contaminated settings. In contrast to the MoR and MoR-HQ2 methods, which exhibit considerable sensitivity to noise, the remaining robust estimators maintain relatively stable performance across varying contamination levels. As shown in Figure 7(a), for a prevalence of 1%, the robust estimator MoR-Myriad achieves RMSRE values of at least one order of magnitude lower than those produced by the non-robust MoR method. This highlights the robustness of this approach, particularly in scenarios involving small hidden populations, a setting that often represents a worst-case scenario in ER networks. Similar performance gains are observed for prevalence levels of 5% and 10%. Figures SM11, SM13, and SM15 in the Supplementary Material depict the boxplots of the unknown population rate estimates generated by the MoR-type estimator at prevalences 1%, 5%, and 10%.

Among the robust methods, MoR-Myriad, MoR-TM10, and MoR-HQ1 estimators consistently exhibit low error values across the entire contamination interval for the three prevalence levels, highlighting their potential for application in noisy, large-scale networks. Although MoR-Huber and MoR-Median show improved performance at higher prevalence levels, they exhibit significant errors at a 1% prevalence, which can be problematic in estimating small unknown groups. To provide a deeper insight, Figures SM17(a)-(c) show $|\text{NB}|$ versus the contamination levels produced by

the MoR-type methods for prevalences of 1%, 5%, and 10%, respectively.

Figures 7(d)-(f) show the RMSRE yielded by RoS-type methods versus contamination levels across prevalences of 1%, 5%, and 10%, respectively. In addition, Figures SM12, SM14, and SM16 in the Supplementary Material illustrate the boxplots obtained by the RoS-type techniques against contamination levels and prevalence 1%, 5%, and 10%, respectively. Moreover, Figures SM17(d)-(f) in the Supplementary Material depict the $|\text{NB}|$ against the contamination level yielded by the RoS-type estimators for prevalences of 1%, 5%, and 10%, respectively. It is observed that both errors and bias values generated by the non-robust RoS and RoS-HQ2 methods significantly increase as the contamination levels increase, demonstrating their sensitivity to noise. The remaining RoS-type methods, however, exhibit consistent error and bias values throughout the contamination interval. Among the robust methods, the RoS-HQ1 and RoS-TM10 estimators perform best for the three prevalence values. Figures SM18, SM19, SM20, SM21 in the Supplementary Material illustrate the behavior of various NSUM estimators in response to perturbations simulating other NSUM biases (precisely, transmission error and recall error) for the ER network model.

4.2.2. Scale-Free

This section evaluates the performance of the NSUM estimation methods using Scale-Free networks characterized by a power-law degree distribution with a scaling exponent

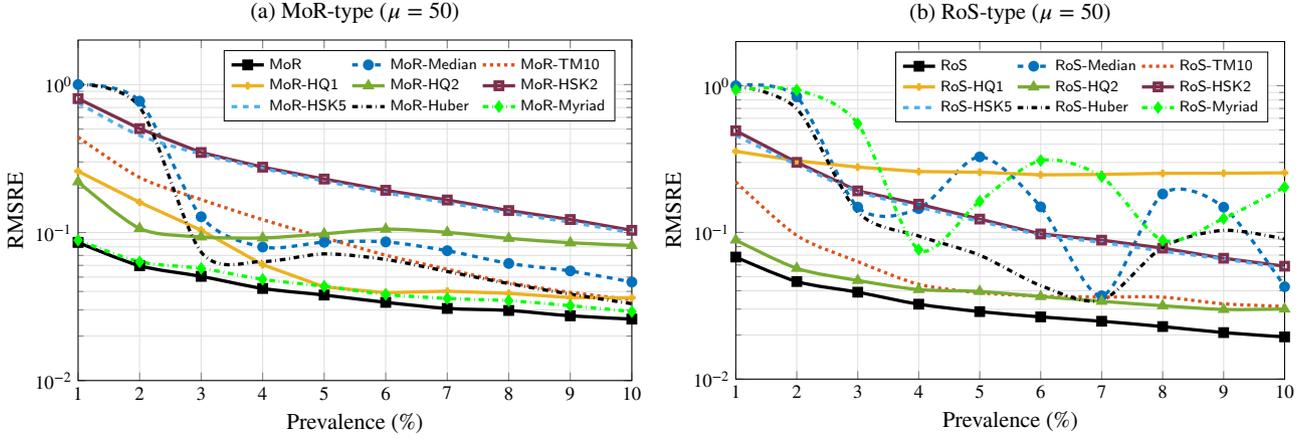


Figure 8: Scale-Free ($\mu = 50$). RMSRE versus the prevalence without contamination for the groups of robust estimators: (a) MoR-type and (b) RoS-type.

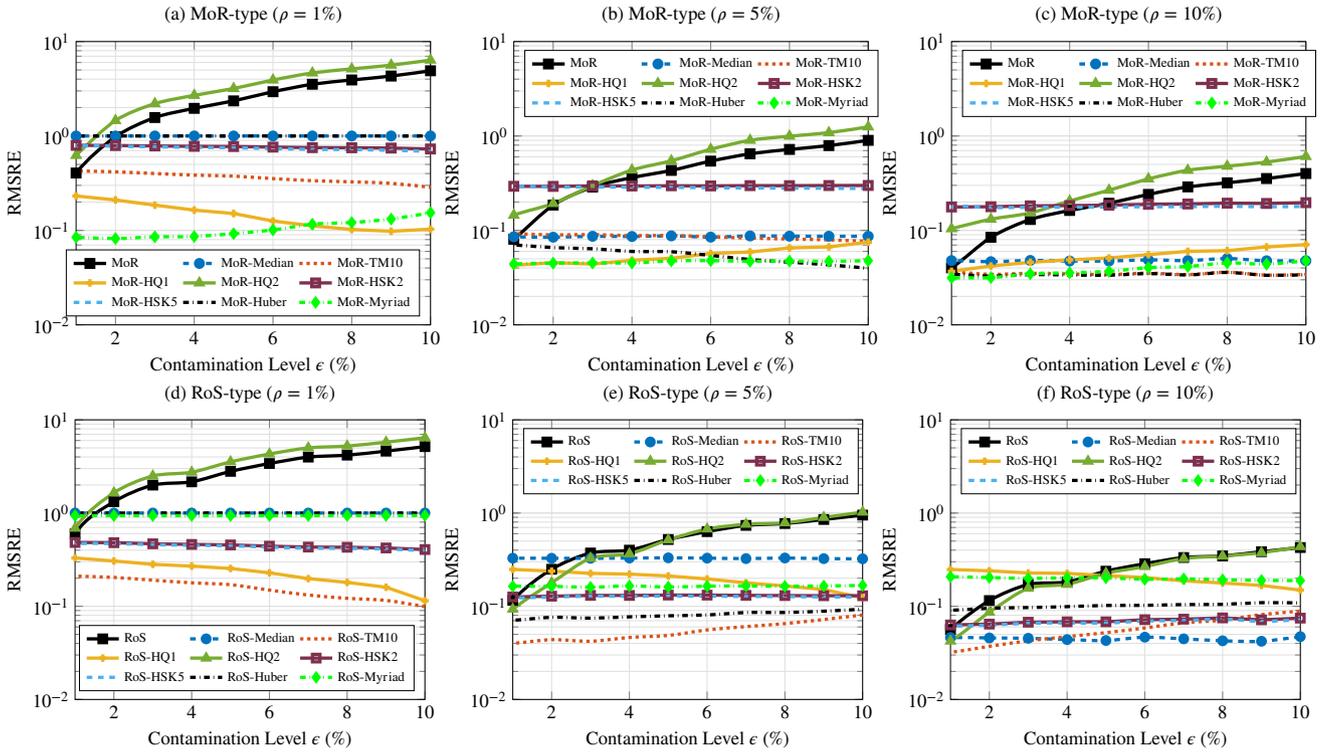


Figure 9: Scale-Free network model ($\gamma = 2.50$). RMSRE versus the contamination level for MoR-type methods and prevalences: (a) 1%, (b) 5%, (c) 10%; RoS-type methods and prevalences: (d) 1%, (e) 5%, (f) 10%.

$\gamma = 2.50$ and mean degree $\mu = 50$. Figure 2 illustrates the histograms of ARD responses and ratios for three prevalence levels: 1%, 5%, and 10%. First, we focus on assessing the accuracy of the estimations without contamination. Figure 8 displays the RMSRE obtained by the different estimators versus the prevalence in the absence of noise. Additionally, the boxplots of the unknown population rate estimated by the various estimators for prevalence values ranging from 1% to 10% without contamination are displayed in Figures SM22 and SM23 in the Supplementary Material.

Figure 8(a) shows the RMSRE generated by MoR-type methods versus the prevalence. In addition, Figure SM24(a)

in the Supplementary Material illustrates the $[NB]$ values yielded by the MoR-type methods for different prevalence levels. In SF networks with $\gamma = 2.50$ and no contamination, the non-robust MoR method gives the lowest error and bias values across the entire prevalence interval, similarly to the ER network results. These results indicate that this method maintains high accuracy in scenarios without contamination despite the non-robust MoR estimator not being the optimal operator when the network degree follows a power-law distribution. Among robust estimators, the MoR-Myriad method performs closest to the non-robust MoR technique. Notice also that MoR-TM10 and MoR-HQ1 slowly improve

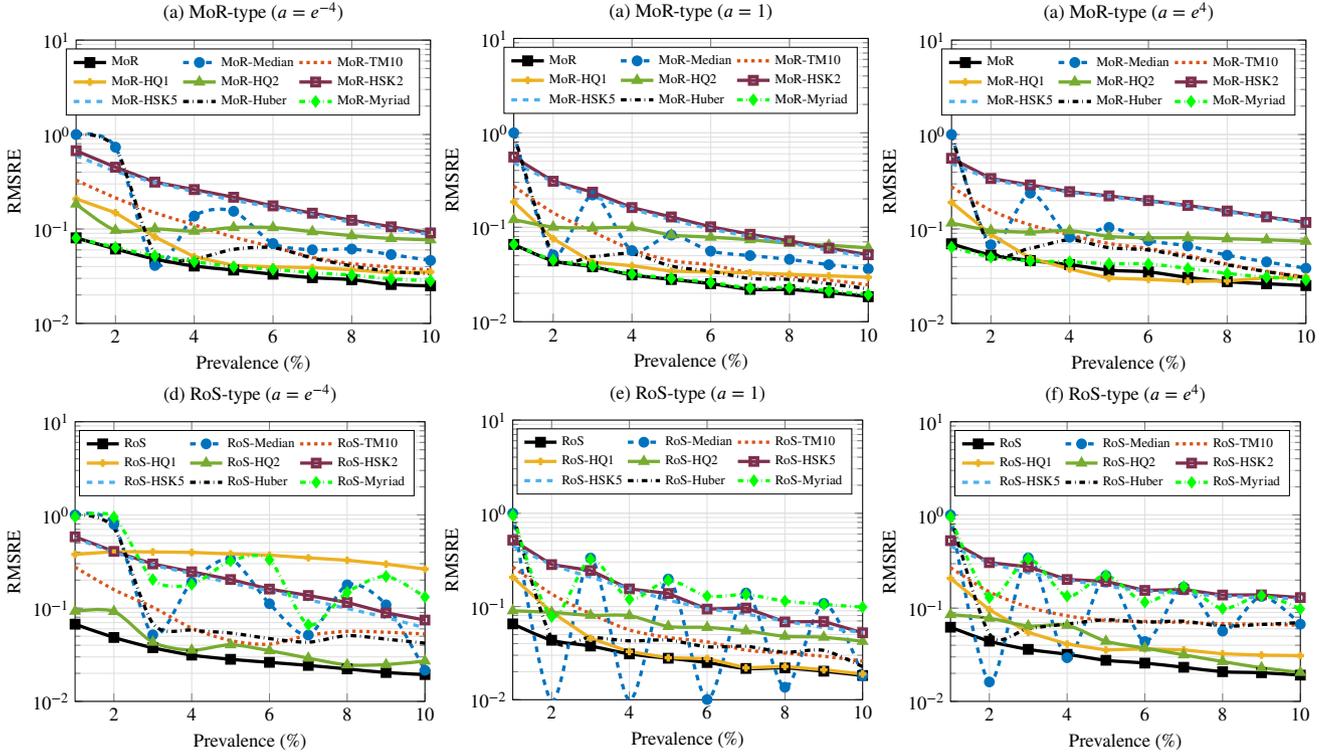


Figure 10: Stochastic Block Model. RMSRE versus the prevalence level for different values of the connectivity factor a . MoR-type methods: (a) $a = e^{-4}$, (b) $a = 1$, (c) $a = e^4$; RoS-type methods: (d) $a = e^{-4}$, (e) $a = 1$, (f) $a = e^4$.

their error values as the prevalence increases. Observe also that MoR-Median and MoR-Huber yield the largest errors for low prevalence values (1% and 2%), making them unreliable for estimating the prevalence of small unknown populations. In scenarios without contamination, MoR-HSK2 and MoR-HSK5 show the poorest performances regarding RMSRE and NB.

Figure 8(b) illustrates the RMSRE produced by RoS-type estimation methods as a function of the prevalence. Further, Figure SM24(b) in the Supplementary Material includes the $|\text{NB}|$ values yielded by the RoS-type methods across the same prevalence levels. The non-robust RoS, RoS-HQ2, and RoS-TM10 methods exhibit the lowest RMSRE and NB across all tested prevalence levels. These results indicate that RoS, RoS-HQ2, and RoS-TM10 methods produce accurate and consistent estimates in uncontaminated environments. Conversely, the RoS-Median, RoS-Myriad, and RoS-Huber methods demonstrate variability in their performance across the prevalence axis. More precisely, these methods exhibit unstable behavior, particularly noted in their fluctuating error rates along the prevalence levels. A notable decline in effectiveness is observed at lower prevalence levels, a significant limitation for applications involving small, unknown groups.

We have also analyzed the performance of NSUM estimators using Scale-Free networks under noisy conditions and varying prevalence levels. Figures 9(a)-(c) illustrate the RMSRE produced by the MoR-type estimators as a function

of the contamination level at three different prevalence values: 1%, 5%, and 10%. It is worth noting that the error values decrease for all estimation methods as prevalence levels increase, which suggests that higher node activation within the network leads to more accurate estimations regardless of noise. For the MoR and MoR-HQ2 methods, error values increase significantly as contamination levels rise, highlighting their limitations in applications with highly contaminated environments. On the other hand, the MoR-Myriad operator and the MoR-HQ1 method consistently deliver the best performance across the contamination range for the three prevalence levels. Notably, at the 1% prevalence level, MoR-Median, MoR-Huber, MoR-HSK2, and MoR-HSK5 exhibit significant errors, underlying the weakness of these methods when dealing with small unknown groups.

Figures 9(d)-(f) depict the RMSRE yielded by the RoS-type estimators versus the contamination level at three different prevalence values: 1%, 5%, and 10%. In this case, the RoS-TM10 estimator consistently performs best along the contamination interval for the three prevalence levels. Finally, Figure SM25 in the Supplementary Material displays the $|\text{NB}|$ produced by robust estimators versus the contamination level for the three prevalence values. Figures SM26, SM27, SM28, SM29 in the Supplementary Material depict the behavior of NSUM estimators in response to perturbations simulating other NSUM biases (transmission error and recall error) for the Scale-Free network model.

4.2.3. Stochastic Block Model

In this section, we analyze the performance of the NSUM estimators using two-group SBM graphs with varying connectivity factor a . These graphs maintain a constant mean degree of $\mu = 50$. We apply nine distinct factors $a = e^M$ with $M = -4, -3, \dots, 3, 4$. Positive values of M represent more connections between the unknown group, while negative values of M represent more connections between the unknown group and the rest of the individuals of the population. This set of values enables us to assess the robustness of NSUM estimation methods concerning the barrier effect. All graphs have the same mean degree, ensuring that the average connectivity across the network remains consistent. Figure 3 displays the histograms of ARD responses and ratios for assortativity $a = e^4$ and prevalence values of 1%, 5%, and 10%. Figures 10(a)-(c) display the RMSRE versus the prevalence yielded by the MoR-type methods for three different connectivity factors $a: e^{-4}, 1, \text{ and } e^4$. These factors indicate various levels of connectivity within the group. To provide a comprehensive analysis, Figures SM30(a)-(c) in the Supplementary Material present the |NB| values generated by the MoR-type methods versus the prevalence for connectivity factors $a = e^{-4}, a = 1, a = e^4$, respectively. The non-robust MoR and the MoR-Myriad methods consistently show superior performance across the entire prevalence level interval for the three connectivity factors. Conversely, the MoR-HQ2 method consistently produces errors about one order of magnitude larger than those obtained by the non-robust MoR method across the prevalence interval for the three connectivity factors. In addition, the MoR-HQ1 and MoR-TM10 exhibit improved accuracy for the three connectivity factors as the prevalence increases. In other words, these estimators perform poorly at low prevalence values but demonstrate better accuracy as the unknown group rate rises. The MoR-Median and MoR-Huber methods demonstrate unstable trends when the prevalence is below 6%, indicating potential difficulties in handling low unknown group rates. The MoR-HSK2 and MoR-HSK5 techniques consistently yield the highest error values, suggesting these methods may be less suitable for networks modeled by SBM with the specified parameters.

Figures 10 (d)-(f) provide a visual representation of the RMSRE generated by the RoS-type estimators as a function of the prevalence under three distinct connectivity factors $a: e^{-4}, 1, \text{ and } e^4$. Figures SM30(d)-(f) in the Supplementary Material complement this analysis by displaying the |NB| values for these methods when the connectivity factor $a = e^4$. In this case, the non-robust RoS method performs best across the prevalence interval for the three connectivity factors. In networks with a low connectivity factor (where $a = e^{-4}$), the RoS-HQ1 method shows high RMSRE values across all prevalence levels. When the connectivity factor is $a = 1$, the RoS-HQ1 method significantly improves performance, generating the lowest errors for prevalence values exceeding 3%. Similarly, in SBM networks with a connectivity factor of $a = e^4$, RoS-HQ1 shows a substantial decrease in errors. The RoS-HQ2 method performs well for

$a = e^{-4}$ and $a = e^4$ connectivity factors. On the other hand, the MoR-TM10 and MoR-Huber methods gradually reduce their errors for the three connectivity factors as the prevalence increases. Finally, the RoS-Median, RoS-Myriad, RoS-HSK2, and MoR-HSK5 display the more significant error values in SBM networks.

Figures 11(a)-(c) present the RMSRE values obtained by the MoR-type estimators as a function of contamination levels for a prevalence rate of $\rho = 5\%$, under three different connectivity scenarios defined by the parameter a : (a) $a = e^{-4}$, (b) $a = 1$, and (c) $a = e^4$. In addition, Figures SM32(a)-(c) in the Supplementary material depict the absolute normalized bias (|NB|) for the same settings, offering complementary insight into estimator accuracy. Among the robust methods, MoR-TM10, MoR-HQ1, MoR-Huber, and MoR-Myriad consistently demonstrate superior performance across all contamination levels and connectivity factors. Their ability to maintain low RMSRE values, even as contamination rises and group connectivity diverges, highlights their robustness under structurally heterogeneous conditions. These results indicate that such methods are well-suited for real-world network settings with uneven contact probabilities between groups.

Figures 11(d)-(f) display the RMSRE results for the RoS-type estimators across contamination levels, again with prevalence fixed at $\rho = 5\%$ and connectivity factors set to (d) $a = e^{-4}$, (e) $a = 1$, and (f) $a = e^4$. Among the robust RoS variants, RoS-TM10 and RoS-Huber estimators consistently exhibit lower RMSRE values, demonstrating robustness to contamination across the different structural configurations. Finally, Figure SM31 in the Supplementary Material depicts the RMSRE obtained by the robust NSUM approaches against the connectivity factor for different prevalences. In essence, this figure displays the response of robust NSUM methods to the barrier effect for SBM graphs.

4.3. Discussion

In this section, we provide an overview of the main findings regarding the performance of the NSUM estimation methods for different network models and contamination scenarios. First, the non-robust MoR and RoS estimators demonstrate the best performance in terms of RMSRE and NB metrics in environments free of contamination across the tested prevalence ranges for the three network models. This superior performance suggests that the non-robust MoR and RoS methods are the most effective approaches for estimating unknown group rates within networks with different characteristics in uncontaminated scenarios.

Among the MoR-type robust methods, the MoR-Myriad approach consistently exhibits superior performance across tested prevalence levels for the three network models when data is uncontaminated. Specifically, the accuracy generated by the MoR-Myriad method is comparable to those obtained by non-robust approaches, which is essential for practical applications where minimal contamination may occur. In addition, the MoR-HQ1, MoR-Huber, and MoR-TM10 methods improve performance as prevalence increases. However,

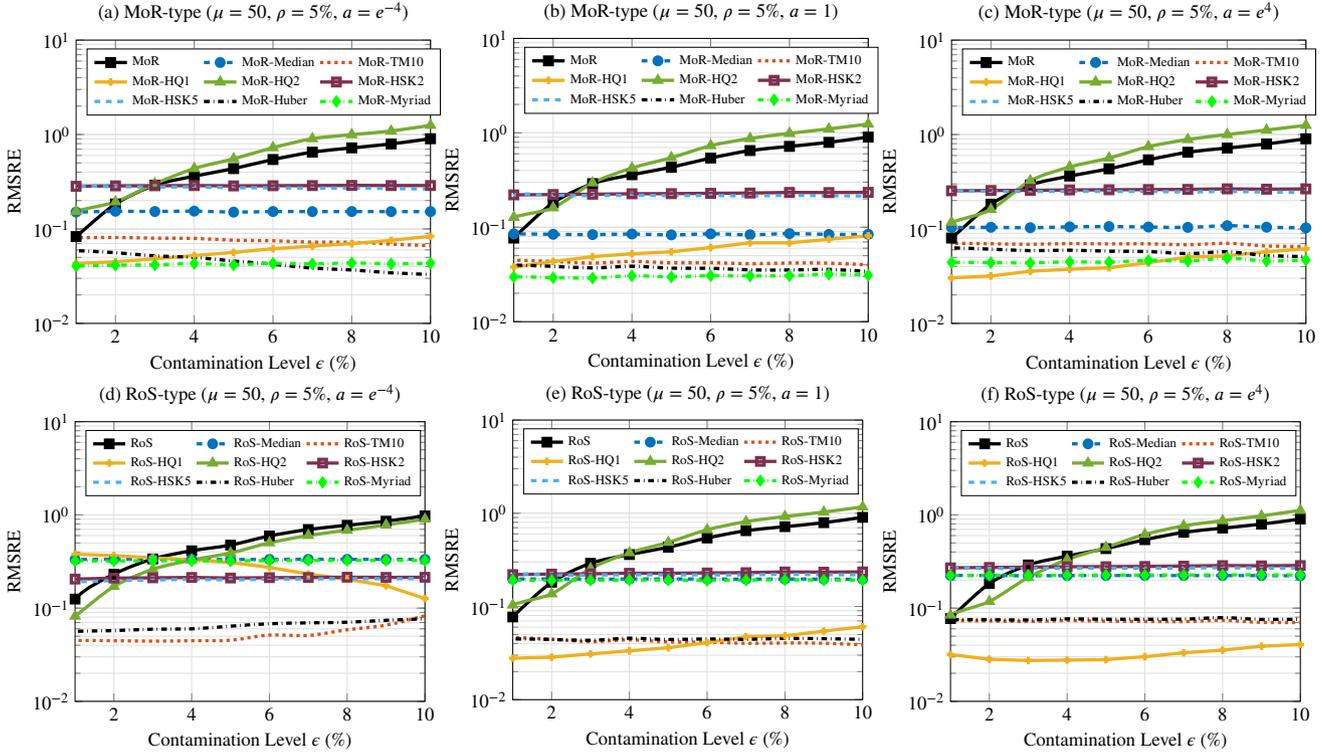


Figure 11: Stochastic Block Model ($\gamma = 2.50$). RMSRE versus the contamination level for MoR-type methods, prevalence $\rho = 5\%$ and connectivity factors: (a) $a = e^{-4}$, (b) $a = 1$, (c) $a = e^4$; RoS-type methods, prevalence $\rho = 5\%$ and connectivity factors: (d) $a = e^{-4}$, (e) $a = 1$, (f) $a = e^4$.

these methods experience a significant decline in performance at low prevalence values, making them less suitable for estimating the size of small unknown groups. In robust RoS-type methods, the RoS-HQ1 estimator behaves similarly to non-robust approaches when prevalence exceeds 3%. This performance indicates a significant degree of reliability in scenarios without contamination but also highlights a potential limitation: RoS-type methods may not be as effective in networks with hard-to-reach unknown groups, especially in prevalence levels below 3%.

Non-robust methods severely degrade their performance in contaminated scenarios as the data noise level increases. Therefore, finding a robust alternative that enables accurate estimations when different kinds of perturbations corrupt data is necessary. For robust MoR-type methods, the MoR-Myriad approach consistently performs best for the three network models, different prevalence values, and contamination levels. Therefore, we can select the MoR-Myriad operator as a reliable option to estimate the unknown group rate in contaminated scenarios. Furthermore, MoR-HQ1, MoR-Huber, and MoR-TM10 are also efficient methods compared to non-robust approaches with contamination, in particular, for prevalence values exceeding 5%. In the case of the RoS-type robust method, RoS-HQ1, RoS-TM10, and RoS-Huber perform best in contaminated scenarios for the ER model. However, these methods degrade their effectiveness

for the SF model. Hence, in contaminated scenarios, RoS-type methods should consider both the network structure and the unknown group characteristics.

5. Case Study

This section evaluates the performance of the NSUM estimators in real scenarios. For that purpose, we use two datasets containing ARD from surveys about COVID-19 cases and voting intention in Spanish elections.

The dataset of COVID-19 exhibits skewness distributions and a type of contamination similar to our simulation study, with contamination at the extreme of the distribution. In contrast, the election dataset provides a particularly unfavorable scenario to validate our approach. Due to political polarization, extreme values—such as participants knowing only contacts with the same political affiliation—are more likely to occur.

This section is structured as follows. First, we introduce the COVID-19 dataset and present the results of the NSUM estimator. Next, we examine the robust NSUM through voting intention data. Finally, we conclude with a summary of key findings.

5.1. The United Kingdom COVID-19 dataset

We use the COVID-19 dataset of Ramirez et al. (2023), consisting of responses from an online survey in the United Kingdom during the period of January 18-26, 2023. The

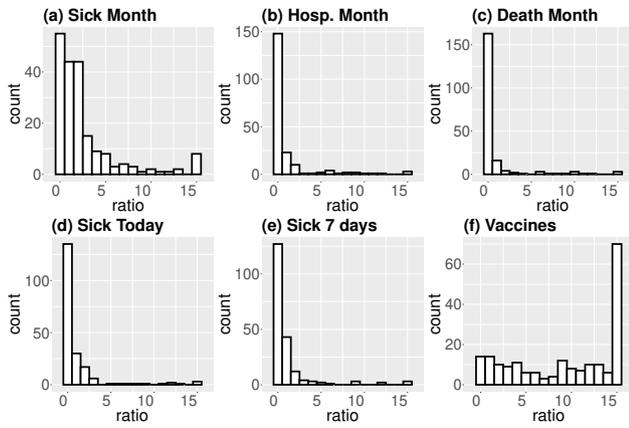


Figure 12: Histograms of ARD responses for (a) Sick Month, (b) Hospitalization Month, and (c) Death Month (d) Sick Today, (e) Sick 7 Days, and (c) Vaccines.

Table 2

Estimated percentages of COVID-19 incidence and mortality for UK.

	sick month	hosp. month	deaths month	sick today	sick 7d	vacc.
Official	9.663	0.044	0.005	1.458	1.116	88.2
MoR	17.367	6.767	5.800	7.000	6.767	62.433
RoS	17.367	6.767	5.800	7.000	6.767	62.433
MoR-Median	13.333	0.000	0.000	0.000	0.000	73.333
MoR-TM10	11.875	1.708	0.750	2.500	2.625	65.292
MoR-HQ1	14.161	3.478	2.153	3.395	3.602	67.246
MoR-HQ2	19.710	8.406	7.250	8.696	8.406	63.768
MoR-HSK2	8.654	0.500	0.000	<u>1.042</u>	<u>1.375</u>	70.683
MoR-HSK5	8.820	0.621	0.000	1.159	1.491	67.826
MoR-Huber	11.409	0.000	0.000	0.000	0.000	62.669
MoR-Myriad	10.052	2.158	1.399	2.862	3.086	84.370
RoS-Median	13.333	0.000	0.000	0.000	0.000	73.333
RoS-TM10	11.875	1.708	0.750	2.500	2.625	65.292
RoS-HQ1	14.161	3.478	2.153	3.395	3.602	67.246
RoS-HQ2	19.710	8.406	7.250	8.696	8.406	63.768
RoS-HSK2	8.654	0.500	0.000	<u>1.042</u>	<u>1.375</u>	70.683
RoS-HSK5	8.820	0.621	0.000	1.159	<u>1.491</u>	67.826
RoS-Huber	11.409	0.000	0.000	0.000	0.000	62.669
RoS-Myriad	6.698	0.054	<u>0.033</u>	0.081	0.108	99.900

survey contains information about vaccinations, hospitalizations, fatalities, and known cases of acquaintances in several time slots. The main characteristic of this dataset is that the degree is fixed in advance to the 15 closest contacts, based on the good-friends support group of Dunbar's theory (Dunbar, 2010). Another characteristic of this data is the extreme skewness of the variables (Ramirez et al., 2023). We observe in Figure 12 that their values concentrate at one extreme of the distribution. Some of these variables, such as hospitalizations and contacts sick in the last month, which have a small prevalence in the population (see Table 2), show responses close to the degree of the nodes (15), suggesting malicious responses or an overrepresentation of people affected by the illness

Table 2 displays the results obtained in the UK dataset. The table columns correspond to the percentage of monthly infection cases, monthly hospitalizations, fatalities in a month, daily infections, weekly infections, and vaccinations. Specifically, the bold numbers in the table are the best performing results, and the underlined values are the second best results, with respect to the official data. The official data was obtained from the *Office for National Statistics*

(ONS)¹. Bold numbers represent the best method for each variable, and underlined values correspond to the second-best method.

The non-robust estimators, MoR and RoS, consistently overestimate small-sized groups while underestimating the larger group, vaccination. Overall, these estimators perform worse across all variables compared to the robust estimators, except for MoR-HQ2 and RoS-HQ2. MoR-Myriad is the best estimator when the size of the targeted group is not small, particularly in the case of vaccination, where it outperforms other methods. However, the performance of MoR-Myriad diminishes for smaller groups. In such cases, MoR-HSK2, RoS-HSK2, MoR-HSK5, RoS-HSK5, and RoS-Myriad yield the best results.

The adaptive trimmed means, with the exception of MoR-HQ2 and RoS-HQ2, are better than the classical estimators. While MoR-HQ1, RoS-HQ1, MoR-HQ2, and RoS-HQ2 do not perform as well as the non-adaptive MoR-TM10 and RoS-TM10; MoR-HSK2, RoS-HSK2, MoR-HSK5, and RoS-HSK5 offer clear improvements over MoR-TM10 and RoS-TM10 across all variables. Notably, MoR-HSK2, RoS-HSK2, MoR-HSK5, and RoS-HSK5 prove to be the best estimators in the low prevalence groups, such as weekly and daily infection rates, as well as monthly fatalities.

MoR-Median and RoS-Median yield relatively better results for vaccination prevalence compared to most other methods. However, when the prevalence decreases, these estimators return zero prevalence estimates. The same issue arises with MoR-Huber and RoS-Huber. In addition, we observe that MoR-type methods yield values identical to the corresponding RoS-type methods, which use the ratios of the location estimator, with the exception of MoR-Myriad and RoS-Myriad.

5.2. Spanish Elections Dataset

We employ a dataset of voting intention in Spanish general elections in July 2023 (Arealillo et al., 2024). This dataset includes the ARD of the 4 main political groups and the ARD associated with the blank vote and other parties. These parties are *Partido Popular* (PP), *Partido Socialista Obrero Español* (PSOE), *Vox*, and *Sumar*. This case is an example containing barrier effects since the participants are more likely to form connections with people with the same political views. While participants were not directly asked for their degree, we derive this by summing their responses since the questions are disjoint.

Table 3 shows the estimated percentage for each of the four most popular political parties. Columns represent the political parties. The real data corresponds to the published official data.

The results indicate that non-robust NSUM estimators generally outperform robust methods in predicting voting intentions. Specifically, MoR performs best across the four main parties, excelling in Sumar and ranking second for PSOE. Certain robust MoR-type estimators show strong performance for specific parties but fall short for others. For

¹<https://www.ons.gov.uk/>.

Table 3

Estimated percentages of vote intention for the different parties in Spain.

	PP	PSOE	Vox	Sumar
Real	40.2	27.6	13.9	15.3
MoR	32.1	27.0	15.4	15.2
RoS	31.7	20.6	14.4	11.5
MoR-Median	25.0	25.0	8.3	8.7
MoR-TM10	28.9	24.9	11.3	11.3
MoR-HQ1	31.4	26.8	<u>13.0</u>	13.0
MoR-HQ2	<u>35.3</u>	29.3	17.2	<u>17.1</u>
MoR-HSK2	24.7	21.8	7.6	7.4
MoR-HSK5	25.9	22.1	8.1	7.7
MoR-Huber	30.2	25.8	11.5	11.4
MoR-Myriad	24.0	23.8	9.1	8.8
RoS-Median	27.8	27.8	5.6	11.1
RoS-TM10	24.4	23.1	9.6	9.8
RoS-HQ1	16.6	15.9	6.8	7.1
RoS-HQ2	36.8	22.3	16.8	12.6
RoS-HSK2	22.8	21.5	7.5	8.2
RoS-HSK5	22.8	22.4	7.5	8.3
RoS-Huber	26.2	25.9	7.7	10.3
RoS-Myriad	11.8	11.8	0.2	0.1

instance, MoR-HQ1 ranks second for Vox but diverges for Sumar, while MoR-HQ2 performs well for PP and Sumar but tends to overestimate for PSOE and Vox. Among RoS-type estimators, RoS generally surpasses its robust alternatives, except for RoS-HQ2. Notably, some robust MoR alternatives, like MoR-TM10, MoR-HQ1, and MoR-HQ2, outperform RoS in several cases. Additionally, we note that the voting intention for PP, shown in the first column, is consistently underestimated, with only RoS-HQ2 coming close to the real outcome. However, RoS-HQ2 is worse than MoR because of the deviations for other parties.

In conclusion, MoR emerges as the most effective estimator across all political parties, with no robust MoR-type method outperforming it. The closest alternative is MoR-HQ2, which provides the second-best predictions for two political parties. RoS demonstrates an inferior performance in comparison with MoR and the MoR robustifications. While RoS-HQ2 offers a performance improvement over RoS, especially being the only one good at PP, it does not surpass several robust MoR-type estimators.

5.3. Discussion

The results of the COVID-19 dataset show that the robust estimators perform better in contaminated scenarios. This dataset has a marked presence of outliers (Ramirez et al., 2023), resulting in poor performance of MoR and RoS. The performance of the robust estimators is lower with respect to the simulation study. However, the results are not different if you compare them with each other, especially for groups with non-small prevalence, such as monthly, weekly, and daily infections and vaccinations. For small prevalence, the adaptive methods are better. However, one possible cause is the lack of precision of these methods since they estimate

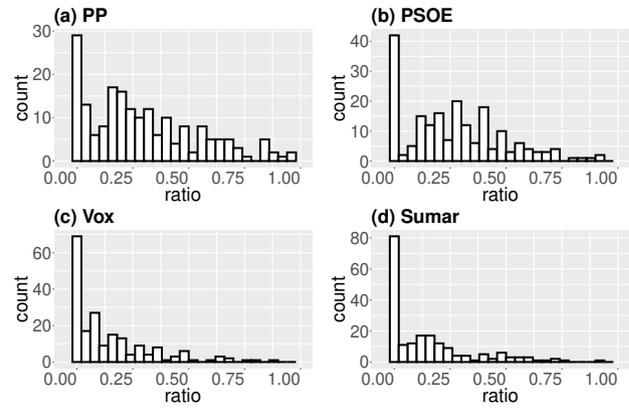


Figure 13: Histograms of ratios (a) PP, (b) PSOE, (c) Vox, and (d) Sumar.

the prevalence as 0. Additionally, the performance of MoR-Median, and RoS-Median, provide relatively similar results. The MoR-Huber and RoS-Huber estimate all the groups as zero, indicating a possible more conservative behavior than the Myriad, focusing more on safety.

We observe that the values for MoR-type methods are equal to the corresponding RoS-type method, which uses the ratios of the location estimator, except for MoR-Myriad and RoS-Myriad. This consistency arises because the degree is fixed at 15, causing location estimators to return identical values when all input values are constant. In addition, the MoR-Myriad estimator performs well for groups without small prevalence, as in the simulations. However, in this case, MoR-Myriad suffers with small groups. Due to the small linearity parameter ($K = 0.2$), the Myriad operator approximates the mode (Arce, 2004), particularly in datasets with low variability. This may lead to a behavior similar to the RoS-Median, approximating to the closest integer. This may explain the bad performance of MoR-Myriad and RoS-Myriad, as well as the discrepancy between their estimates.

The UK COVID-19 dataset shows that some adaptive estimators are resistant to deviations in skewness and contamination. Notably, MoR-HSK2 and MoR-HSK5 seem more adequate than the other adaptive estimators, having good performance compared to the previous simulations, presenting closer values while the simulations underestimate consistently. This contrasts with the fact that MoR-HQ1 performs poorly. The estimators MoR-HSK2 and MoR-HSK5 use the factors that involve the smallest and largest values, in addition to the mean and median. One possible factor is that the smallest and largest values are limited to 15. Another factor is that MoR-HQ1 and MoR-HQ2 use sums over a proportion of the smallest and largest. Therefore, they may consider more outlying values when calculating the trimming proportions.

The analysis of the voting intention dataset reveals that the non-robust estimators are better in the presence of barrier effect, conclusions that align with our findings with the simulation experiments. Specifically, the non-robust MoR estimator consistently outperforms its corresponding robust

MoR-type alternatives, while the robust RoS-type estimators generally perform worse than the non-robust RoS operator, except for RoS-HQ2. This observation corroborates our simulation results, where robust RoS-type methods, except RoS-HQ2, performed poorly under high assortativity and prevalence. Importantly, the results from real data suggest that the improvements seen in RoS-HQ2 during the simulations may become even more pronounced, potentially surpassing RoS in scenarios with higher prevalences. In this example, the variables have distributions with values on the extremes that are less likely to be produced by contamination. Identifying whether the extreme values are due to contamination is much harder. This may be the reason why the estimators underestimate all the groups.

Another feature in the Spanish elections is the outperforming of MoR with respect to RoS, which is consistent with our previous theoretical analysis. RoS assumes that every contact has the same possibility of knowing someone in the unknown group, but regarding political opinions, this rarely happens. In addition, the methods based on the ratios of the location estimators perform worse than the procedures based on the ratios of the responses. Indeed MoR-HQ1 and MoR-HQ2 are better than RoS, and MoR-TM10 and MoR-Huber are very close to RoS.

Another relevant fact is that even MoR-Myriad yields bad estimates. In this dataset, people of one party are more likely to have all their acquaintances voting for the same political option. This produces distributions that differ from the SBM distributions of Figure 3. We observe in Figure 13 that, unlike high prevalences, there are numerous participants whose social circles do not include anyone voting for a particular party. Conversely, some social circles tend to vote exclusively for a single party. Our robust proposals are based on giving less weight to extreme values. However, these extreme values are meaningful, and giving them less weight is counterproductive. The better performance of the RoS-HQ2 with respect to other robust estimators may be produced because it gives more weight to the extreme data, and this feature is the cause of the overestimation in the simulations, especially in the contaminated scenarios. MoR-HQ2 employs the same statistic for the trimming proportions, but in this case, the values of the ratios are bounded.

The real data suggests that, in the absence of a barrier effect, MoR-Myriad is the preferred estimator for groups with significant prevalence. However, MoR-Myriad struggles with small prevalence rates and networks of lower degrees, as the discrete nature of responses complicates accurate estimation. In these cases, alternatives such as MoR-HSK2, MoR-HSK5, RoS-HSK2, and RoS-HSK5 may be better options. However, due to the poor performance of these estimators in the simulation, it may be good to select other estimators, such as MoR-TM10 to safeguard against bad estimates. In the presence of barrier effects in groups with high prevalence, where participants may have all or none of their acquaintances belonging to the same group, leading to meaningful extremes of the distribution,

impacting our proposals. In this case, the MoR is the preferred option, while the RoS and RoS-type robustifications suffer from bad performance since their assumptions require the absence of barrier effects. For a combination of barrier effect and contamination, the selected estimates would be a robustification of MoR such as MoR-TM10, MoR-Myriad, or MoR-Huber, balancing strong simulation results with reasonable results in real data.

6. Conclusions

This work focused on the area of robust statistics, analyzing the behavior of NSUM techniques in the face of outliers and unusual data. The purpose of this work was to study the reliability and accuracy of the methods, which is highly relevant in fields as diverse as finance, bioinformatics, and environmental sciences, since real-world data often do not meet the ideal conditions assumed by many NSUM techniques. This analysis of statistical robustness in NSUM techniques has shown its significant impact, as illustrated with simulations. Real-world cases have been presented with COVID-19 and voting intention datasets.

This work proposes the first robustifications of the classical NSUM estimators most frequently employed in the literature. We propose robustifications substituting the sample means of the classical estimators with location estimators. To assess their effectiveness, we conduct a series of experiments in networks with real-world properties. Specifically, we employ Erdős-Rényi networks as a suitable model for classical NSUM, analyzing deviations varying prevalence levels. We include simulations in Scale-Free networks, providing insights into the effects of skewed distributions with heavy tails. We evaluate the performance of the estimators under contamination, simulating participant misreporting by replacing a fraction of unknown group ARD values with zeros or personal network sizes. We also measure the barrier effect through the Stochastic Block Model, considering scenarios in which the members of the unknown group have different propensities to form ties with other groups. Finally, we validate our findings through a comparison of NSUM estimators on real datasets.

The main findings regarding the robust NSUM estimators across various scenarios are as follows. Firstly, non-robust estimators consistently outperform robust approaches across all network models and tested prevalence levels in uncontaminated environments. These results indicate that non-robust estimators are particularly effective in scenarios without contamination, delivering optimal performance for different network structures. In contrast, in networks with noisy data, the MoR-Myriad method stands out by outperforming other estimation methods across various prevalence values and contamination levels. Its robustness makes the MoR-Myriad method especially valuable in situations where external noise may distort data, particularly at low prevalence levels. Additionally, the techniques MoR-TM10, MoR-Huber, RoS-Huber, and RoS-TM10 show improved performance as prevalence levels increase, both in contaminated and uncontaminated environments. This

trend highlights the potential limitations of these methods when estimating the sizes of hard-to-reach groups. Finally, sample-based methods such as MoR-Median and RoS-Median demonstrate instability, especially at low prevalence levels.

This work has some limitations. Firstly, conclusions are based primarily on simulation environments, and although we have validated experiments with two datasets, these may not entirely capture the complexity of real-world distributions, biases, and contamination. Another limitation is the use of undirected graphs, which assume reciprocal relationships, potentially oversimplifying the asymmetric nature of some real-world social networks. Moreover, while our results show that robust estimators perform consistently well in various scenarios, we recognize that using fixed parameter values may limit their full capabilities. In this study, we utilized standard settings for key estimators (Myriad, Huber, and trimmed mean), which might not be optimal for the specific types of contamination we examined. For example, although there are parameter selection strategies available for α -stable noise models (Gonzalez and Arce, 2001), our research focuses on a different type of impulsive contamination. Therefore, a systematic investigation into the effects of these hyperparameters and the potential for tuning them to enhance performance remains an important area for future research.

To the best of our knowledge, this is the first study to introduce robust NSUM estimators and provide a comprehensive comparative evaluation of their performance. The first further work is to explore robust versions of traditional NSUM applications where participants' degrees are unknown, requiring estimation from responses related to groups with known prevalence. Another future alternative consists of the analysis of the contamination of the responses of the reported network size. Further studies could focus on additional contamination types, such as contamination in one extreme of the distributions, in a neighborhood of the data, or representing the transmission and recall error. Given the impact of prevalence, network characteristics, and contamination on estimator choice, further development is needed to tailor robust estimators to these specific features. Another further work is the proposal of methods that give more weight to extreme values in scenarios with a high barrier effect. Since many NSUM studies focus on low-prevalence groups, future research on estimators resistant to very low prevalence rates is also warranted. Finally, a subsequent line of research is the analysis of the effect of the sample size and the robustness of the procedures designed for selecting it (Josephs et al., 2024).

Declaration of interest statement

The authors have no conflicts of interest to disclose.

Code availability

The source code used for the simulations can be downloaded from the following link:

https://github.com/JuanMarcosRamirez/Robust_NSUM_IMDEA.

References

- Agostinelli, C., Basu, A., Filzmoser, P., Mukherjee, D., et al., 2016. Recent advances in robust statistics: theory and applications. Springer.
- Arce, G.R., 2004. Nonlinear signal processing: a statistical approach. John Wiley & Sons.
- Arevalillo, J.M., Ramírez, J.M., Díaz-Aranda, S., Aguilar, J., Fernández Anta, A., Lillo, R.E., 2024. Network scale-up methods on aggregated relational data to estimate the outcome of elections. (Submitted article).
- Cheng, S., Eck, D.J., Crawford, F.W., 2019. Estimating the size of a hidden finite set: large-sample behavior of estimators. URL: <https://arxiv.org/abs/1808.04753>, arXiv:1808.04753.
- Díaz-Aranda, S., Aguilar, J., Ramírez, J.M., Rabanedo, D., Fernández-Anta, A., Lillo, R., 2024. Performance analysis of nsum estimators in social-network topologies. *The American Statistician* 0, 1–18. doi:10.1080/00031305.2024.2421361.
- Dunbar, R., 2010. How many friends does one person need? Dunbar's number and other evolutionary quirks. Harvard University Press.
- Feehan, D.M., Salganik, M.J., 2016. Generalizing the network scale-up method: a new estimator for the size of hidden populations. *Sociological methodology* 46, 153–186.
- García-Agúndez, A., Ojo, O., Hernández-Roig, H.A., Baquero, C., Frey, D., Georgiou, C., Goessens, M., Lillo, R.E., Menezes, R., Nicolaou, N., et al., 2021. Estimating the covid-19 prevalence in spain with indirect reporting via open surveys. *Frontiers in Public Health* 9, 658544.
- Goh, K.I., Kahng, B., Kim, D., 2001. Universal behavior of load distribution in scale-free networks. *Physical review letters* 87, 278701.
- Gonzalez, J., Arce, G., 2001. Optimality of the myriad filter in practical impulsive-noise environments. *IEEE Transactions on Signal Processing* 49, 438–441. doi:10.1109/78.902126.
- Habecker, P., Dombrowski, K., Khan, B., 2015. Improving the network scale-up estimator: Incorporating means of sums, recursive back estimation, and sampling weights. *PLoS one* 10, e0143406.
- Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W., 2011. Robust Statistics: The Approach Based on Influence Functions. Wiley Series in Probability and Statistics, Wiley. URL: <https://books.google.es/books?id=XK3uhrVefXQC>.
- Hogg, R.V., 1974. Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association* 69, 909–923.
- Huber, P.J., Ronchetti, E.M., 2009. Robust statistics.
- Jing, L., Lu, Q., Cui, Y., Yu, H., 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.
- Josephs, N., Feehan, D.M., Crawford, F.W., 2024. A sample size formula for network scale-up studies. *Sociological Methods & Research* 53, 1252–1289.
- Keselman, H., Wilcox, R.R., Lix, L.M., Algina, J., Fradette, K., 2007. Adaptive robust estimation and testing. *British Journal of Mathematical and Statistical Psychology* 60, 267–293.
- Killworth, P.D., Johnsen, E.C., McCarty, C., Shelley, G.A., Bernard, H.R., 1998a. A social network approach to estimating seroprevalence in the united states. *Social networks* 20, 23–50.
- Killworth, P.D., McCarty, C., Bernard, H.R., Shelley, G.A., Johnsen, E.C., 1998b. Estimation of seroprevalence, rape, and homelessness in the united states using a social network approach. *Evaluation review* 22, 289–308.
- Kunke, J.P., Laga, I., Niu, X., McCormick, T.H., 2024. Comparing the robustness of simple network scale-up method (nsum) estimators. URL: <https://arxiv.org/abs/2303.07490>, arXiv:2303.07490.
- Laga, I., Bao, L., Niu, X., 2021. Thirty years of the network scale-up method. *Journal of the American Statistical Association* 116, 1548–1559.

- Laga, I., Bao, L., Niu, X., 2023a. A correlated network scale-up model: Finding the connection between subpopulations. *Journal of the American Statistical Association* , 1–18.
- Laga, I., Kunke, J., McCormick, T., Niu, X., 2023b. The role of scaling and estimating the degree ratio in the network scale-up method. *arXiv preprint arXiv:2305.04381* .
- Maltiel, R., Raftery, A.E., McCormick, T.H., Baraff, A.J., 2015. Estimating population size using the network scale up method. *The annals of applied statistics* 9, 1247.
- Maronna, R.A., Martin, R.D., Yohai, V.J., Salibián-Barrera, M., 2019. *Robust statistics: theory and methods (with R)*. John Wiley & Sons.
- McCormick, T.H., 2020. The network scale-up method. *The Oxford Handbook of Social Networks* , 153.
- Newman, M., 2018. *Networks*. Oxford university press.
- Nyarko-Agyei, A., Moser, S., Seymour, R.G., Brewster, B., Li, S., Weir, E., Landman, T., Wyman, E., Torres, C.B., Fell, I., et al., 2024. A partially pooled nsum model: Detailed estimation of csem trafficking prevalence in philippine municipalities. *arXiv preprint arXiv:2407.13267* .
- Ocagli, H., Azzolina, D., Lorenzoni, G., Gallipoli, S., Martinato, M., Acar, A.S., Berchiolla, P., Gregori, D., Group, I.S., 2021. 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* 18, 5713.
- Ramirez, J.M., Diaz-Aranda, S., Aguilar, J., Ojo, O., Lillo, R.E., Fernandez Anta, A., 2023. A Snapshot of COVID-19 Incidence, Hospitalizations, and Mortality from Indirect Survey Data in China in January 2023. *medRxiv* , 2023–02.
- Ramirez, J.M., Paredes, J.L., 2016. Recursive weighted myriad based filters and their optimizations. *IEEE Transactions on Signal Processing* 64, 4027–4039.
- Ramirez, J.M., Paredes, J.L., 2017. Recursive myriad-mean filters: Adaptive algorithms and applications. *Signal Processing* 139, 12–24.
- Reed III, J., Stark, D., 1996. Hinge estimators of location: Robust to asymmetry. *Computer Methods and Programs in Biomedicine* 49, 11–17.
- Staudte, R.G., Sheather, S.J., 2011. *Robust estimation and testing*. John Wiley & Sons.
- Sully, E., Giorgio, M., Anjur-Dietrich, S., 2020. Estimating abortion incidence using the network scale-up method. *Demographic Research* 43, 1651–1684.
- Wilcox, R.R., 2011. *Introduction to robust estimation and hypothesis testing*. Academic press.
- Zheng, T., Salganik, M.J., 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.