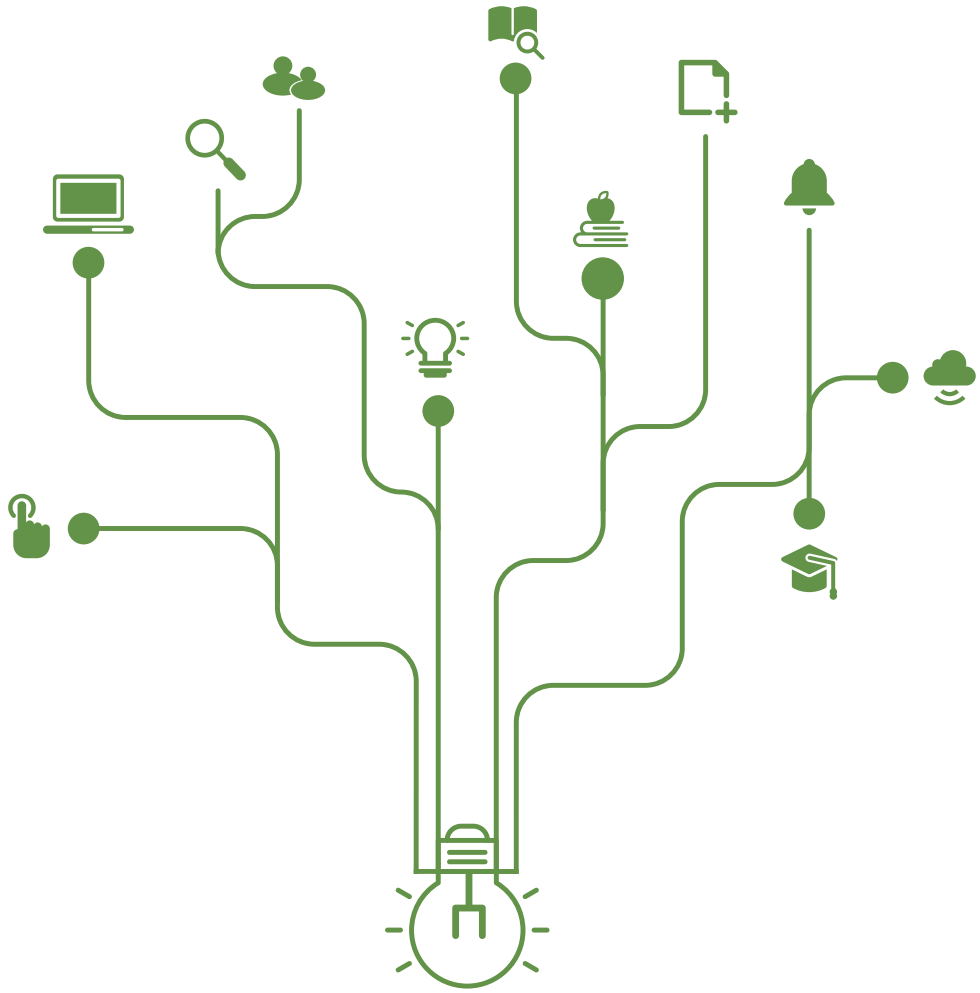
Development Studies Series DP21-14

The Network Labeling Optimization for Hidden Population Size Estimation: A Case Solution for the Bangladesh kidney Sellers Problem

Narae Lee (George Mason University)
Abu Bakkar Siddique (George Mason University)
Meng-Hao Li (George Mason University)
Manzur Ahmad (EXIM Bank Agricultural University Bangladesh)
Tariq Lutfay (Dhaka University)
Reaz Haque (Dhaka University)
Hadi El-Amine (George Mason University)
Naoru Koizumi (George Mason University)



KDI SCHOOL
KDI School of Public Policy and Management

The Network Labeling Optimization for Hidden Population Size Estimation:

A Case Solution for the Bangladesh kidney Sellers Problem

Narae Lee[1] *, Abu Bakkar Siddique[1] *, Meng-Hao Li[1] *, Manzur Ahmad[2], Tariq Lutfay[3], Reaz Haque[3], Hadi El-Amine[4], Naoru Koizumi[1]

[1]Schar School of Policy and Government, George Mason University, [2]Business Administration, EXIM Bank Agricultural University Bangladesh, [3]Department of Development Studies, Dhaka University, [4]Volgenau School of Engineering, George Mason University
**The three authors equally contributed.*

Abstract

Estimating the prevalence of hidden population is a challenging but important task for policymakers. Without knowing the precise scale of the problem, it is difficult to design a sharp remedy. Existing tools such as facility-based sentinel surveillance, snowball sampling, respondent-driven sampling, and network scale-up methods are prone to respondents' misinformation, false responses, and sample misrepresentation. Therefore, this paper proposes a novel analytical framework to overcome such weaknesses and derive better estimates. Specifically, our optimization-based mathematical model employs the Integer Programming (IP) and Social Network Analysis (SNA) to directly remove double-counting from the survey of more accessible subjects of the general public. To validate the model, the study implemented a survey on kidney trafficking in the kidney selling hotspot of Bangladesh. Reflecting the survey responses of 400 residents in a Ward of one Union in Kalai Upazila, we simulated an Exponential Random Graph Models (ERGMs) driven network. Although the model validation using the simulated network showed some signs of over-representation, a secondary validation using other data showed that the model estimates are fairly accurate.

Keywords: hidden population, social network, network density, sampling, illegal behavior, integer-programing optimization

1. Introduction

Estimating the prevalence of hidden population has been a challenging task for researchers and policymakers. Not only that information asymmetry hinders the detection of self-hiding people, but it also poses potential harm to investigators as the population may be involved in illegal, stigmatizing, or even criminal activities. Nonetheless, the size estimation of hidden population is important to combat several policy problems, especially from the resource allocation perspective. One of such examples that we address in this paper is kidney sales, in which many poor people are victimized or facing a constant threat from illicit businesses. While the kind of difficulty applies to the broader problems of general human trafficking, sexual violence, and drug abuse, kidney sales has emerged to be a significant global security issue particularly in recent years.

There are several existing approaches for hidden population size estimation. Facility-based sentinel surveillance, snowball sampling (Frank & Snijders, 1994; Goodman, 1961), respondent-driven sampling (RSD) (Heckathorn, 1997, 2002), network approach (Crawford et al., 2018), and network scale-up approach (Bernard et al., 2010; Shelton, 2015) are among them. However, these approaches are prone to respondents' misinformation, false responses, and sample misrepresentation. These weaknesses could result in the inaccurate measure of the problem, failing to produce quality data for scientific research.

To potentially improve these methodological issues, this paper introduces a novel analytical framework for hidden population size estimation. Combining the Integer-Programming (IP) Optimization and Social Network Analysis, our methodology aims to overcome some of the existing approaches' weaknesses. Our approach differs from the conventional methods on two accounts: First, our technique relies not on the information given by the members of the hidden population but also on the information provided by the survey respondents who are general public. The non-affiliated reports the number of hidden population members s/he knows in one's community. Although our methodology is an extension of network scale-up methods in this regard, our model also directly detects the multiple counting from different respondents instead of averaging several survey outcomes. The direct removal of multiple counting from the reported survey reduces respondent bias. Lastly, our methods allow creating a sample network that can better represent the population network.

To demonstrate our model for the hidden population size estimation, we administered a survey on the kidney sales problem in one ward of Bangladesh's northwestern region, where the kidney sales issue is prevalent. The actual field survey of nearly 400 residents is another contribution of our paper, with which we created the testing bed for the methods in realistic settings.

In section two, we describe the current situation of kidney sales with the problem's unique features making the size estimation difficult. Section three discusses the existing approaches to hidden population size estimation and introduces our new model. Section four elaborates on our survey design and implementation, and section five discusses the network simulation methods and the network properties of the outcomes. Lastly, we report our findings in section six and discuss the limitations and possible remedies we will test forward in the conclusion.

2. Kidney Sales Problem and the Challenges in the Size Estimation

Recent evidence suggests that Middle Eastern criminal organizations exploit refugees who sell their kidneys to pay for their passage (Fraser & Koizumi, 2017). Similarly, a recent web scaping analysis from the South Asian regional newspapers and applying machine learning approaches shows that the network of kidney trading hubs exists in South Asia that not only serves the kidney demand for regional buyers but also attracts buyers from all over the world (Li et al., 2021). Ethnographic approaches have also well documented the sufferings of kidney sellers and bio violence in the kidney Bazar in Bangladesh (Moniruzzaman, 2012, 2016, 2019). While these studies provide information that major hubs for kidney sales exist and more of their intra- and inter-networking, it does not provide detailed information about the extent of the problem in each of these countries or in regions. This is because there are no effective methods to estimate the prevalence of the victims who tend to be too hidden.

In contrast to other hidden population such as sex workers, estimating the size of kidney sellers are particularly challenging due to the following reasons: first, we know that the kidney sellers fall in the sub-group of the population that are socially deprived, low-income people, however, we do not know what level of poverty can lead an individual to decide to sell their body parts. Complexities are further added due to the ambiguous definition of poverty and if they reside in low-income economies which tend to have no good census data that can track individual households. Second, these socially deprived groups who are likely to become kidney sellers are often stigmatized, and therefore conventional household surveys are unlikely to produce accurate inspection data. Moreover, since they represent a very small proportion of the general population, obtaining statistically reliable data for such subpopulations through household surveys would additionally need a large sample. Third, there is no such list or sampling frame of the poor population that can be useful since it can be either infeasible or excessively expensive (Magnani et al., 2005). Ethnographic surveys through snowball sampling can be too dangerous to reach those sellers who are often within the surveillance of the trafficking network namely brokers (Moniruzzaman, 2016). It can be also labor-intensive.

Such trafficking like trading human kidneys functions in a particular way capitalizing the social network of traffickers, buyers, sellers, and even the medical facilities that can be too hidden to policymakers and researchers.

3. A New Methodology for Hidden Population Size Estimation

1) The Review of Existing Approaches

Prior research developed several methods to estimate the size of different hidden populations such as the population of human trafficking victims (Cruyff et al., 2017), illegal migrants, drug sellers/users, rape (Killworth et al., 1998), homeless people (Dávid & Snijders, 2002), sexually transmitted disease (STD) (Rubin et al., 1992), human immunodeficiency virus (HIV) (Magnani et al., 2005), and others. These methods include snowball sampling (Goodman, 1961), capture-recapture method (Hook & Regal, 1995), respondent-driven sampling (RDS) (Crawford et al., 2018), and network scale-up method (NSUM) (Maltiel et al., 2015). While these methods can estimate the size of some groups of hidden populations better, they cannot estimate other hidden populations accurately when the target populations are not too hidden and thus the sampled respondents are not amenable to reveal their true status to strangers including researchers/surveyors because their status is associated with stigma as well as penalties or other negative consequences.

The most commonly used technique is to use snowball sampling in which subjects enumerate their social contacts, each of whom enters the study, and the process repeats until either a target sample size has been reached or the sample has become ‘saturated’ (Goodman, 1961). Estimating hidden population size from snowball samples requires homogeneity assumptions about the underlying social network and respondents are required to reveal their networks (Frank & Snijders, 1994). However, participants may decline to enroll in the study, which is more likely to occur for kidney sellers, and thus, the subgraph of respondents may be incomplete, and the estimation of the size of the kidney sellers can be wrong. Moreover, the initial seeds in snowball sampling are supposed to be randomly chosen, but it is often not feasible, especially for kidney sellers. Like other non-probability sampling methods, another drawback of snowball sampling is sampling bias not representing a larger population (Magnani et al., 2005).

Like snowball sampling, respondent-driven sampling, a widely used technique for counting the size of a hidden population (Heckathorn, 1997), also involves chain referral. It suffers from the same limitations as snowball sampling suffers except it does not require all social contacts of the subject to be surveyed. In respondent-driven sampling, the subjects decide which of their contacts to recruit. It can also minimize the influence of initial seeds on the final sample composition more than it is possible for snowball sampling

(Magnani et al., 2005). The initial seeds are interviewed and given a reward for participation and then they receive a small number of coupons that they can use to recruit other qualified members. Coupons are usually marked with a unique ID that allows tracing back to the recruiter. Thus, the recruitment process in the respondent-driven sampling is designed to spread through the social network of the interested hidden population (Salganik & Heckathorn, 2004).

Alternative to the chain referral samplings is the network scale-up method which does not require access to the hidden population. However, its validity relies on subjects' knowledge of their contacts' membership in the target population (Killworth et al., 1998). It surveys members of the general population to determine how many people they know and how many people they know who are members of the hidden population (Bernard et al., 2010; Shelton, 2015). It assumes that the proportion of surveyed people's contacts who are part of the hidden population is equal to the population proportion. Then, the known general population size produces an estimate of the hidden population size after multiplying the proportion (Maltiel et al., 2015; Wang et al., 2015). However, the membership in the hidden population is often not differentiable from the non-members, or groups within the general population may have different probabilities of ties to the hidden population which can preclude the success of applying the network scale-up method (Feehan & Salganik, 2016; Shelley et al., 2006; Zheng et al., 2006). Therefore, a simple network scale-up approach can innate over- or under-representation issues for the hidden population.

In the paper, we have developed an alternative optimization-based mathematical model that does not rely on the aforementioned assumptions. The model aims to estimate the size (N_H) of a hidden population in a given general target population from which the sample is taken. The proposed optimization approach is an extension of the network scale-up methods in that it relies on information from the general public or non-affiliated people. However, the unique contribution of our model is to estimate the size of the hidden population by measuring and addressing potential multiple counting by an Integer-Programming (IP). The following section describes the model we have developed.

2) Our Model: Integer-Programming (IP) for Social Network Mapping

Figure 1 represents a hypothetical network, which includes both the members of the hidden population (nodes 3 and 5) and the general public (nodes 1, 2, 4, 6, 7, and 8). There are two pieces of information available for each node: i) the total number of connections to individuals of the hidden population not including himself (values R_i associated with each node in the figure); and ii) the connections to other members (i.e., the network structure). Suppose that we do not know the actual members of the hidden

population in this hypothetical framework. If the two pieces of information are available for each node, then we can deduce the nodes that belong to the hidden population with certainty.

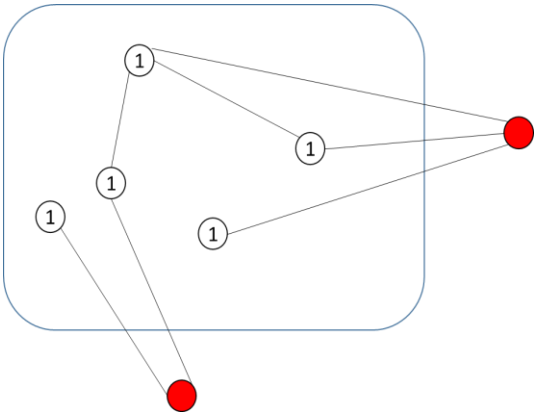


Figure 1. Multiple Counting of the Hidden Population in Network

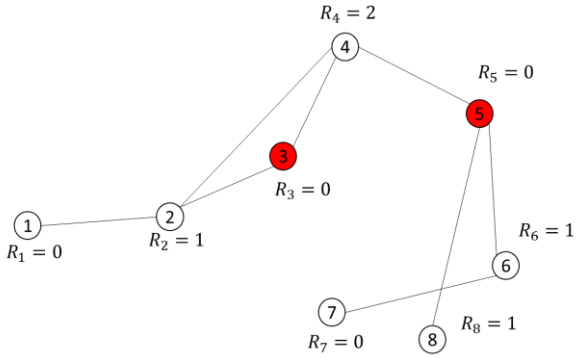


Figure 2. A Hypothesized Network of the Hidden Population and General Public

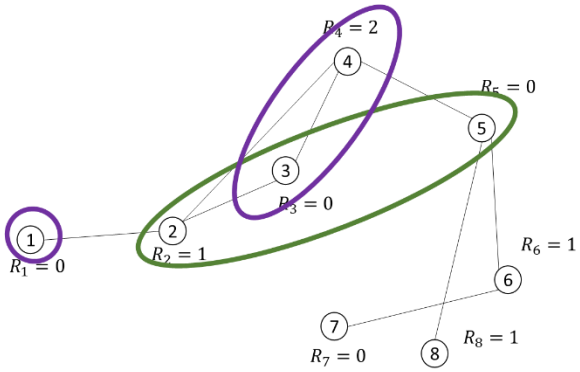


Figure 3. The Illustration of the Hidden Population Member Identification

The basic concept of our approach is described in Figures 2 and 3, which represent the social network of a given target population of size N (in the example network in the figures $N = 8$), each node is represented by a label $1, \ldots, N$. If two nodes, say i and j, know each other then a link (i, j) connects the two nodes (in the example networks there are 8 links in the network). In Figure 3, the members of the hidden population are unknown, and thus there are no red nodes displayed. However, we know that, based on the response given by node four, two of the nodes in the green circle must belong to the hidden population. Similarly, one of the nodes in the purple circles must be the member of the hidden population because of the response of node 2. In this fashion, we can identify analogous conditions based on the responses of other nodes. In this example, two nodes that satisfy these conditions are the red nodes (node 3 and 5) in Figure 2.

The optimization model uses the knowledge of the social network structure and the values R_i to assign labels x_i to each node such that $x_i = 1$ if person i in the target population belongs to the hidden population, and $x_i = 0$ otherwise. Once such an assignment is determined then the size of the hidden population can be computed as

$$N_H = \sum_{i=1}^{N} x_i$$

The following optimization model looks for such an assignment:

$$\text{minimize} \quad \sum_{i=1}^{n} |\delta_i| \tag{1}$$

$$\text{subject to} \quad \delta_i = R_i - \sum_{j \in N(i)} x_j \qquad \forall i \in \{1, \cdots, n\} \tag{2}$$

$$x_i \in \{0,1\} \qquad \forall i \in \{1, \cdots, n\} \tag{3}$$

$$\delta_i \in \mathbb{R} \qquad \forall i \in \{1, \cdots, n\} \tag{4}$$

In the formulation, the set $N(i)$ denote the neighbors of node i, $i \in \{1, \cdots, n\}$, that is the set of nodes that are connected to node i with a link. Constraints (2) in the optimization model compute, for a given assignment x_i, the discrepancy δ_i for each node, which is the difference between the value R_i associated with the node and the total number of neighbors of node i whose associated value $x_j = 1$. The objective function (1) requires minimizing the total discrepancy; hence the resulting optimal values x_i, will be such that the total discrepancy is a low as possible. If R_i's is not noisy, then such total discrepancy should be 0. Note that, even with a total zero discrepancy, the resulting assignment could return false positives (nodes identified as belonging to the population when they are not) and false negatives (nodes NOT identified as part of the hidden population even if they are). The degree of false identification depends on the amount and the accuracy of information as well as the assumptions underlying the model. The model relies on the knowledge of three pieces of information. First, we assume to know the size N of the target population. Second, we assume that the knowledge of the network structure of the target population is known from the sampled network. Third, we assume to know the responses R_i for each member of the population. The first assumption is a common assumption of many models that estimate the size of the hidden populations (Crawford et al., 2018; Maltiel et al., 2015). They usually assume the existence of a network and that the

"target population social network is a finite network G = (V, E) with no parallel edges or self-loops" (Crawford et al., 2018). In the current study, we define *N* as the total population in a community of interest.

The second assumption is perhaps the most difficult assumption to meet. The most common approach to construct the complete network from a sampled network is to employ the Exponential Random Graph Models (ERGMs), which are well-established models to statistically analyze social and other network data. In essence, ERGMs allow us to predict the existence of ties between each pair of the nodes in the network by estimating the degree of homophily, i.e., the tendency that two nodes with specific characteristics, e.g., gender, age, are more likely to be connected. This prediction process is a model-based simulation that we calculated a set of network configurations of the ego-centric network, applied the ERGMs with the alignment of sample size and network configurations to obtain the scaled-up model coefficients (Kolaczyk & Krivitsky, 2015; Krivitsky et al., 2019; Krivitsky & Morris, 2017), and then used the model coefficients to simulate the population network. The general form of the ERGMs can be expressed as:

$$P(Y = y) = \frac{\exp(\theta' g(y))}{k(\theta)}$$

where Y is the random variable for the state of the network with an observed network y, g(y) is a vector of sufficient statistics for the observed network y, θ is the vector of model coefficients, and k(θ) is a normalized constant representing the summation of all possible networks with the same node set of the observed network y. The ERGMs applied to estimate the network structure for our sample is detailed in Section 5. Lastly, for the third assumption, we took the survey responses *Ri* from each member of the sampled population as the face values. Next chapter elaborates our survey design and questionnaire which focus on obtaining the three pieces of information required by the methodology.

4. Creating a Testing Bed: The Field Survey of Bangladesh Northwestern Region

As stated in chapter 2, the kidney sales problem has several unique features. A naïve network simulation might reveal the properties of general public, but not the properties of the infiltrated kidney sellers and brokers. The team performed a field survey and created a simulation network based on the survey outcomes. By doing so, we tested the model validity in a real context.

1) Survey Design: Sampling and Implementation

A pilot survey of 30 respondents in March 2020 for a week. Although the original schedule for the complete survey was May-June 2020 for a one-month duration, the schedule was delayed due to the COVID-19 pandemic. The final survey was administered in October 2021. More than 400 responses were collected by

seven interviewees within the short time span under the leadership of the field coordinator. Our target sample size was 300 over a total adult population of 2,520 in Ward-2 in Matrai Union of Kalai Upazila, a known region for the kidney sales problem in Bangladesh.

The field survey followed a convenient sampling approach in which 30 interviewees walked around the Ward-2 and requested the responses to local encounters who provided us full consent. The target population was the adults of age 18 and older in Ward-2. For the ease of data input and processing, the interviewees used a survey software 'Qualtrics' installed on an electronic tablet to manually input the respondents' answers. The survey link was shared with interviewees via the local coordinator. The software recorded each tablet device's Internet Protocol address, input location, and input time, which allowed to evaluate the credibility of the inputs.

An interview guide was verbally communicated by the local administrator before the survey began, and interviewers' feedbacks and additional questions were communicated via the local coordinator in the every-other-day meetings and the 24 hours active group chat with researchers. At the end of each survey day, the local coordinator and researchers checked the survey responses. After completing the field visit, the data processing was done by the local coordinator and the researchers in collaboration. Additional notes on the unique contexts in the field were drafted by the local coordinator.

2) Questionnaires

The survey questionnaire is composed of six sections: 1) demographic information (age, gender, educational attainment, marital status, number of household member and child), 2) financial and professional life (work status, employment status, wage, land and livestock owned), 3) general network information (people chat last week, people sharing secrets, list of three light and three best friends totaling six), 4) kidney seller information (whether heard of/know kidney sellers and how many, the demographic of sellers/the reasons of selling the respondents think/broker existence if heard of/know sellers, the specific network information such as age/sex/profession/reason for selling/broker existence/how close with the respondents for up to five kidney sellers if know sellers), 5) broker information (whether know brokers, the broker's specific demographic and whether the broker sold own kidney for up to five brokers if know broker), and 6) various socioeconomic perceptions (perception on kidney sales, social trust, happiness, and health).

Sections 3) - 5) ask the key network information such as the network with friends, network with sellers, and the network with brokers. Because the number of the friends the respondents report can heavily depend

on their personalities and the style of interview, we distinguished the network questions into how many they know and what are the demographics of each friend for up to three (light friends, best friends) and five (sellers, broker). By limiting the number of neighbors for each respondent, we intended to avoid too weak network ties. Indeed, some people reported that they chatted with 1,000 people in the work in the past week (professional shopper, etc.). The full survey questionnaires in both English and Bangladesh Languages are attached in Appendix I.

3) Final Data and Descriptive Statistics

The data, after cleaning, contained 392 observations, which represents 15.5% of the total adult population in Ward-2. With additional removals of the observations with missing data, the final data included 320 egos and subsequent alters which account for 13% of the total population.

Table 1 shows the descriptive statistics of the final data. The mean age was 38 years old with the minimum age 18 and the maximum age 85. Our sample contained slightly more female than male, and the average years of education they finished is 7 years. More than 90% of the sample were married and had on average 2 children in a 4 people household. Only half of the respondents were currently working (many females are housewives), but the work income was lower than the average wage in Bangladesh. Many people were farmers owning livestock and land in their households.

Table 1. Descriptive Statistics for the Adult Sample Age 18+ in the Matrai Union Ward-2

Variable	Mean (S.E.)	Min	Max
Demographics			
Age	38.02 (0.68)	18	85
Gender (Female)	0.428 (0.03)	0	1
Educ (Years)	6.93 (0.24)	0	18
Married	0.9 (0.02)	0	1
Number of Child	1.99 (0.68)	0	8
Number of Household	4.23 (0.08)	1	14
Currently Working	0.48 (0.03)	0	1
Monthly Work Income (Country Avg) *(Category 1;less to 5;more)*	1.61 (0.06)	1	5
Household Owns Livestock	0.79 (0.02)	0	1
Household Owns Land (Bighas)	74.35 (6.68)	0	1320
Network Information			
People Chatted in the Work (Past Week)	29.18	0	1000

	(4.66)		
People Chatted in Leisure Time (Past Week)	31.54	0	600
	(2.90)		
Number of Family Sharing Secrets	2.40	0	25
	(0.17)		
Number of Friends Sharing Secrets	2.07	0	10
	(0.08)		
Kidney sales			
Heard of Kidney Sellers	0.51	0	1
	(0.03)		
Know Some of the Kidney Sellers	0.55	0	1
	(0.04)		
The Number of Sellers Know	3.09	1	30
	(0.36)		
Know Brokers	0.157	0	1
	(0.03)		

The survey found that, in one week, the respondents on average chat with 30 people in the workplace and outside the workplace. However, they share secrets only with 2 family members and 2 friends. Half of the respondents have heard about kidney sellers, while half of the heard people actually knew the sellers, which is about 25% of the total sample. These people know on average 3 sellers, mostly ranging between 1 to 3, except for some outlier cases. Only 15% of the respondents who heard about sellers knew brokers, which is 31 (or 8%) of the total 392 observations.

Table 2. Respondent Characteristics and the Awareness of Kidney Sales, T-test Estimates

- Male (Not Female) is more likely to have heard (-.129***) /know (-.160**) sellers
- Working people are more likely to have heard (.148***) sellers, but not knowing (.106)
- A person owns livestock is more likely to have heard (.146***) sellers but not knowing (.042)
- A person shares secret with one more family member has 1.7% more chance to have heard (.017**) sellers but not knowing (.014). Friends secret sharing has no significant correlation.
- Among the seller heard people, those share secret with one more family member have 2.5% more chance to know brokers (.025**).
- Among the seller heard people, those who actually know the sellers have 25.9% more chance to know brokers (.259***).

Table 3. The Awareness of Kidney Sales and Socioeconomic Perception, T-test Estimates

- Those heard of kidney sellers are likely to agree that kidney selling should be illegal (-.264***), but not among the sellers knowing people (-.001).
- Similarly, those knowing brokers are likely to agree that kidney selling should be illegal (-.140*).
- Those knowing the kidney sellers do not think financial reasons as key drivers (.431***), but not among the broader group of sellers heard people (.093).
- Similarly, those knowing brokers do not think financial reasons as key drivers (-.427***)
- If someone knows sellers, they are likely to feel their life happy (-.252**). But heard of sellers do not have significant correlations (-.041).

Table 2 presents the summary of bivariate correlational statistics investigating the associations between respondent characteristics and their awareness of kidney sales. Table 3 shows the analogous summary statistics regarding the perception of kidney sellers among those who are aware of kidney sales.

5. Global Network Simulation Reflecting the Survey Data Properties

1) The Network Simulation Using the Survey Data

To estimate the network among the sampled population, we utilized the ego-centric network information from the survey questions 2, 3, 15, and 17.3 (Appendix I). The questions reflect the sex and age of egos (respondents) and alters (the respondent's friends and kidney sellers). Age was classified as four groups, age_group.2 (18-29), age_group.3 (30-39), age_group.4 (40-49), and age_group.5 (over 50). We removed isolated nodes from the analysis because those nodes contribute little to the ERGMs and our proposed algorithm. The final network data contained 320 egos and 1,689 alters. The egos and alters' distributions of sex and age groups are presented in Table 4.

Table 4. The Sex and Age by Ego and Alter

		Ego		Alter	
		N	%	N	%
Sex	Male	192	60	1,137	67
	Female	128	40	552	33
Age	18-29	97	30	588	35
	30-39	95	30	579	34
	40-49	71	22	332	20
	Over 50	57	18	190	11
Total		320	100	1,689	100

In terms of degree distribution, each ego nominated, on average, 5 alters with minimum 1 and maximum 11. The degree distribution is displayed in Figure 4.

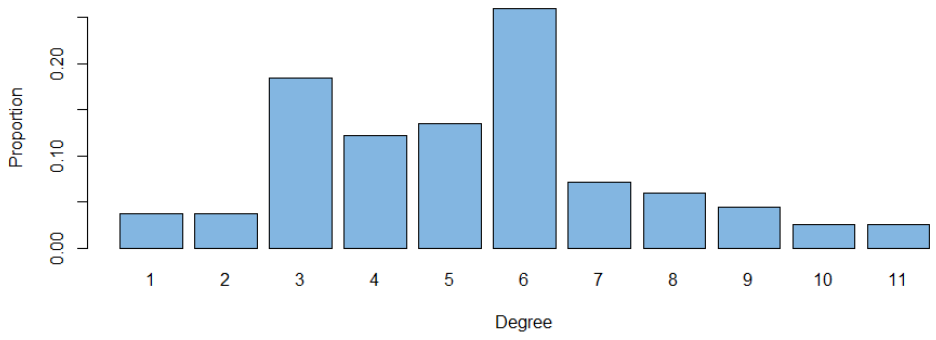


Figure 4. Degree Distribution

We used the '*ergm.ego v1.0.0 package*' in R to estimate an ERGM and simulate a global network. Table 5 shows the ERGM estimates. The results indicate that the observed network contains more nodes with degree 3 ($P < 0.001$) or degree 6 ($P < 0.001$) than expected by chance. Sex homophily shows that male is more likely to connect with male, and female is more likely to initiate a relationship with female ($P < 0.001$). Age group homophily illustrates that people prefer to connect with the nominees in the same age group ($P < 0.001$).

Table 5. The ERGM Estimation

	Estimate	Std. Error	z value	Pr(>|z|)	
Netsize.adj	-5.768	0.000	-Inf	< 0.001	***
Edges	0.189	0.200	0.945	0.344	
Degree3	0.708	0.188	3.768	< 0.001	***
Degree6	0.853	0.155	5.496	< 0.001	***
Male	0.045	0.091	0.500	0.617	
Sex homophily	1.728	0.088	19.723	< 0.001	***
Age_group.3	0.020	0.098	0.206	0.837	
Age_group.4	-0.034	0.138	-0.247	0.805	
Age_group.5	-0.191	0.160	-1.191	0.234	
Age_group homophily	0.927	0.070	13.182	< 0.001	***

* Netsize.adj is fixed by offset and is not estimated.; *** P < 0.001

The goodness-of-fit diagnostics of the ERGM estimate shown in Figure 5 indicate that the estimations from the model center around the values of the observed model terms, indicating that the estimates match the target statistics. Likewise, we observed that the estimated degree distribution fits to the observed degree distribution reasonably well in Figure 6. Both results together suggest that the network properties in the estimated global network do not differ significantly from the observed ego-centric network.

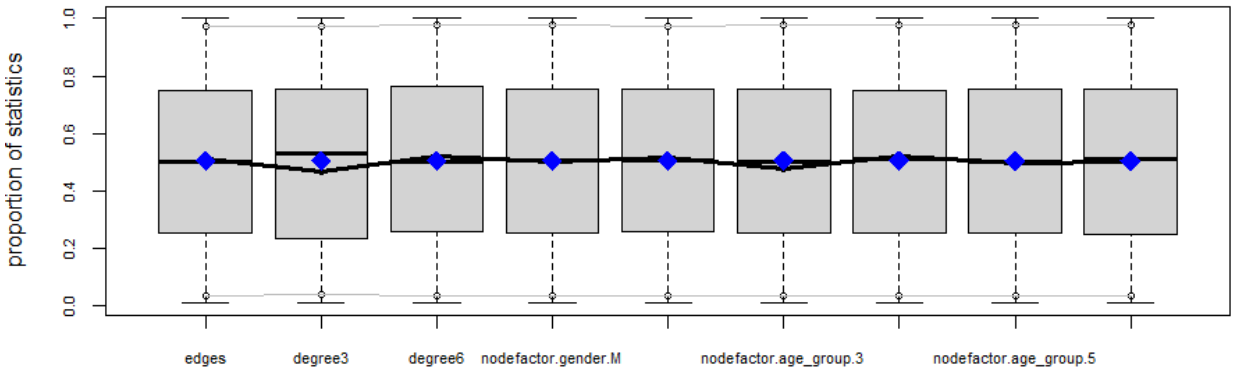


Figure 5. Goodness of-Fit Diagnostics Based on Model Statistics

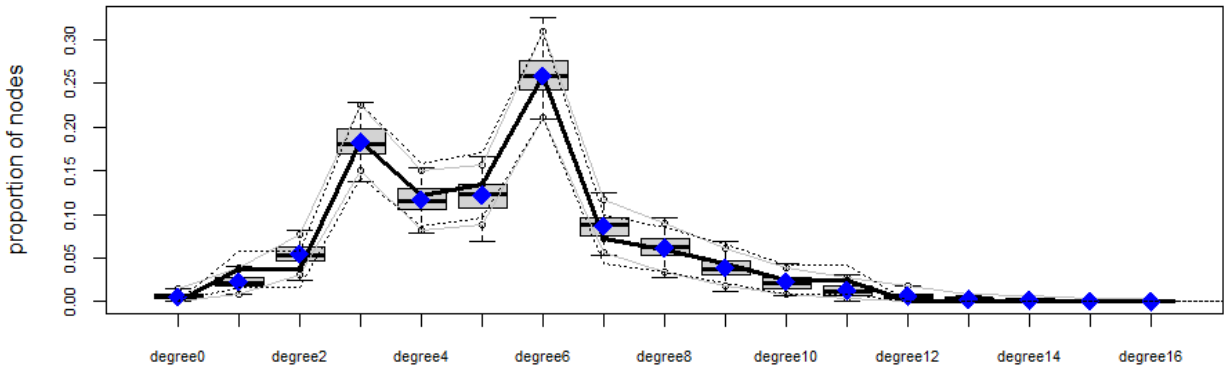


Figure 6. Goodness-of-Fit Diagnostics Based on Degree Distribution

2) **The Homophily Estimation for the Respondents and the Kidney Sellers**

Table 6 presents the distribution of sex and age for respondents who nominated at least one kidney seller in question 17.3. The total number of respondents who nominated at least one kidney seller is 103. According to the respondents, most kidney sellers are male (77%) aged between 18 and 39 (79%).

Table 6. The Sex and Age by Respondent and Kidney Seller

		Respondent		Kidney Seller	
		N	%	N	%
Sex	Male	73	71	198	77
	Female	30	29	60	23
Age	18-29	27	26	79	31
	30-39	32	31	125	48
	40-49	21	21	44	17
	Over 50	23	22	10	4

We further applied ERGMs to explore the relationship between respondents and kidney sellers. Table 7 reveals that the female-female relationship is more likely to be formed as compared to the female-male relationship ($P < 0.05$). Age_group.2 homophily effect is more likely to be observed than the relationship between Age_group.2 and Age_group.5 ($P < 0.05$) or between Age_group.5 and Age_group.5 ($P < 0.05$).

Table 7. The Homophily Estimation by ERGM

	Estimate	S.E.	Z	P-value	
Netsize.adj	-4.635	0.000	-Inf	< 0.001	***
Edges	1.309	0.366	3.572	0.000	***
Female - Female		(reference)			
Female - Male	-0.603	0.283	-2.127	0.033	*
Male - Male	0.066	0.309	0.213	0.831	
Age_group.2 - Age_group.2		(reference)			
Age_group.2 - Age_group.3	0.340	0.303	1.121	0.262	

Age_group.3 - Age_group.3	0.119	0.409	0.291	0.771	
Age_group.2 - Age_group.4	-0.535	0.389	-1.375	0.169	
Age_group.3 - Age_group.4	-0.144	0.386	-0.374	0.709	
Age_group.4 - Age_group.4	0.043	0.464	0.093	0.926	
Age_group.2 - Age_group.5	-0.919	0.403	-2.278	0.023	*
Age_group.3 - Age_group.5	-0.142	0.377	-0.376	0.707	
Age_group.4 - Age_group.5	-0.810	0.433	-1.869	0.062	
Age_group.5 - Age_group.5	-1.975	0.835	-2.364	0.018	*

* Netsize.adj is fixed by offset and is not estimated.; *P < 0.05; *** P < 0.001

6. The Results: The Model Estimates of the Hidden Population and Implications

1) The Model Estimates of the Number of Kidney Sellers

Solving our proposed IP model on the simulated network of 320 respondents yielded a solution with 40 sellers, which accounts for 13% of respondents. The total number of connections R_i in the network is 258, and the discrepancy δ_i is 164.

$$N_H = \sum_{i=1}^{320} x_i = 40,$$

$$\sum_{i}^{320} R_i = 258,$$

$$\sum_{i=1}^{320} |\delta_i| = 164$$

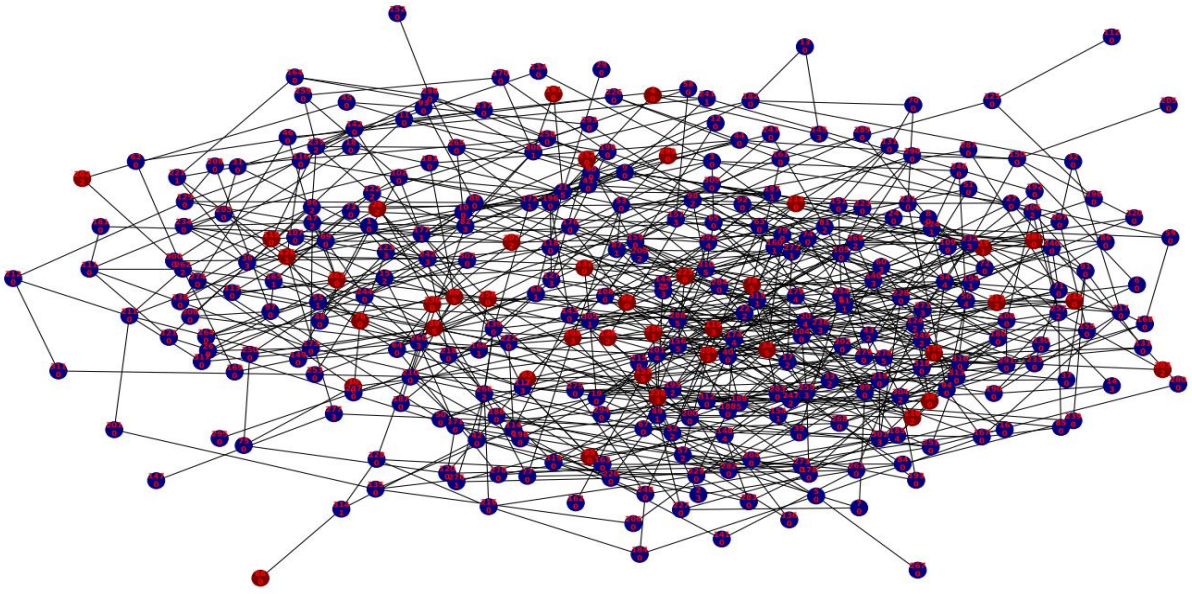


Figure 7. The Hidden Population in the Simulated Network

Figure 7 presents the structure of the network and marks the node of the hidden populations as red. Each node has two numbers of which the upper one is the node number and below one is the assigned response R_i in the simulation. Scaling up the network to the entire adult population of 2,520 results in an estimate of a total hidden population of 315.

7. Discussion and Conclusion

We performed a post-hoc literature search to validate our estimate, i.e., kidney sellers represent about 13% of the total population. The statistics on the topic is, however, very scarce, and according to a news article published in 2021, the Kalai Upazila's assembly (Parishad) chairman reported that approximately 500 people sold their kidneys in Kalai in the past 10 years (Daily-Sun, 2021). This figure is roughly 1.6% of the total population, which indicates that either our model overestimates the prevalence significantly or the local knowledge regarding the extent of the problem is incomplete. Several causes for our overestimation are possible. First and foremost, our model could overestimate as the survey questions concerning the network structure failed to capture an accurate picture of their social networks. Such errors are not necessarily due to respondents' misinformation, false responses, and the sample misrepresentation found in other methodologies, but possibly due to the privacy and confidentiality issues for the respondents and their friends and the potential reluctances of giving out information that will lead to the identification of close persons who are sellers suggested by the field survey coordinator. Methodologically, it is possible that some of the assumptions, with regard to the network structure of the population, are violated. Specifically, as we were unable to collect unique identifiers of friends, we could not detect common and in-common neighbors, which could have significantly affected the ability to estimate the network structure accurately. However, obtaining a unique identifier could cause various privacy and confidentiality issues to the residents due to the sensitivity of the topic. We believe that the two other assumptions are more likely to hold. The first assumption on the size N of the target population is known from the Bangladesh census data, while the third assumption on R_i is likely to hold given that the sampled data appear to represent the general population of the village well.

We should also note that validating the model using the Bangladesh data is an unrealistic endeavor. Statistics on highly hidden population are known to be unreliable even if they are available. For instance, the earlier estimate of kidney sellers in Nepal was between 500 to 1,000, while a survey-based number that became available at a later stage put this number somewhere between 120 and 160 (The Asia Foundation, 2015). In light of this, we validated our model using different network data before applying the model to the Bangladesh context. The description of the validation method and the result are available in Appendix II. The validation result indicated that the model estimates the prevalence of the selected population fairly

accurately as long as the three assumptions are met. Further research is warranted to examine the sensitivity of these results to the violation of each assumption as well as the survey instruments that can accurately capture the parameters used in the model.

Appendix I. Survey Questionnaires

A. English Version (Before Qualtrics Input)

Hidden Population Detection Survey

Demographic Characteristics

1. Code:
2. How old are you? …….. (answer in years)
3. Sex (a) male (b) female (c) others
4. Education level …….. (answer in class or degree)
5. Marital status (a) married (b) single (c) others:……..
6. How many children do you have? ……..
7. How many people do you have in your household? ……..

Financial/Professional Life

8. Which sentence describes your current situation the best?
    a) I am employed (1)
    b) I am unemployed but looking for a job (2) **---- skip to question 12**
    c) I am not working, and I am not looking for a job (3) **---- skip to question 12**
    d) If someone gives me an opportunity, I will work (4) **---- skip to question 12**
    e) Other (5) (Please specify:) **---- skip to question 12**

    8.1. Please choose the sentence describing your current situation the best. I am:
        a) self-employed (1)
        b) salary worker (2)
        c) unpaid worker (3)
        d) other (4)
9. My main profession is a:
    a) farmer (1)
    b) day laborer (2)
    c) shop keeper (3)
    d) construction worker (4)
    e) hotel or restaurant worker (5)
    f) driver (truck driver, rickshaw driver, taxi driver, van driver, etc.) (6)
    g) fisher man (work in the river or sea) (7)
    h) factory worker (garment factory, machine parts factory, etc.) (8)
    i) teacher (9)
    j) government official (10)
    k) housewife (11)
    l) other (12) (Please specify:)

    9.1. The average salary is about 26,000 BDT per month in Bangladesh. Is your job pay higher or lower than the average salary?
        a) much lower (1)
        b) lower (2)
        c) similar (3)
        d) higher (4)
        e) much higher (5)

10. How many bighas/acres of land do your family own? …….....

11. Do your family raise any domestic animal? (a) Yes (b) No

Network Detection

12. How many people did you chat at work last week? ……. number

13. How many people did you chat for leisure last week? ……. number

14. How many people do you normally share your personal worries?
    a) friends……. number
    b) family .....number

15. Please list 3 friends with whom you share your personal worries and additional 3 friends with who you do NOT share personal worries. Please then identify if those friends know each other. Please just write their pseudonym instead of their real names.

				Do your friends know each other?					
	Pseudonym	Age	Sex	Friend (A)	Friend (B)	Friend (C)	Friend (D)	Friend (E)	Friend (F)
Please list friends with whom you share your personal worries and identify if those friends know each other. Please just write their pseudonym and do not write their real name.	Friend (A)……								
	Friend (B)……								
	Friend (C)……								
Please list friends with whom you do **NOT** share your personal worries **but had chat last week**. Please just write their pseudonym and do not write their real name.	Friend (D)…….								
	Friend (E)…….								
	Friend (F)…….								
				Friend (A)	Friend (B)	Friend (C)	Friend (D)	Friend (E)	Friend (F)

Kidney Seller Information

16. Have you heard of anyone who sold kidney in your Union?
    a) No --- **skip to question 23**
    b) Yes

17. Do you know any of these kidney sellers in person? (Note: We are not interested in knowing who they are. Rather we would like to know about their profile and the reasons behind.)
    a) No --- **skip to question 18**

b) Yes

17.1. How many kidney sellers do you know in person? ……. number

17.2. Can you write the age, gender, profession, and reasons for selling kidneys (if known to you) for these people? Just write as much you feel comfortable.

17.3. Also, please select one from the list below that best describe your relationship with each of the kidney seller.

a) A close friend (with whom you share worries)
b) Not so close friend
c) Acquaintance

(Please do NOT write their names. The purpose of this survey is not to identify the sellers)

	Sex (M/F) (1)	Age (2)	Profession (3)	Reason for selling kidney (4)	Had a broker/s? (5)	Your relationship with the seller (6)
Person 1						
Person 2						
Person 3						
Person 4						
Person 5						

Questions 18-20 ask about the kidney sellers you have heard about, rather than those sellers you know in person.

18. Do you know why these people sold their kidneys?

18.1. No --- **skip to question 19**
18.2. Yes (select all it applies)

a) Poverty
b) Dowry
c) Micro-credit
d) To repay other loans
e) Drug addiction
f) Other reasons (Please specify:)

19. You think most of these sellers are:

19.1. Sex
a) Male
b) Female
c) Don't know

19.2. Age
a) Young
b) Middle age
c) Old

d) Don't know

19.3. Income level
a) Low
b) Medium
c) High
d) Don't know

20. Do you know if any of those sellers sold their kidney through brokers?
a) Yes
b) No
c) Don't know

Broker Information

21. Do you know any brokers of kidney sales?
a) No --- **skip to question 23**
b) Yes

22. Please write the age, gender, and profession of these brokers whom you know. Just write as much you feel comfortable. (Please do NOT write their name).

	Sex (M/F) (1)	Age (2)	Profession (3)	Key features (respected?, know many people?, powerful?, rich?, etc.) (4)	Did the broker sell his/her own kidney too? (5)
Person 1					
Person 2					
Person 3					
Person 4					
Person 5					

Opinion / Religion

23. Please rate your level of agreement with the following statements.

Statement	Strongly Agree (1)	Agree (2)	Neutral (3)	Disagree (4)	Strongly Disagree (5)
(a) Kidney sales should be illegal.	○	○	○	○	○
(b) If someone has severe financial need, kidney sales is an acceptable option.	○	○	○	○	○

(c) If I sold kidney, I would keep it as a secret.	○	○	○	○	○
(d) I can trust anyone in this Union.	○	○	○	○	○
(e) Practicing religion is important.	○	○	○	○	○
(f) Police generally do right things.	○	○	○	○	○
(g) Government generally does right things.	○	○	○	○	○
(h) I am happy with my life.	○	○	○	○	○
(i) I am healthy.	○	○	○	○	○

Thank you for your participation.

B. Bangladesh Language Version (Qualtrics Input Final)

লুকানো জনসংখ্যা সনাক্তকরণ জরিপ

Start of Block: ভূমিকা

- আর্থ -সামাজিক জরিপ: অংশগ্রহণের জন্য আপনাকে ধন্যবাদ

End of Block: ভূমিকা

Start of Block: ক. জনসংখ্যাগত বৈশিষ্ট্য

1. কোড: _____

4. শিক্ষাগত যোগ্যতা: ..... ক্লাস বা ডিগ্রী। *(Some question numbers and orders were changed due to survey flow and cultural context.)*

2. আপনার বয়স কত? ..... বছর।

3. লিঙ্গ:
পুরুষ (1)
মহিলা (2)
অন্যান্য (3) _____

5. বৈবাহিক অবস্থা:
বিবাহিত (1)
অবিবাহিত (2)
অন্যান্য (3) _____

6. আপনার সন্তান সংখ্যা কয়জন ? ..... জন।

7. আপনার পরিবারে (খানা) সদস্য সংখ্যা কয়জন? ..... জন।

End of Block: ক. জনসংখ্যাগত বৈশিষ্ট্য

Start of Block: খ. আর্থিক/ পেশাগত জীবন

8. কোন বাক্যটি আপনার বর্তমান পরিস্থিতিকে সর্বোত্তমভাবে বর্ণনা করে?
(We added 'housewife' option in the first question of the block to avoid confusions.)
আমি চাকরি করি (1)
আমি বেকার কিন্তু একটি চাকরি খুঁজছি (প্রশ্ন 10 সরাসরি চলে যান) (2)
গৃহিণী (প্রশ্ন 10 সরাসরি চলে যান) (6)
আমি কাজ করছি না এবং চাকরিও খুঁজছি না (প্রশ্ন 10 সরাসরি চলে যান) (3)
যদি কেউ আমাকে সুযোগ দেয়, আমি কাজ করব (প্রশ্ন 10 সরাসরি চলে যান) (4)
অন্যান্য: অনুগ্রহ করে নির্দিষ্ট করুন ..... (প্রশ্ন 10 সরাসরি চলে যান) (5)

End of Block: খ. আর্থিক/ পেশাগত জীবন

Start of Block: খ. আর্থিক/ পেশাগত জীবন - 1

8-1. আপনার বর্তমান অবস্থাকে সর্বোত্তমভাবে বর্ণনা করে নিচের কোনটি? আমি:
নিজেই মালিক (1)
বেতনভুক্ত কর্মী (2)
বেতনছাড়া কর্মী (3)
অন্যান্য (4) _____

9. আপনার প্রধান পেশা নিচের কোনটি:

কৃষক (1)
দিনমজুর (2)
দোকানদারী (3)
নির্মাণ শ্রমিক (4)
হোটেল বা রেস্টুরেন্ট কর্মী (5)
ড্রাইভার (ট্রাক ড্রাইভার, রিকশা চালক, ট্যাক্সি ড্রাইভার, ভ্যান ড্রাইভার ইত্যাদি) (6)
জেলে (7)
কারখানার শ্রমিক (পোশাক কারখানা, মেশিন পার্টস কারখানা ইত্যাদ) (8)
শিক্ষক (9)
সরকারী চাকরী (10)
গৃহিণী (11)
অন্যান্য (12) _____

9-1. বাংলাদেশে প্রতি মাসে গড় বেতন প্রায় ২৬,০০০ টাকা। আপনার চাকরির বেতন কি গড় বেতনের চেয়ে বেশি বা কম?

অনেক কম (1)
কম (2)
একই (3)
বেশি (4)
অনেক বেশি (5)

End of Block: খ. আর্থিক/ পেশাগত জীবন - 1

Start of Block: খ. আর্থিক/ পেশাগত জীবন - 2

10. আপনার পরিবারের কত শতাংশ জমি আছে? ..... শতাংশ।

11. আপনার পরিবার কি কোনও গৃহপালিত পশু লালন পালন করে?

হ্যা (1)
না। (2)

End of Block: খ. আর্থিক/ পেশাগত জীবন - 2

Start of Block: গ. নেটওয়ার্ক সনাক্তকরণ

12. গত সপ্তাহে আপনি কতজনের সাথে কর্মস্থলে কথা বলেছিলেন?.....জন।

13. গত সপ্তাহে আপনি কত জনের সাথে অবসর সময়ে কথা বলেছিলেন? .....জন।

14. আপনি সাধারণত কতজন সাথে আপনার সুখ-দু:খের কথা শেয়ার করেন?

পরিবারের সদস্য ..... জন। (2)
বন্ধু-বান্ধব ..... জন। (1)

End of Block: গ. নেটওয়ার্ক সনাক্তকরণ

Start of Block: গ. নেটওয়ার্ক সনাক্তকরণ – 1

15. অনুগ্রহ করে 3 জন বন্ধুকে তালিকাভুক্ত করুন যাদের সাথে আপনি আপনার ব্যক্তিগত সুখ-দু:খগুলো শেয়ার করেছেন এবং অতিরিক্ত 3 জন বন্ধু যাদের সাথে আপনি ব্যক্তিগত সুখ-দু:খগুলো শেয়ার করেন নি। তারপরে সনাক্ত করুন যে সেই বন্ধুরা একে অপরকে চেনে কিনা। দয়াকরে তাদের আসল নামের পরিবর্তে কেবল তাদের ছদ্মনাম লিখুন।

1) অনুগ্রহ করে এমন বন্ধুদের তালিকা করুন যাদের সাথে আপনি আপনার ব্যক্তিগত সুখ-দু:খগুলো শেয়ার করেছেন এবং সেই বন্ধুরা একে অপরকে চেনে কিনা তা সনাক্ত করুন। দয়া করে কেবল তাদের ছদ্মনাম লিখুন এবং তাদের আসল নাম লিখবেন না।

	বয়স (1)	লিঙ্গ (2)	বন্ধু (এ) (3)	বন্ধু (বি) (4)	বন্ধু (সি) (5)	বন্ধু (ডি) (6)	বন্ধু (ই) (7)	বন্ধু (এফ) (8)

বন্ধু (এ) (ছদ্মনাম) (1)								
বন্ধু (বি) (4)								
বন্ধু (সি) (5)								

2) অনুগ্রহ করে এমন বন্ধুদের তালিকা করুন যাদের সাথে আপনি আপনার ব্যক্তিগত সুখ-দু:খগুলো শেয়ার করেন নি, তবে গত সপ্তাহে কথা বলেছিলেন। দয়া করে কেবল তাদের ছদ্মনাম লিখুন এবং তাদের আসল নাম লিখবেন না।

	বয়স (1)	লিঙ্গ (2)	বন্ধু (এ) (3)	বন্ধু (বি) (4)	বন্ধু (সি) (5)	বন্ধু (ডি) (6)	বন্ধু (ই) (7)	বন্ধু (এফ) (8)
বন্ধু (ডি) (ছদ্মনাম) (9)								
বন্ধু (ই) (10)								
বন্ধু (এফ) (11)								

End of Block: গ. নেটওয়ার্ক সনাক্তকরণ - 1

Start of Block: ঘ. কিডনি বিক্রেতার তথ্য

16. আপনি কি আপনার ওয়ার্ডে এমন কারও কথা শুনেছেন যিনি তার নিজ কিডনি বিক্রি করেছেন?

না (প্রশ্ন 23 সরাসরি চলে যান) (1)
হ্যাঁ (2)

End of Block: ঘ. কিডনি বিক্রেতার তথ্য

Start of Block: চ. মতামত/ ধর্ম

23. অনুগ্রহ করে নিম্নলিখিত বিবৃতিগুলির বিষয়ে আপনার মতামতটি প্রদান করুন।

	দৃঢ়ভাবে একমত (1)	একমত (2)	নিরপেক্ষ (3)	একমত না (4)	দৃঢ়ভাবে একমত নই (5)
ক) কিডনি বিক্রয় অবৈধ হওয়া উচিত। (1)					
খ) যদি কারও গুরুতর আর্থিক প্রয়োজন হয়,					

তাহলে নিজ কিডনি বিক্রয় করতে পারে। (2) গ) কেউ কিডনি বিক্রি করলে, এটি তার গোপন করা দরকার। (3) ঘ) এই ওয়ার্ডের প্রত্যেকের উপর প্রত্যেক ব্যক্তির বিশ্বাস রাখা উচিত। (4) ঙ) ব্যক্তিজীবনে ধর্ম পালন করা গুরুত্বপূর্ণ। (5) চ) পুলিশ সাধারণত সঠিক কাজ করে। (6) ছ) সরকার সাধারণত সঠিক কাজ করে। (7) জ) আমি আমার জীবন নিয়ে খুশি। (8) ঝ) আমি একজন সুস্থ ব্যক্তি। (9)	

End of Block: চ. মতামত/ ধর্ম

Start of Block: ঘ. কিডনি বিক্রেতার তথ্য - 1

17. আপনি কি এই কিডনি বিক্রেতাদের কাউকে ব্যক্তিগতভাবে চিনেন? (তারা কারা তা জানতে আমরা আগ্রহী নই। বরং আমরা তাদের প্রোফাইল এবং পিছনের কারণগুলি সম্পর্কে জানতে চাই।)

ক) না (প্রশ্ন 18 সরাসরি চলে যান) (1)

খ) হ্যাঁ (2)

End of Block: ঘ. কিডনি বিক্রেতার তথ্য - 1

Start of Block: ঘ. কিডনি বিক্রেতার তথ্য - 2

17-1. আপনি কতজন কিডনি বিক্রেতাকে ব্যক্তিগত ভাবে চেনে ? ..... জন।

17-2. আপনি কি এই লোকগুলো সর্ম্পকে নিচের তথ্যগুলো বলতে পারবেন? (যদি আপনার জানা থাকে এবং আপনি যতটা সম্ভব স্বাচ্ছন্দ্যবোধ করেন ঠিক ততটাই বলুন)

	লিঙ্গ (পু/ম) (1)	বয়স (7)	পেশা (3)	কিডনি বিক্রির কারণ (4)	বিক্রেতার সাথে আপনার সম্পর্ক (5)	তিনি কি নিজেই মধ্যস্থতাকারী? (6)
ব্যক্তি 1 (4)						
ব্যক্তি 2 (5)						
ব্যক্তি 3 (6)						

ব্যক্তি 4 (7)						
ব্যক্তি 5 (8)						

End of Block: ঘ. কিডনি বিক্রেতার তথ্য - 2

Start of Block: ঘ. কিডনি বিক্রেতার তথ্য - 3

18-20. প্রশ্নগুলি আপনার শোনা কিডনি বিক্রেতাদের সম্পর্কে.....

18. আপনি কি জানেন কেন এই লোকেরা তাদের কিডনি বিক্রি করেছে ?
- না (প্রশ্ন 21 সরাসরি চলে যান) (1)
- হ্যাঁ (কারনগুলো নির্বাচন করুন 18.1) (2)

18-1. কারনগুলো নির্বাচন করুন
- দারিদ্র (1)
- যৌতুক (2)
- ক্ষুদ্র ঋণ (3)
- অন্যান্য ঋণ পরিশোধ করতে (4)
- মাদকাসক্তি (5)
- অন্যান্য কারণ (6) _____

19. আপনি মনে করেন এই বিক্রেতাদের বেশিরভাগই:

19-1. লিঙ্গ
- পুরুষ (1)
- নারী (2)
- জানি না (3)

19-2. বয়স
- অল্পবয়স্ক (1)
- মধ্যবয়স (2)
- বৃদ্ধ (3)
- জানি না (4)

19-3. আয়ের স্তর
- নিম্নবিত্ত (1)
- মধ্যবিত্ত (2)
- উচ্চবিত্ত (3)
- জানি না (4)

20. আপনি কি জানেন, সেই বিক্রেতাদের মধ্যে কেউ মধ্যস্থতাকারীর মাধ্যমে তাদের কিডনি বিক্রি করেছে কিনা?
- হ্যাঁ (1)
- না (2)
- জানি না (3)

End of Block: ঘ. কিডনি বিক্রেতার তথ্য - 3

Start of Block: ঙ. মধ্যস্থতাকারীর তথ্য

21. আপনি কি কিডনি বিক্রির কোনও মধ্যস্থতাকারীকে চেনেন?
- না (প্রশ্ন 23 সরাসরি চলে যান) (1)
- হ্যাঁ (2)

End of Block: ঙ. মধ্যস্থতাকারীর তথ্য

Start of Block: ঙ. মধ্যস্থতাকারীর তথ্য - 1

22. দয়া করে এই মধ্যস্থতাকারীদের বয়স, লিঙ্গ এবং পেশা লিখুন যাদের আপনি জানেন। আপনি যতটা স্বাচ্ছন্দ্যবোধ করেন ঠিক ততটাই লিখুন। (দয়া করে তাদের নাম লিখবেন না)

	লিঙ্গ (পু/ম) (1)	বয়স (6)	পেশা (3)	সমাজে তাদের অবস্থান (সম্মানিত, অনেক লোক চেনেন, ক্ষমতাশালী, ধনী, ইত্যাদি) (4)	মধ্যস্থতাকারী কি তার নিজের কিডনিও বিক্রি করেছে? যদি তাই হয়, আপনি কি 17.3 এর 1-5 ব্যক্তি অন্তর্ভক্ত করেছেন। (অনুগ্রহ করে ব্যক্তিটি নির্দিষ্ট করুন) (8)
ব্যক্তি 1 (1)					
ব্যক্তি 2 (2)					
ব্যক্তি 3 (3)					
ব্যক্তি 4 (4)					
ব্যক্তি 5 (5)					

End of Block: ঙ. মধ্যস্থতাকারীর তথ্য - 1

Appendix II. Model Validation

Experimental data: We looked alternative data on which we can test and validate our method to estimate the size of the hidden population. While the data that record both the information on the hidden population as well as the social network of respondents were not available, we were able to find several data that provides the information about the structure of social networks. The closest data we were able to find was the *Faux.Magnolia.High* data constructed from the National Longitudinal Study of Adolescent Health (Add Health) Wave I dataset (http://www.cpc.unc.edu/addhealth/). The Add Health survey was completed by 90,118 of 119,233 students in grades 7 through 12. 3. The first wave was conducted in 1994-1995, and included a friendship nomination module based upon student rosters.

The raw data of friendship nomination were used to construct *Faux.Magnolia.High* networks, ego-centric networks that are built through a model-based simulation. The specific steps to develop the networks are listed below. These steps were implemented using the programming language R (Goodreau et al., 2008).

1. Extract schools 086 and 186 into an R network object.
2. Remove all vertices that did not take the survey or were not on the school roster.
3. Convert the data from directed to undirected, by defining an undirected friendship Xij as equaling 1 if i named j as a friend and j named i as a friend.
4. Define race based on the answers to two questions (one on Hispanic ethnicity and one on race), allowing Hispanic identity as primary (i.e. all Hispanic are one category, non-Hispanic whites are another, non-Hispanic blacks are another, etc.)
5. Impute any missing Race, Grade and Sex values by random draw with weights determined by the size of the attribute classes in the school.
6. Fit the following model to the resulting school:

R> magnolia.fit <− ergm(magnolia ∼edges + absdiff("Grade") ++ nodematch("Race", diff = TRUE) + nodematch("Grade", diff = TRUE) ++ nodematch("Sex", diff = FALSE) + gwesp(0.25, fixed = TRUE),+ burnin = 10000, interval = 1000, MCMCsamplesize = 2500, maxit = 25,+ control = control.ergm(steplength = .25))

1. Simulate a single network from this model fit with fixed density:

R> faux.magnolia.high <− simulate(magnolia.fit, nsim = 1, burnin = 1e+8,+ constraints = ∼edges)

Figure A presents the final social networks that we obtained as a result. The networks comprised of 1461 nodes and 974 undirected edges, where the undirected edges represent mutual friendships, implying both students nominate the other person as a friend.

Model Validation: We used the *Faux.Magnolia.High* networks to test and validate the method described in the model section. Given that we do not have the information on the hidden population (i.e., kidney sellers), we randomly assigned kidney seller nodes to this network. We took this network as the population network and implemented the following Monte Carlo procedure to test and validate the model.

1. We randomly sampled n nodes for 100 times.
2. For each of the 100 samples, we solved the optimization problem described in the model section and reported the predicted number of sellers scaled up to 1461 nodes.

This process was done for a sample size n=14 (1%), 84 (6%), 170 (12%), and 210 (15%). The following histograms (Figure B.) presents the results of the simulations.

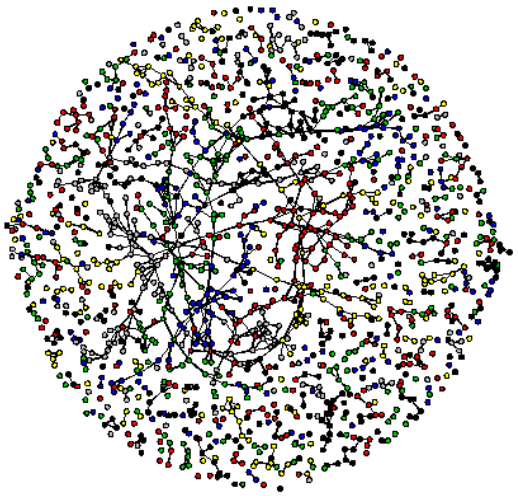


Figure A. Faux.magnolia.high networks

As seen in the figures, all histograms are clustered around 100, indicating that the model is, on average, successful in identifying the true size of kidney sellers (100). The histograms also indicate that the variability of the estimates reduces as the number of sample size increases, implying that the model can predict the number of kidney sellers with increasing accuracy as n increases. Our results indicated that, with the sample size of 84 (6%), the estimates are between 80 and 120 (±20%) 80%+ of the time. This result confirmed that our plan to sample approximately 300 out of 4,000 villagers (8%) for the survey in the field is likely to be sufficient for our model to estimate the size of the hidden population with a reasonable accuracy.

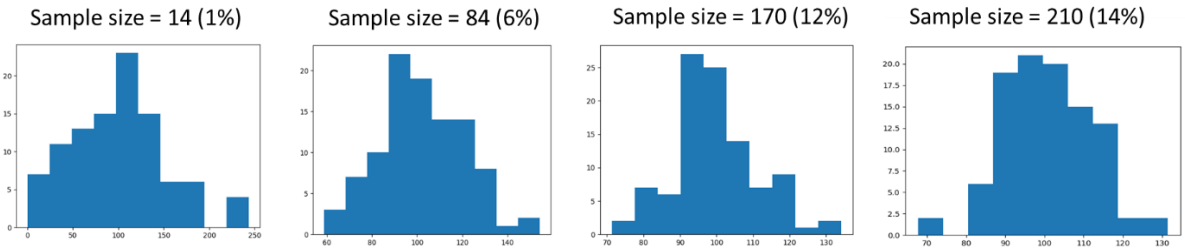


Figure B. Monte-Carlo simulation results to validate the estimation method.

REFERENCES

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

Crawford, F. W., Wu, J., & Heimer, R. (2018). Hidden population size estimation from respondent-driven sampling: A network approach. *Journal of the American Statistical Association*, *113*(522), 755–766.

Cruyff, M., Dijk, J. van, & Heijden, P. G. M. van der. (2017). The Challenge of Counting Victims of Human Trafficking: Not on the record: A multiple systems estimation of the numbers of human trafficking victims in the Netherlands in 2010–2015 by year, age, gender, and type of exploitation. *CHANCE*, *30*(3), 41–49. https://doi.org/10.1080/09332480.2017.1383113

Daily-Sun. (2021). Kidney trafficking goes unchecked | Daily Sun. *Daily-Sun*. https://www.daily-sun.com/post/589485/Kidney-trafficking-goes-unchecked

Dávid, B., & Snijders, T. A. (2002). Estimating the size of the homeless population in Budapest, Hungary. *Quality and Quantity*, *36*(3), 291–303.

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*(1), 153–186.

Frank, O., & Snijders, T. (1994). Estimating the size of hidden populations using snowball sampling. *JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM-*, *10*, 53–53.

Fraser, C., & Koizumi, N. (2017). Human Organ Trafficking: A study of change in the Egyptian kidney market 2008-2016. *Transplantation*, *101*, S13.

Goodman, L. A. (1961). Snowball sampling. *The Annals of Mathematical Statistics*, 148–170.

Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T., & Morris, M. (2008). A statnet Tutorial. *Journal of Statistical Software*, *24*(9), 1.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*, *44*(2), 174–199.

Heckathorn, D. D. (2002). Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems*, *49*(1), 11–34.

Hook, E. B., & Regal, R. R. (1995). Capture-recapture methods in epidemiology: Methods and limitations. *Epidemiologic Reviews*, *17*(2), 243–264.

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

Kolaczyk, E. D., & Krivitsky, P. N. (2015). On the question of effective sample size in network modeling: An asymptotic inquiry. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *30*(2), 184.

Krivitsky, P. N., & Morris, M. (2017). Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. *The Annals of Applied Statistics*, *11*(1), 427.

Krivitsky, P. N., Morris, M., & Bojanowski, M. (2019). *Inference for Exponential-Family Random Graph Models from Egocentrically-Sampled Data with Alter–Alter Relations*.

Magnani, R., Sabin, K., Saidel, T., & Heckathorn, D. (2005). Review of sampling hard-to-reach and hidden populations for HIV surveillance. *Aids*, *19*, S67–S72.

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*(3), 1247.

Moniruzzaman, M. (2012). "Living cadavers" in Bangladesh: Bioviolence in the human organ bazaar. *Medical Anthropology Quarterly*, *26*(1), 69–91.

Moniruzzaman, M. (2016). At the Organ Bazaar of Bangladesh: In Search of Kidney Sellers. In D. Siegel & R. de Wildt (Eds.), *Ethical Concerns in Research on Human Trafficking* (pp. 227–247). Springer International Publishing. https://doi.org/10.1007/978-3-319-21521-1_14

Moniruzzaman, M. (2019). The trade in human liver lobes: Bioviolence against organ sellers in Bangladesh. *Journal of the Royal Anthropological Institute*, *25*(3), 566–586.

Rubin, G., Umbach, D., Shyu, S.-F., & Castillo-Chavez, C. (1992). Using mark-recapture methodology to estimate the size of a population at risk for sexually transmitted diseases. *Statistics in Medicine*, *11*(12), 1533–1549.

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

Shelley, G. A., Killworth, P. D., Bernard, H. R., McCarty, C., Johnsen, E. C., & Rice, R. E. (2006). Who knows your HIV status II?: Information propagation within social networks of seropositive people. *Human Organization*, 430–444.

Shelton, J. F. (2015). Proposed utilization of the network scale-up method to estimate the prevalence of trafficked persons. *Special Issue Researching Hidden Populations: Approaches to and Methodologies for Generating Data on Trafficking in Persons*, *8*, 85.

The Asia Foundation. (2015). *Kidney Trafficking in Nepal: A Study of Selected VDCs in Kavrepalanchowk District*. Forum for Protection of People's Rights Nepal. https://www.scribd.com/document/432599584/Kidney-Trafficking-in-Nepal

Wang, J., Yang, Y., Zhao, W., Su, H., Zhao, Y., Chen, Y., Zhang, T., & Zhang, T. (2015). Application of network scale up method in the estimation of population size for men who have sex with men in Shanghai, China. *PloS One*, *10*(11), e0143118.

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*(474), 409–423.