Modeling, Estimation and Interpretation of Peer Effects in Social Networks

by

Jaekyeong Shin

B.A., Sungkyunkwan University, 2015

M.A., Sungkyunkwan University, 2018

M.A., University of Colorado Boulder, 2020

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Economics

2024

Committee Members:

Xiaodong Liu, Chair

Adam McCloskey

Carlos Martins-Filho

Scott Savage

Stephen Billings

Shin, Jaekyeong (Ph.D., Economics)

Modeling, Estimation and Interpretation of Peer Effects in Social Networks

Thesis directed by Prof. Xiaodong Liu

In this dissertation, I propose a model that estimates the underlying motives of peer effect. With Monte Carlo simulations and actual data, I apply the model and interpret the results by relating to the motives. Additionally, I suggest an algorithm to capture network links with panel data and binary outcome variables for the cases where network data are unavailable.

In the first chapter, a generalized spatial autoregressive model (GSAR) estimation model is presented. In addition to the intensity of peer effect, which is the sole focus of the standard SAR model, I introduce a "conformity parameter" to capture the underlying motives of peer effect: complementarity and conformity. These motives lead to different policy implications as only the former generates the social multiplier effect. I perform Monte Carlo simulations and demonstrate that the GSAR model is robust to model selection among different motives of peer effect. Also, I characterize the threshold for a positive social multiplier effect.

In the second chapter, I examine the microfinance data collected by Bharatha Swamukti Samsthe (BSS) in Karnataka, India. I extend the GSAR model for binary outcome variables and higher-order networks and apply it to the data. As a result, I find strong evidence of the villagers' peer effect with a significant conformity motive. Using the estimated parameters, I conduct a counterfactual analysis with alternative target groups. It is shown that the target group chosen based on the GSAR model is more effective than the original group designated by BSS.

The final chapter studies a case where network data is unavailable. I estimate a network structure using panel data of repeatedly observed samples with binary outcome variables. I present an algorithm that combines the binary particle swarm optimization process with the shrinkage estimator. Simulation results for different types of networks are presented, along with a comparison of the estimation performance with varying numbers of particles.

Dedication

To Seungyoung, who has so much belief in me even when I do not.

To my family, for all the support from a country on the other side of the planet.

Acknowledgements

This dissertation could not have been finished without enormous help from my advisor, Xiaodong Liu. His guidance allowed me to explore the unknown territory of econometrics that eventually became my new intellectual home.

I am also grateful for the discussions that I had with Adam McCloskey and Carlos Martins-Filho. They turned my eyes toward new directions and gave me fresh viewpoints. I thank Scott Savage and Stephen Billings for providing advice from a perspective outside of pure econometrics.

My thanks also go to Alexander Bentz, who shed light on a language that unfortunately I was not born with.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Generalized Spatial Autoregresseive Model: A Unified Model for Complementarity and Conformity Peer Effects

Since the introduction of the spatial autoregressive model (SAR) for estimating geographical relationships (Anselin 1988; Blommestein 1983), the model's usage has been expanded to various fields, including analysis of social networks and peer effects. In earlier studies, simpler specifications such as linear-in-means or Euclidean distance models were used to represent underlying networks. Due to the increased interest in peer effects in empirical studies and recent progression in data collection techniques, more complicated specifications of networks have become available for econometricians, and estimation strategies utilizing them have also been considerably improved.

A peer effect in social networks is interpreted as an interdependency between individuals in their behaviors. Unlike geographical objects and their relationship within geographical spaces, people have more sophisticated motives for influencing each other. Specifically for social networks, the peer effect can be categorized into complementarity and conformity peer effects based on an individual's underlying motives. Under complementarity, people enjoy extra utility by engaging in the same action with their peer groups. On the other hand, under conformity, people get disutility by deviating from their peers. For both cases, one can expect that people will tend to behave similarly as a result of the peer effect. Despite this similarity, their policy implications are vastly different.

Consider an organizer of a microfinance program, which will be employed as an application in the next chapter, who wants to introduce it to a specific village. The core idea of microfinance is that people's living conditions can be improved by creating peer groups that encourage each other to spend more responsibly. To initiate the program, the organizer may want to exploit peer effects: She could spread microfinance more efficiently and quickly by encouraging an influential group of people to join the program in advance. As they actively participate in the program, their friends or relatives will find it more attractive due to the complementary benefit from the already existing pool of participants. Furthermore, the participation of the initial group will be increased even more. Through this positive feedback loop, the overall participation rate will grow exponentially. This scenario is an example of a positive "social multiplier" effect. If a process of positive feedback exists, a policymaker can exploit it to achieve her policy goal more efficiently. In this case, her job will be to carefully choose the target group based on criteria such as network centralities.

However, this story implicitly assumes that the peer effect is based on complementarity. If the individuals in the initially targeted group are willing to conform with their peers, many of them will soon leave the program before generating a multiplier effect. If the conformity motive is overwhelming, all of them will exit, and there will be a negative impact on the participation rate. In this case, there will be no aforementioned positive feedback. Also, the social multiplier effect will be negative because the participation of the target group has decreased.

The key to the successful use of peer effect is, therefore, assessing the exact nature of the peer effect in a target group, which requires an appropriate econometric model to estimate it. Unfortunately, the standard SAR model is unable to distinguish the different motives and assumes only the complementarity motive. Thus, even when an estimate shows evidence of a strong peer effect, it does not necessarily suggest that the policymakers can expect a positive social multiplier.

To address this issue, I propose the generalized spatial autoregressive model (GSAR). The main idea is to introduce a **conformity parameter** and estimate the degree of conformism along with the intensity of the overall peer effect. Specifically, two parameters are of core interest: the peer effect and the conformity parameter. The former is identical to the "endogenous effect" in the SAR literature. It measures the overall intensity of the peer effect. Meanwhile, the latter measures the degree of the conformity motive with a real number between zero and one: a conformity parameter closer to one suggests a higher degree of conformism. As a result, a combination of these two parameters allows an econometrician to capture the true nature of the peer effect. For example, if a conformity parameter is very close to one, while the peer effect parameter is low, it can be interpreted as a weak conformity peer effect. In other words, the conformity parameter determines the "direction" while the peer effect parameter determines the "magnitude" of peer effect.

Another way to view the GSAR model is as a mixture of the standard SAR model and the Laplacian variant of Boucher (2016). Each model corresponds to the case where the conformity parameter is zero and one, respectively. That is, the existing literature only considers the pure complementarity and the pure conformity peer effects separately. Therefore, the GSAR model may be considered as a single generalized framework for the estimation of peer effects.

The most straightforward advantage of this approach is that it is a more realistic characterization of the peer effect. For example, the villagers may join the program partly because of the benefit of cooperating in the microfinance groups and, at the same time, partly because of peer pressure. The GSAR model can capture such situations properly with the flexible conformity parameter.

Another benefit of this flexibility is that no model selection is required. The common approaches either arbitrarily assume a specific motive or rely upon model selection to choose the correct one. Two potential concerns can be raised with these. First, there is no guarantee that one of the pure models will be selected: when each specification is tested against the other, both or neither of the nulls may be rejected. In this case, the researchers are forced to draw another arbitrary conclusion to interpret the result. Second, the pure conformity model requires considerably larger samples due to the smaller variance caused by using the difference of outcome variables as an explanatory variable. Insignificant estimates from small samples may prohibit a researcher from performing model selection reliably.

The estimation performance of the GSAR model becomes clear with Monte Carlo simulations. First, the mean squared error from the true value and the standard error of the conformity parameter are higher when the intensity of peer effect is lower. Second, the mean squared error is higher when the conformity parameter is higher. Lastly, the GSAR model is robust to model selection. In other words, the GSAR model is free from the risk of choosing a wrong peer effect model at the cost of estimation efficiency.

I build a microeconomic foundation for my model with a utility function consisting of network utility and private utility. Specifically, the network utility comprises the complementary benefit and the deviation cost; the former is a utility from taking the same action as the others, and the latter is a disutility from deviating from the others. I show that the parameters for each component in the utility model are identified by the peer effect and the conformity parameters in the econometrics model.

As a next step, I characterize the social multiplier effect as a function of the complementary benefit and the deviation cost. In general, the complementary benefit must be sufficiently higher than the deviation cost to have a positive multiplier. The threshold for the positive multiplier is obtained as a function of both peer effect and conformity parameters; it is characterized by a decreasing function of the sum of squared degrees divided by the sum of degrees, which is parallel to the "tendency to make hubs" centrality of Saberi et al. (2021). The intuition is that it is easier to have a positive multiplier if people are clustered around a small number of individuals.

The rest of the paper is organized as follows. Section 1.1 reviews related literature. Section 1.2 presents the foundation of the GSAR model. In Section 1.3, I conduct a series of Monte Carlo simulations to demonstrate the behavior of the estimator in various conditions. In the next section, its microeconomic foundation is examined with the implication to the social multiplier effect. Lastly, Section 1.5 provides concluding remarks.

1.1 Related Literature

This paper is mainly motivated by the literature on the conformity SAR model. As pointed out by Boucher and Fortin (2016), the two motives cannot be identified with the ordinary SAR model where its weight matrix is an adjacency matrix with zero diagonal elements. This negative identification result is attributed to the fact that the underlying microeconomic foundations for those motives are observationally equivalent. To address this, (Ushchev and Zenou 2020; Patacchini and Zenou 2012) suggest the "local-average" model. In these studies, the authors model conformity as an additional utility from the average effort of one's peer group, which is reflected in a row-sum normalized weight matrix. By doing so, they distinguish conformity from complementarity, where the additional utility comes from the aggregated effort (the "local-aggregate" model). This approach is further sophisticated by Liu, Patacchini, and Zenou (2014). They present a microfoundation for the local-average model and perform the J-test to choose a suitable peer effect for a given dataset. This branch of literature, however, has a limitation in that the social multiplier arises even under the conformity motive. It is attributed to the fact that the marginal cost from deviation is a constant function with respect to the number of one's peers.

Boucher (2016) takes a different approach by adopting the graph Laplacian as a weight matrix for the pure conformity peer effect. In his model, the marginal cost from deviation increases with the number of peers. As a result of this feature, the social multiplier effect does not exist, which differentiates the model from the previous literature.

This paper is related to the line of the social multiplier literature. The inflection problem of Manski (1993) draws attention to the identification of the "endogenous" and the "contextual" effect because only the former generates a positive social multiplier. He shows that those two effects are not identified with the linear-in-means model under complete networks. As a solution, utilizing more detailed specification of networks is suggested (Kelejian and Prucha 1998a, 2010; Lung-fei Lee 2003), and the existence of the multiplier effect is also identified as a result. Nevertheless, the endogenous effect of the standard SAR model is still based on the complementarity peer effect, and the multiplier effects under different motives are not fully explained.

The existence of the social multiplier effect based on the linear-in-means model is empirically studied by Glaeser, Sacerdote, and Scheinkman (2003). In their work, the authors find evidence of the multiplier in the form of $1/(1-x)$ from multiple datasets. The limitation of this conventional form is that, as in Manski (1993), a network is assumed to be complete. Its result is a significantly higher multiplier for most cases, considering social networks are generally sparse. Indeed, (Giorgi, Frederiksen, and Pistaferri 2019) shows that such a measure can be misleading under more realistic network structures. The definition of the social multiplier effect proposed in this paper is rooted in the marginal effect analysis employed by many social network studies (Liu, Patacchini, and Rainone 2017; Chomsisengphet, Kiefer, and Liu 2018). This common notion of the multiplier under general networks is extended to incorporate the distinct impacts of the complementarity and conformity motives with a single metric.

1.2 Model

Consider a set of samples with N individuals who are linked to each other by a non-stochastic and exogenous social network. The network is assumed to be symmetric, which means that if an individual i is a peer of the other individual j, j is also a peer of i. Such a network is described by a graph on a set of vertices, V, and a nonempty set of edges, E, which is denoted as $G_N = G_N(V, E)$, where the edges are unweighted and undirected. Each vertex corresponds to an individual in a network. Self-loops or multiple edges between the same pair of vertices are prohibited, but vertices may be isolated. A set of vertices that are connected, or adjacent, to vertex i is called a neighborhood, and its size is referred to as a degree of i, d_i. The adjacency between the vertices is denoted as a binary variable, $a_{ij} \in \{0, 1\}$, which is zero if there is no edge between i and j and unity if there exists one. This relationship is summarized into a single $N \times N$ adjacency matrix, A. Note that, by construction, its diagonal is all zero. A degree matrix, D, is a diagonal matrix of which i-th diagonal element is equal to i's degree, d_i. More specifically, I define the degree of i as a sum of the i-th row of A.[1] In short, a network, G_N, is summarized into an adjacency matrix and a degree matrix defined as follows.

1. This is equivalent to the "out-degree" of the graph theory literature. This specification does not make any difference for the rest of the arguments, as I assume a symmetric network, where $a_{ij} = a_{ji}$.

Definition 1.1.

$$A = \begin{cases} [a_{ij}], & \textit{if } i \neq j \\ 0, & \textit{otherwise} \end{cases} \qquad \textit{(Adjacency matrix)}$$

$$D = \begin{cases} [d_{ij} = \sum_{k=1}^{N} a_{ik}], & \textit{if } j = i, \\ 0, & \textit{otherwise} \end{cases} \qquad \textit{(Degree matrix)}$$

For the SAR models, the specification of peer effect is summarized into a weight matrix. For my model, it is defined as a difference between the adjacency and the degree matrices multiplied by **a conformity parameter**, $\gamma \in [0, 1] \subset \mathbb{R}$.[2]

Definition 1.2. *A generalized weight matrix derived from the network structure, G_N, is an $N \times N$ square matrix defined as follows:*

$$W(\gamma) = A - \gamma D,$$

I propose the generalized spatial autoregressive model (GSAR) using the weight matrix defined above.

$$y = \rho W(\gamma) y + X\beta + \varepsilon, \tag{1.1}$$

where X is an $N \times K$ covariate matrix. The outcome variable, y, is an $N \times 1$ vector of real numbers. The vector of direct coefficients, β, is a $K \times 1$ vector and measures the direct effect of one's attributes on y. The idiosyncratic error, ε, is a random variable with a zero mean independent of the individual characteristics and the network structure, and no particular distribution is imposed. The parameter of interest is $\theta = (\rho, \gamma, \beta')'$.[3]

The intensity of peer effect is measured by the parameter, ρ, in (1.1). The first assumption is a constraint on this **intensity parameter**. A peer effect is referred to as "explosive" if it is so strong that there is no unique equilibrium, which should be avoided.

2. This definition can be considered as a type of the negative generalized graph Laplacian, where its diagonal may be any nonnegative numbers.

3. Note that this definition encompasses both the ordinary SAR model ($\gamma = 0$) and the graph Laplacian variant of Boucher (2016) ($\gamma = 1$) as special cases.

Assumption 1.1. *(Model stability) The domain of the peer effect parameter is defined by the inverse of the spectral radius of $W(\gamma)$. That is,*

$$0 < \rho < 1/|\xi(\gamma)|,$$

where $\xi(\gamma)$ is the largest eigenvalue of $W(\gamma)$ in absolute value.

Since the weight matrix is symmetric, its eigenvalues are all real numbers, and, accordingly, the spectral radius can always be found. If Assumption 1 is satisfied, then the inverse of $I - \rho W(\gamma)$ exists, and the model outcome is derived as a reduced form,

$$y = \big(I - \rho W(\gamma)\big)^{-1}(X\beta + \varepsilon). \tag{1.2}$$

The IV estimation of Kelejian and Prucha (1998a) is not feasible for the GSAR model because the proposed instrument variables are also the objects of estimation. Therefore, the IV estimator is infeasible for the GSAR model. Also, any estimators that rely upon the weight matrix, such as the GMM estimator (Lung-fei Lee 2007), are also unusable. Therefore, I employ the nonlinear least square estimator for the GSAR model.

The parameters of the GSAR model is estimated by minimizing the following sum of squared residuals:

$$SSR(\theta) = \Big(y - \big(I - \rho W(\gamma)\big)^{-1} X\beta\Big)'\Big(y - \big(I - \rho W(\gamma)\big)^{-1} X\beta\Big).$$

Assumption 1.2. *Let T_θ the derivative of $(y - (I - \rho W(\gamma))^{-1}X\beta)'$ with respect to each parameter in θ. Then, $T_\theta\big(I - \rho W(\gamma)\big)^{-1}[X, AS^{-1}X\beta_0, -DS^{-1}X\beta_0]$ has full rank for feasible values of ρ and γ in their compact parameter space Ω.*

Proposition 1.1. *The parameters of the GSAR model, $\theta = (\rho, \gamma, \beta')'$, are identified under Assumption 1.1-2.*

Proof. See Appendix A. □

1.3 Monte Carlo Simulation

In this section, I demonstrate the GSAR model with a series of Monte Carlo simulations. The main goal is twofold: to investigate the finite sample behavior of the conformity parameter estimator and to delineate the bias caused by choosing a wrong peer effect model.

Regarding the first, the estimation of the conformity parameter depends on the magnitude of the intensity parameter since the former is multiplied by the latter. In principle, it is identified as long as the intensity is strictly greater than zero. However, a significantly larger sample will be necessary in practice if the peer effect is weak in intensity.

Next, the GSAR model is robust to model selection. Expectedly, a correct model yields the most efficient result if a true motive is known. On the other hand, it will lead to biased estimates if a wrong peer effect model is used, or even the OLS estimator is used, ignoring the peer effect. Through simulations, I show the directions of the biases caused by the wrong models and their implications.

1.3.1 GSAR Estimation

For the following simulations, I generate exogenous, random symmetric networks. I consider the following DGP:

$$y = \rho_0(A - \gamma_0 D)y + X\beta_0 + \varepsilon, \tag{1.3}$$

where X is an $N \times 3$ vector with $\beta_0 = (1, 0, -1)$. Each link is generated by a Bernoulli distribution with $P(a_{ij} = 1) = P(a_{ji} = 1) = 0.05$, which reflects the sparsity of typical social networks. The k-th exogenous covariate for an individual i is generated by the normal distribution with a mean of zero and a standard deviation of 4. The error term for i is i.i.d, following the normal distribution with a mean of zero and a standard deviation of 2.

In Table 1.1, I summarize the simulation result from 1,000 samples with 1,000 replications.[4]

4. A constraint is imposed on the conformity parameter so that its estimates stay between zero and one. While it does not alter the result for larger samples, it prevents having unreasonable values for the intensity or conformity parameter from smaller ones.

Table 1.1: Monte Carlo simulation of the GSAR model

	$\rho_0 = .05$		$\rho_0 = .1$		$\rho_0 = .2$	
	Estimates	MSE	Estimates	MSE	Estimates	MSE
$\gamma_0 = .2$						
ρ	0.050 (.009)	0.000	0.100 (.010)	0.000	0.201 (.012)	0.000
γ	0.265 (.249)	0.066	0.221 (.145)	0.021	0.203 (.081)	0.006
β_1	1.006 (.026)	0.001	1.004 (.029)	0.001	1.002 (.031)	0.001
β_2	0.000 (.017)	0.000	0.000 (.016)	0.000	-0.001 (.015)	0.000
β_3	-1.005 (.026)	0.001	-1.004 (.029)	0.001	-1.001 (.031)	0.001
$\gamma_0 = .5$						
ρ	0.051 (.010)	0.000	0.100 (.011)	0.000	0.200 (.014)	0.000
γ	0.526 (.286)	0.083	0.498 (.188)	0.035	0.497 (.098)	0.010
β_1	1.002 (.028)	0.001	1.000 (.034)	0.001	1.000 (.035)	0.001
β_2	0.001 (.016)	0.000	-0.001 (.017)	0.000	-0.001 (.017)	0.000
β_3	-1.002 (.028)	0.001	-1.000 (.035)	0.001	-1.000 (.035)	0.001
$\gamma_0 = .8$						
ρ	0.051 (.010)	0.000	0.099 (.012)	0.000	0.200 (.018)	0.000
γ	0.732 (.280)	0.083	0.770 (.187)	0.036	0.791 (.122)	0.015
β_1	0.996 (.027)	0.001	0.996 (.033)	0.001	0.998 (.041)	0.002
β_2	0.000 (.018)	0.000	0.000 (.018)	0.000	0.001 (.019)	0.000
β_3	-0.995 (.028)	0.001	-0.996 (.032)	0.001	-0.998 (.042)	0.002

Standard errors in parentheses
$MSE = (\hat{\theta} - \theta_0)^2/1000$
$N = 1000$ with 1,000 replications

As mentioned earlier, the estimation efficiency of the conformity parameter varies upon the intensity of peer effect. Higher intensities tend to reduce the standard error and MSE of the estimates of ρ and γ, but not necessarily β.

The additional dimension imposed by the conformity parameter reduces the finite sample performance of the estimator compared to the pure models, and it is more pronounced under 1) higher γs and 2) lower ρs. Due to its interaction with ρ, γ appears to have higher MSE.

1.3.2 Misspecified Models

Also, I show the simulation results for misspecified models to clarify the direction of biases depending on the estimation models. First, I estimate the model with the OLS estimator. This shows the bias caused when the peer effect is ignored by a researcher. Moreover, the direction of bias appears different depending on the underlying motive of peer effect.

The simulation environments are identical to the former one, except that the intensity parameter is fixed at 0.15. According to the true model, the true conformity parameter, γ_0, will be either zero or one.

The results are summarized in Table 1.2.

If the true model is pure complementarity ($\gamma_0 = 0$), the OLS tends to overestimate the direct effect of one's attributions. This means that the higher outcome, which results from the social multiplier effect of the complementarity peer effect, is interpreted as if it is directly from one's characteristics. This type of bias leads a researcher to a false conclusion that these individuals will show similarly high outcomes even when separated from their social networks.

When it is estimated by the conformity model, the estimates show a weaker intensity and much higher direct effects. As Boucher (2016) suggests, the conformity motive tends to reduce the variance of outcome variables. In other words, individuals will move toward the social norm of their social networks. As a compensation for this tendency, the contribution of individual characteristics is severely overestimated.

The weaker intensity of the estimates can be explained by the structure of the weight matrix. The graph Laplacian matrix, which is used as the weight matrix in the pure conformity model, also reflects the interdependency between the samples. However, it will be suppressed by the existence

Table 1.2: Simulation: misspecified models

	OLS		Complementarity		Conformity		GSAR	
	Estimates	Bias	Estimates	Bias	Estimates	Bias	Estimates	Bias
$\gamma_0 = 0$								
ρ			0.150 (.007)	-.001	0.160 (.020)	.010	0.153 (.008)	.003
γ							0.042 (.060)	.042
β_1	1.041 (.020)	.041	1.000 (.016)	.000	1.274 (.042)	.274	1.011 (.022)	.011
β_2	0.000 (.019)	.000	0.000 (.015)	.000	0.000 (.020)	.000	0.001 (.015)	.001
β_3	-1.041 (.021)	-.041	-1.000 (.016)	.000	-1.274 (.043)	-.274	-1.011 (.022)	-.011
$\gamma_0 = 1$								
ρ			0.113 (.010)	-.037	0.151 (.016)	.001	0.149 (.016)	-.001
γ							0.908 (.128)	.089
β_1	0.839 (.021)	-.161	0.824 (.021)	-.176	1.000 (.022)	.000	0.982 (.033)	-.018
β_2	0.000 (.017)	.000	0.000 (.016)	.000	0.000 (.019)	.000	0.000 (.018)	.000
β_3	-0.839 (.022)	.161	-0.824 (.022)	.176	-1.000 (.022)	.000	-0.982 (.033)	.018
$\gamma_0 = .14$								
ρ			0.141 (.008)	-0.009	0.160 (.016)	0.010	0.151 (.010)	0.001
γ							0.155 (.095)	0.005
β_1	1.001 (.019)	0.001	0.968 (.016)	-0.032	1.223 (.036)	0.223	1.004 (.028)	0.004
β_2	0.000 (.018)	0.000	0.000 (.015)	0.000	0.000 (.020)	0.000	0.001 (.016)	0.001
β_3	-1.002 (.020)	-0.002	-0.968 (.017)	0.032	-1.224 (.037)	-0.224	-1.004 (.029)	-0.004

Standard deviations in parentheses
$\rho_0 = .15$, $\beta_{10} = 1$, $\beta_{20} = 0$, $\beta_{30} = -1$

of the degree matrix and reflected in the weaker intensity of peer effect.

The opposite case, where the true model is conformity, shows the exact opposite tendency. The OLS underestimates the direct effect. The outcome variables were attenuated by the conformity motive. When the pure complementarity model is used, it shows an even lower direct effect. This is because the model attempts to explain the outcome variables as a consequence of the social multiplier effect. For both cases, the estimates are unbiased when correct models are used. More importantly, the GSAR model captures the motive with the conformity parameter without any noticeable bias. Moreover, the bias is more noticeable as the intensity of the peer effect increases.

The last example, $\gamma_0 = .14$, is a coincidental case where the OLS estimates are unbiased. Both of the two pure models show biases, while only the GSAR model correctly captures both intensity and motive.

In all of the cases, the GSAR model is robust to the model specification at the cost of higher standard errors.

1.4 Microfoundation

Now, I provide and analyze a microfoundation of the econometric model (1). The utility function of any given individual i is represented as a quadratic utility function that consists of linear benefits and quadratic costs.

$$u_i(y_i, y_{-i}, x_i, \varepsilon_i) = \underbrace{\underbrace{\lambda\left(\sum_{j\neq i} a_{ij} y_j y_i\right)}_{\text{Complementary benefit}} - \underbrace{\frac{\delta}{2}\left(\sum_{j\neq i} a_{ij}(y_j - y_i)^2\right)}_{\text{Deviation cost}}}_{\text{Network net benefit}} + \underbrace{\left((x_i\beta + \varepsilon_i)y_i - \frac{1}{2}y_i^2\right)}_{\text{Private net benefit}}, \tag{1.4}$$

where $\lambda > 0$, $\delta \geq 0$ and a_{ij} is the (i, j) element of A. Given this, player i chooses an action, y_i, to maximize his or her utility, u_i. The network net benefit of i is maximized for the set of actions of the other players, y_{-i}. The private net benefit is determined by his or her own attributes, x_i, and an idiosyncratic shock, ε_i. The benefit arises from complementarity, which has been considered by a few studies (Ballester, Calvó-Armengol, and Zenou 2006; Bramoullé, Kranton, and D'Amours 2014). The players benefit from each others' engaging in the same activity if they are linked by an edge, a_{ij}, in a network. On the other hand, the deviation cost measures the disutility from taking different actions from one's peers with a quadratic function (Boucher 2016).

The first-order condition of (3) induces each player's best-response function.

$$\begin{aligned} y_i^* &= \Bigg(\sum_{j\neq i} a_{ij}\big((\lambda+\delta)y_j - \delta y_i\big)\Bigg) + x_i\beta + \varepsilon_i \\ &= (\lambda+\delta)\Bigg(\sum_{j\neq i} a_{ij}\big(y_j - \frac{\delta}{\lambda+\delta}y_i\big)\Bigg) + x_i\beta + \varepsilon_i \\ &= \rho\Bigg(\sum_{j\neq i} a_{ij}\big(y_j - \gamma y_i\big)\Bigg) + x_i\beta - \varepsilon_i, \end{aligned}$$

where $\rho = \lambda + \delta$ and $\gamma = \delta/(\lambda+\delta)$. The pair of parameters from the utility function, (λ, δ) is identified by those from the estimation model, (ρ, γ).

Proposition 1.2. *The complementary benefit, λ, and the deviation cost, δ, from the microfoundation (3) uniquely determine the peer effect parameter, ρ, and the conformity parameter, γ, from the econometric model (1), and vice versa.*

Proof. Given (ρ, γ), the complementary benefit and the deviation cost are obtained by solving the following system of equations:

$$\begin{cases} \lambda = \rho - \rho\gamma \\ \delta = \rho\gamma. \end{cases} \tag{1.5}$$

Suppose that there exists another pair, $(\rho', \gamma') \neq (\rho, \gamma)$, which satisfies (4). Then, $\delta = \rho\gamma = \rho'\gamma'$ and $\lambda = \rho - \rho\gamma = \rho' - \rho'\gamma'$. Using the first equation, the latter is rewritten as $\rho'\gamma' + \rho = \rho'\gamma' + \rho'$. Then, it is obtained that $\rho = \rho'$, which is a contradiction and proves the uniqueness of ρ. Trivially, γ is also uniquely determined. Similarly, the opposite direction also holds. □

Then, if the conformity parameter, γ, is equal to zero, it corresponds to the case where the deviation cost parameter, δ, is zero. In that case, the standard SAR model estimates the pure complementarity peer effect. In contrast, if γ is equal to one, the entire portion of the peer effect, ρ, is equal to the deviation cost, and the model connects to the pure conformity case of Boucher (2016).

This identification result is derived from the distinguished structure of the complementary benefit and the deviation cost. In Boucher and Fortin (2016), the deviation cost is modeled as a quadratic function, $-\frac{\delta}{2}(y_i - \sum_{j \neq i} a_{ij} y_j)^2$. Note that the square of the difference is taken with y_i and the aggregated outcome of the peers. Therefore, the marginal deviation cost does not increase with respect to d_i, which is the same structure with the complementary benefit. Then, this pure conformity model leads to a SAR model, which is observationally equivalent to the pure complementarity model. As noted in Boucher (2016), this issue is shared by the local-average model as well. Due to its distinguished structure of the benefit and the cost, the GSAR model can incorporate the two components together and identify them at the same time.

An advantage of this approach is that the net benefit from the peer effect can be described as a tension between the complementary benefit and the deviation cost. By allowing the relative magnitudes of these components to be flexible, the two pure models become two extrema of the proposed generalized SAR model, and the characteristic of a peer effect is determined by the dynamics between the network benefit and cost from one's peers. The other advantage is that the GSAR model does not require model selection. The model selection procedure assumes that either of the models must be true and would not provide any useful information if both models are rejected. Allowing the mixture of the two models can prevent the complexity accompanying model selection.

Figure 1.1: A star network with $N = 3$

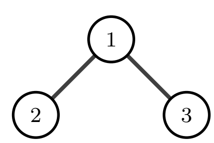


Figure 1.2: The pure complementarity peer effect

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}; \quad X\beta = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \quad (I - 0.15A)^{-1} X\beta = \begin{bmatrix} 1.05 \\ 0.16 \\ 0.16 \end{bmatrix}$$

A simple star network with three vertices is focused on further elucidating the difference between the existing models (Figure 1). From this, an adjacency matrix and a graph Laplacian

Figure 1.3: The pure conformity peer effect

$$A - D = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}; \quad X\beta = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \quad (I - 0.15(A - D))^{-1} X\beta = \begin{bmatrix} 0.80 \\ 0.10 \\ 0.10 \end{bmatrix}$$

can be induced, with associated reduced forms (Figure 2 and 3). Suppose everyone's ex-ante outcome without peer effect is $(1, 0, 0)$. The first vertex may be considered as an "injection" point, which means that he or she is a target individual of a policymaker. Then, the ex-post outcomes are obtained by multiplying the inverse resolvent matrix respectively, as $(1.05, 0.16, 0.16)$ for the complementarity and $(0.80, 0.10, 0.10)$ for the conformity model. The evidence of a positive social multiplier is found from the increased outcome of the first vertex in the former. Meanwhile, in the latter, everyone has moved toward each other, and the outcome of the first vertex has decreased, which suggests a negative social multiplier.

Next, consider the local-average model with a row-normalized adjacency matrix (Figure 4). This simple example demonstrates that the local-average model does not precisely describe the conformity peer effect. The aggregated outcome of the vertices is still higher than the ex-ante outcome despite the conformity motive it attempts to describe.[5]

Figure 1.4: The local-average model

$$A_{norm} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}; \quad X\beta = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \quad (I - 0.15A_{norm})^{-1} X\beta = \begin{bmatrix} 1.02 \\ 0.15 \\ 0.15 \end{bmatrix}$$

1.4.1 Social Multiplier Effect

As previously illustrated, the magnitude of the social multiplier effect depends not on the structure of networks but also on the underlying motive of peer effect. However, the earlier literature

5. This, however, does not suggest that the local-average model is not useful in any case. If one is interested in the complementarity benefit from the average behavior of a group, it shows distinguished patterns compared to the local-aggregate model.

on the social multiplier only considers the simplest form of multiplier, which is equivalent to $1/(1-\rho)$. This is rooted in the linear-in-means model with complementarity peer effect. Indeed, one of its limitations is that it is based on a complete network because it is well-known that typical social networks are sparse. Furthermore, if the true motive of peer effect is conformity, any attempts to capture peer effect using this type of multiplier will fail because there can be significant conformity peer effect even when the multiplier is close to zero.

To address this complexity, I propose an alternative measure for the social multiplier incorporating the different motives.

The nature of the multiplier effect can be further investigated by focusing on the marginal effect of covariates on outcome variables.[6] Consider the reduced form of the model (2),

$$y = (I - \rho W(\gamma))^{-1}(X\beta + \varepsilon).$$

For simplicity, assume that X is an $N \times 1$ vector and β is a scalar. Then, the marginal effect of i's covariate on all of the players is,

$$\frac{\partial y}{\partial x_i} = (I - \rho W(\gamma))^{-1}\beta. \tag{1.6}$$

Intuitively, the marginal effect from β is multiplied by the inverse matrix of $I - \rho W(\gamma)$, which means that the inverse matrix dictates the shape of the overall multiplier effect. By definition of a matrix inverse, it can be decomposed into

$$(I - \rho W(\gamma))^{-1} = \frac{1}{\det(I - \rho W(\gamma))} W(\gamma)_{adj},$$

where $W(\gamma)_{adj}$ is an adjugate matrix of $W(\gamma)$. Since the ij-th element of an adjugate matrix represents the number of paths between i and j, the overall multiplier effect will be rooted in the magnitude of the determinant. From this, a definition of the social multiplier is proposed as follows:

Definition 1.3. *Given a nonstochastic, exogenous G_N, the social multiplier is defined as a function of ρ and γ, which is*

$$\mu(\rho, \gamma) = 1/|(I - \rho W(\gamma))|.$$

6. The idea of the marginal effect multiplied by $(I - \rho W)^{-1}$ has been widely used in the literature of the SAR model (Chomsisengphet, Kiefer, and Liu 2018) (add more citations)

If $\mu > 1$*, there is a positive social multiplier effect. If* $0 < \mu < 1$*, there is a negative social multiplier effect.*

The boundaries for the pure models are derived as follows:

Corollary 1.1. $\mu(\rho, 0) > 1$ *and* $0 < \mu(\rho, 1) < 1$

Proof. For $\gamma = 1$, it is suffice to show that $|I - \rho W(1)| > 1$. For any arbitrary matrix, M, and its eigenvalues, $\lambda_i(M)$, a determinant of M is equal to $\Pi_{i=1}\xi_i$. For any eigenvalues of M, $1 + \xi_i(M) = \xi_i(I + M)$ and $c\xi_i(M) = \xi_i(cM)$ hold. Since a graph Laplacian matrix is positive semidefinite, all of its eigenvalues are nonnegative. Then, $|\rho M| = \rho(\Pi_{i=1}\xi_i) > 0$ and $|I - \rho W(1)| > 1$.

For $\gamma = 0$, $|I - \rho W(0)|$ is equal to the difference between 1 and the sum of paths of all lengths discounted by ρ. Therefore, $0 < |I - \rho W(0)| < 1$ by Assumption 2. □

As the conformity parameter is flexible in the GSAR model, the social multiplier is hinged upon the relative magnitudes of the two motives. Accordingly, a threshold where the social multiplier is exactly equal to 1 can be found.

Proposition 1.3. *The threshold for the positive social multiplier is a solution of* $\mu(\rho, \gamma) = 1$*, which is*

$$\bar{\rho}(\gamma) = \frac{2\gamma}{1 + \gamma^2 \sum_{i=1}^N d_i^2 / \sum_{i=1}^N d_i}.$$

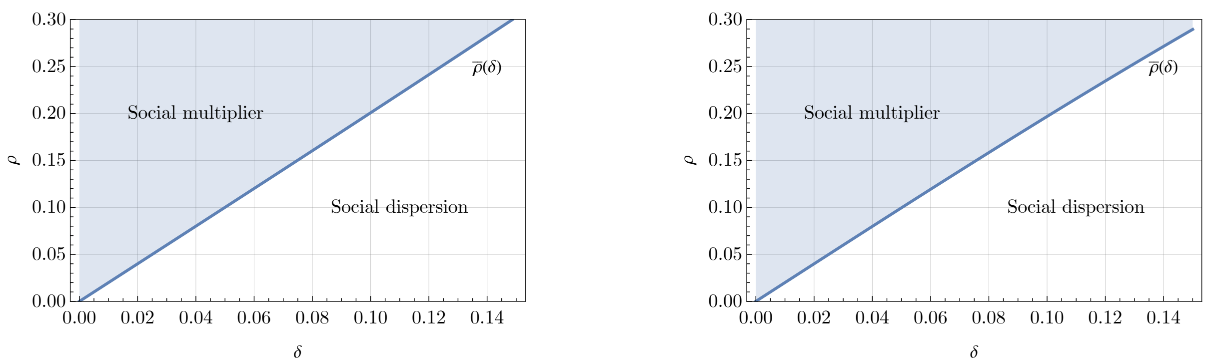
If $\rho > \bar{\rho}$*, there exists a positive social multiplier effect, and* $\rho < \bar{\rho}$*, there exists a negative social multiplier effect.*

Proof. See Appendix A. □

This result suggests that, for the positive social multiplier effect, the complementary benefit must be sufficiently higher than the conformity motive of individuals. Note that the coefficient, $\sum_{i=1}^N d_i^2 / \sum_{i=1}^N d_i$, is parallel to the "tendency to make hubs centrality" proposed by Saberi et al. (2021). The implication is that the more vertices are linked with a specific "hub" vertex, the easier for the social multiplier effect to arise.

The threshold is based on the sparse approximation of matrix determinants proposed by Ipsen and Lee (2011). The performance of approximation depends on the sparsity of non-diagonal elements of a matrix, and, indeed, most of social networks are qualified as sparse. A graphical presentation of a numerical and an approximated solution for a random graph with $N = 10$ and a density of 10% is shown in Figure 5. The shaded areas represent where the multiplier effect

Figure 1.5: The numerical (left) and the approximated solution (right)



Note: The graph is generated randomly with $N = 10$ and density of 10%.

is positive. The example shows that the approximated solution is tracking the actual solution relatively well, even for the higher density than the graphs in the real world.

1.5 Conclusion

To integrate the pure complementarity and conformity models, I propose the generalized SAR model that simultaneously estimates the underlying motives and intensity of peer effect. The identification result for the model parameters is provided, and the model is estimated by the nonlinear least squares estimator.

To show the model's estimation performance, I perform a set of Monte Carlo simulations. The conformity parameter is estimated more accurately when peer effect is stronger. Furthermore, the GSAR model is robust to model selection. If a true motive is unknown or does not belong to either of the pure cases, the GSAR performs better than the existing models despite the efficiency loss caused by the additional conformity parameter.

I support the GSAR model by the microfoundation where an individual's utility is described as a function that consists of both benefit and cost from the actions of her peers. Moreover, the relative magnitude of the benefit and cost corresponds to the conformity parameter in the estimation model. I provide a threshold of a positive social multiplier, which is a function of the intensity and conformity parameters.

Chapter 2

Revisiting the Microfinance in India: An Application of the GSAR Model

The authors of Banerjee et al. (2013), BCDJ, henceforth, collaborated with Bharatha Swamukti Samsthe (BSS) to study the diffusion process of microfinance in rural villages in Karnataka, India. They reveal the role of the villagers who do not take up the microfinance program but only pass through the program-related information. Although they transferred the information with a possibility not as high as the people who joined the program, their larger number played a more significant role in the diffusion of microfinance.

In contrast, their result implies that the peer effect, or the "endorsement effect" in their terminology, does not exist among the villagers. In other words, they do not change their decision depending on whether the information is coming from actual participants or not. Moreover, depending on their diffusion parameters, the peer effect even appears to be slightly negative.

Considering that microfinance is new to the villagers, and not many of them were aware of the real benefit of it, it is not unfathomable that there should be extra credibility in the information when it is coming from actual users. Indeed, while no clear explanation is provided in the original paper, Chandrasekhar and Lewis (2011) implies that underspecified networks can be a possible explanation of the weak peer effect.

The purpose of this chapter is to propose an alternative, but not exclusive, approach to identify and measure peer effect in the case of BCDJ by applying the GSAR model introduced in the first chapter. It is distinguished from the diffusion model by tri-fold: 1) an assumption of stable equilibrium, 2) consideration of the underlying motives of peer effect and 3) inclusion of multiple dimensions of social networks.

First, the diffusion model used in BCDJ is the SIR (Susceptible-Infected-Recovered) model. Here, no one is initially aware of the microfinance program (susceptible). Once an individual receives the information about the program, she makes a decision (infected). After that, she passes the information to her peers and is removed from the network (recovered). A limitation of this diffusion model is that an individual has no further chance to revert her decision later, so the direction of diffusion is unidirectional. While this is a necessary compromise for estimating a diffusion pattern, it does not allow the social multiplier effect to arise.

In contrast, by the stable equilibrium assumption, everyone is aware of the existence of the microfinance program and participates depending on their own characteristics and the expected actions of their peers. This is an essential feature of the family of spatial autoregressive models. At the cost of forgoing the explanation of the diffusion pattern over time, it allows a longer time period between receiving information and making decisions.

For estimation, I extend the GSAR model in Chapter 1 for binary outcome variables by adopting the rational expectation model (Lee, Li, and Lin 2014; Yang and Lee 2017; Yang, Lee, and Qu 2018). This line of literature assumes that the players have complete information on network structures and attributes of their peers.

Second, the function used in BCDJ to estimate each individual's choice variable implicitly imposes the complementarity motive. In the previous chapter, I discussed that peer effect may be motivated by either complementarity or conformity, and using a wrong model may lead to biased estimates. Indeed, estimation results from the GSAR model show that the conformity motive dominates the villagers' peer effect.

Last, many observations imply that one's social network consists of multiple dimensions, such as coworkers, friends or relatives. In fact, studies suggest that such dimensions may channel different types of peer effect (Kitts and Leal 2021). For example, people may talk about microfinance with their friends but not relatives. Even further, it is possible that separate dimensions are driven by distinguished motives.

Despite the advantage of including multiple network dimensions and measuring the peer effect on each of them, applying this approach to diffusion models is challenging. At the cost of forgoing modeling a diffusion process, focusing on a stable equilibrium may allow a researcher to investigate the peer effect that differs by the separate dimensions.

For this purpose, I extend the GSAR model by employing the higher-order SAR model (Lee and Liu 2010; Debarsy and J. LeSage 2018; Debarsy and J. P. LeSage 2022). In this model, an aggregate weight matrix is represented as a weighted sum or a convex combination of weight matrices that represent the multiple dimensions.

In many cases, an obstacle to the use of the higher-order model is data limitation. Fortunately, the dataset collected by BSS is a rare case containing rich data on the multiple dimensions of the villagers' social networks. The original study has left this opportunity unexplored, as its primary interest is on the diffusion process.

I summarize the results from the GSAR estimation as follows: First, the peer effect of high intensity on participation in the microfinance program exists among the villagers. Second, the dominating motive of the peer effect is conformity. Third, only three out of eight dimensions of the villagers' social network channel the peer effect. Last, the different dimensions may be driven by different motives. These findings contrast the original study of BCDJ, which implies that there is no peer effect and the villagers only pass information.

In the last section of this chapter, I conduct a counterfactual analysis to identify the most influential individuals in networks. Before BSS began introducing the program to the villagers, the organizers had chosen a group of "leaders" who were considered high-profile based on their socioeconomic status. These leaders are expected to spread the information and bring their peers to the program. I evaluate their performance as injection points for utilizing peer effect and compare theirs with that of the counterfactual leaders found by the "key player" approach of Ballester, Calvó-Armengol, and Zenou (2006).

The key players are found by removing each person one by one and assessing their impact on an aggregate outcome. In this case, the outcome of interest is the aggregate probability of taking up the program, computed by the GSAR model estimates. Using this method, I assign the same number of counterfactual leaders as the BSS leaders and compare the outcomes. It appears that the performance of the counterfactual leaders is 11% higher than that of the leaders designated by BSS.

The rest of this chapter is organized in the following order: in Section 2.1, I describe the data used for the study. In Section 2.2, the methodology is explained with a focus on the difference from the diffusion model. In Section 2.3, the estimation model is presented by extending the GSAR model for binary outcome variables. Section 2.4 provides estimation results. In Section 2.5, I conduct a counterfactual analysis by choosing alternative leader groups. The last section concludes the chapter.

2.1 Data Description

This study utilizes the data set collected by Banerjee et al. (2013) in collaboration with Bharatha Swamukti Samsthe (BSS), a non-profit organization that conducted the microfinance program in Karnataka, India. The authors had conducted a household-level survey prior to the entry of BSS into 43 rural villages in the target districts, including household characteristics such as roof materials, ownership of houses, access to electricity, religion, and castes. An individual-level survey was conducted for the members of households randomly selected among eligible ones to collect more detailed information. As a result, additional data were obtained for about 46% of the sample households. BSS had reported the take-up rate periodically.

The main interest of the individual survey was the social network information. Respondents were asked to provide the names of their peers who belong to a total of twelve dimensions of their peer network respectively as follows: (1) from whom they would borrow money, (2) to whom they would lend money, (3) to whom they give advice, (4) from whom they find help for important decision making, (5) from whom one would borrow kerosine and rice, (6) to whom one would lend kerosine and rice, (7) to whom they visit for free time, (8) who visits them for free time, (9) from whom they seek medical advice, (10) relatives in the village, (11) non-relative people with whom they socialize and (12) people with whom they go to temples together. Note that some networks, such as (1) and (2), are examined bidirectionally. The average overlapping rate of the twelve networks is roughly 50%, which satisfies the linear independence assumption. These network data were collected for each individual and then collapsed into household-level in a way that a pair of households is considered connected if any of the household members are connected. Due to the random sampling, it is possible that a household designated as a neighborhood may not have been sampled, and there is no way to verify if the relationship is reciprocal. For this reason, the surveyed networks are symmetrized. The dimensions surveyed by a set of questions with "directions" such as "from whom they would borrow money" and "to whom they would lend money," are also flattened to a single matrix throughout this paper.[7]

Some villagers were designated as "leaders," mostly consisting of self-help group leaders, teachers, and shop owners. They were initially informed about the microfinance program in private meetings held by the institution and encouraged to invite their neighbors to the following sessions. Households with the leaders are considered leader households.

The entries without additional details such as caste or religion are removed to control the household characteristics.[8] As a result, a total of 7,919 households are used. The descriptive statistics can be found in Figure 1. There is no noticeable difference in the characteristics depending on the sample sizes. The additional characteristics are crucial due to the potential endogeneity between the socioeconomic status variables and the peer networks. In the original paper, individual-level demographics were not used for the analysis. In this paper, caste and religion are controlled by imposing the attributes of the heads of households on each sample. This is not a significant issue, as the characteristics of the other family members are highly likely to be identical to those of their heads.

As pointed out by many studies, random removal of nodes on a network does not guarantee unbiasedness and consistency. If such an issue cannot be avoided, it is advisable to clarify its

7. Note that BSS had surveyed the networks before they entered the villages. Because of this, it is unlikely that the villagers have adjusted their networks according to their decision to join the microfinance program.

8. Christianity is excluded due to its extremely small sample size

expected impacts. Comparing the density and the spectral radius of the networks under different sample sizes, it appears that the density increases while the radius decreases. It may be interpreted as a result of more nodes with smaller degrees and some bridging the others being removed. In this case, a downward bias is expected on the estimates of the peer effects. Consequently, if any estimates of peer effects are found significant under this smaller sample, it can be believed as a lower bound. Also, all 1,672 of the excluded samples are only from the first two villages, so its impact will be contained in those two.[9]

The dummy variables for the income proxies are constructed to present lower statuses, considering that the most demand is from underprivileged households.

2.1.1 Network Characteristics

Definitions and graph statistics of the surveyed networks are summarized in Table 2.2. For the individual networks, the average degree is 3.284, and the average density is 0.023, which is common sparsity for social networks.[10] [11] The union network is constructed by the villagers who are connected through any of the individual dimensions.

2.2 Methodology

By adopting the SAR model, this study departs from Banerjee et al. (2013) of which the main interest is the diffusion process and the rate of transmission of the microfinance program. Exploiting the panel data provided by BSS, they simulate the process over the surveyed network and find that non-participants play a crucial role in the diffusion of the program. While those who adopt the program are more likely to pass the information, the larger number of non-participants is more pivotal, even with the lower transmission rate. In terms of the peer effect, however, they find that the number of participants in one's neighborhood does not significantly affect the diffusion

9. See Table B.1 for full descriptive statistics.

10. The respondents were asked to report up to four persons for each network. It can potentially cause a mismeasurement problem suggested by Griffith (2022). Its potential impact, however, is only an underestimation of peer effect, not an overestimation. If a peer effect turns out to be positive, it is not an issue anymore, which is the case in this study.

11. The degrees over four are due to the directionally surveyed networks merged into a single one.

Table 2.1: Descriptive statistics

Variables	Description	Mean	S.D.
Outcome variable			
Microfinance take-up	Yes = 1, No = 0	0.18	0.38
Income proxies			
Roof material: tile	Yes = 1, No = 0	0.37	0.48
Number of rooms		2.41	1.30
Number of beds		0.82	1.23
No latrine in house	Yes = 1, No = 0	0.73	0.45
House not owned	0, if the house is privately owned, 1, otherwise	0.10	0.30
No access to electricity	Yes = 1, No = 0	0.38	0.48
Social statuses			
Leader group	1, if the household belongs to the leader group, 0, otherwise	0.13	0.33
Other backward castes	Base caste variable	0.53	0.50
General caste	Yes = 1, No = 0	0.12	0.33
Minority castes	Yes = 1, No = 0	0.20	0.13
Scheduled caste	Yes = 1, No = 0	0.28	0.45
Scheduled tribe	Yes = 1, No = 0	0.05	0.22
Religions			
Hinduism	Base religion variable	0.95	0.21
Islam	Yes = 1, No = 0	0.05	0.21
Village			
Village Population		184.16	86.08

Sample size = 7,919
Number of villages = 43

rate. Moreover, depending on the value of their parameters, it even turns out to be negative. While this is not impossible in theory, no plausible explanation of what causes the negative peer effect is provided in the original research.[12] As a related study, an attempt is made by Chandrasekhar and Lewis (2011), where the issue is attributed to a sampled network and consequent missing network links.

In this paper, however, a distinct approach is taken with the specification of a network. As mentioned in the previous section, they surveyed a total of twelve dimensions of social networks

12. Complementarity does not imply that the extra utility must come from using the same goods and services together at the same time. It may come from mitigating any uncertainties of unknown products. (Bursztyn et al. 2014)

Table 2.2: Networks description

Network	Description	Degree	Density	Spectral radius
Money	From/to whom they may borrow/lend money	4.244 (.804)	.029 (.016)	7.493 (1.793)
Advice	From/to whom they may find/give advice	3.263 (.627)	.024 (.015)	6.598 (1.250)
Kerosine	From/to whom they may borrow/lend kerosine or rice	3.740 (.901)	.025 (.012)	6.771 (2.072)
Medical	From whom they seek medical advice	2.947 (.735)	.020 (.012)	5.354 (1.752)
Non-relatives	Non-relative people with whom they socialize	4.360 (.649)	.031 (.018)	7.645 (1.364)
Relatives	Relatives in the village	2.172 (.479)	.015 (.009)	4.607 (.875)
Temple	With whom they go to temples together	.351 (.172)	.003 (.002)	1.287 (1.884)
Visit	Who they visit/Who visit them for free time	5.278 (1.247)	.035 (.016)	8.570 (2.340)
	Average	3.284	0.023	6.040
Union	Union of all networks	9.22 (1.69)	.06 (.04)	14.07 (3.2)

Standard errors in parentheses

and took a union of them for their study. By doing so, it is ignored that an individual's social network has multiple dimensions, and one communicates through a specific dimension with only specific topics. If this is the case, the misspecification problem of the original paper, in fact, is an overly dense network, not the opposite. In Table 2.2, it can be confirmed that the union network is significantly different from its individual dimensions in terms of graph statistics. The spectral radius (an eigenvalue of the largest absolute magnitude) of the union network is 15.08, which is significantly higher than 5.76 on average over the twelve networks. Also, density is four times higher than the individual average as 4% to 1%. To address this, the higher-order SAR model is used, and the exact dimensions that channel the peer effect on microfinance are identified.

Another departing point is the underlying assumption on the observed samples. The SAR model mainly differs from the diffusion model by assuming that the observed state is a stable equilibrium. Under this assumption, an equilibrium is an outcome of rational expectation formed by players based on complete information on the networks and the other players. To obtain a stable equilibrium, the magnitude of the peer effect must not dominate the marginal effect of each player's characteristic. On the other hand, the original paper views the diffusion process mainly as a function of graph statistics and individual characteristics as secondary. Regardless of individual characteristics, if one is not reached by an informer, they do not have any chance to make a decision. Consequently, the central role is played by the position of the leaders, who are the initial injection points on a graph. In addition to this, in the SIR (Susceptible-Infected-Recovered) diffusion model, one has only one opportunity to make a decision at the time of receiving information, and there is no further chance to change one's action in the future. In contrast, in the SAR model, it is assumed that the information is already common knowledge. Each player's concern is to expect their neighbors' possibilities of participation and decide theirs accordingly.

2.3 Model

For this chapter, I extend the GSAR model introduced in Chapter 1 to the binary outcome variables with rational expectation. Also, I adopt the higher-order specification for the weight matrices to incorporate the multiple dimensions of networks. Under the rational expectation assumption, players maximize their utility without directly observing the others' actions. To do so, they form rational expectation on the actions based on the network structure and their neighbors' attributes. Consequently, samples are viewed as equilibrium outcomes from a simultaneous-move game (Lee, Li, and Lin 2014; Yang and Lee 2017).

For this study, the peer effect is defined village-wise. The network structure of village v, G_v, is defined in a similar way to the continuous outcome variable case of the previous chapter. Each village's connectivity is represented by an undirected graph consisting of N_v villagers and the edges connecting them.

An assumption distinguished from the previous case is rational expectation. That is, the villagers are assumed to form rational expectation on the others' choices based on complete information.

Assumption 2.1. *(Complete information) The network structure, G_v, and the population characteristics, X_v, of village v are public information among the individuals in v.*

This means that an individual is fully aware of the links connected to her and the attributes of her peers connected through those links. Since the network structure is self-reported by the villagers, this assumption is automatically satisfied. Note that they do not have to know the networks of those with whom they are not connected.

Also, the margin of error from incorrect characteristics of one's neighborhoods becomes smaller as the walk on a network increases because it is more heavily discounted by the intensity parameter, which is less than one.

To control unobserved differences in each village, a group fixed effect variable τ_v, is included. Then, the model is represented as follows:

$$y_{i,v}^* = \sum_{q=1}^{Q} \rho_q W_{q,v}(\gamma) E[y_v | G_v, X_v] + x_{i,v}\beta + \mathbb{I}_{(v>1)}\tau_v - \varepsilon_{i,v}, \quad \varepsilon_{i,v} \sim \text{Logistic}(0,1),$$

where $Q = 8$, which is the number of networks, and $\mathbb{I}_{(v>1)}$ is an indicator variable that is one for every village other than the first one. The number of villages is $V = 43$. The vector of parameters of interest is an $(Q + 1 + K + V) \times 1$ vector, $\theta = (\rho', \gamma, \beta', \tau')'$. Note that $\rho = (\rho_1, \ldots, \rho_Q)$, which implies that the intensity of peer effect varies upon the network dimensions.

An observed outcome variable, $y_{i,v}$, is a binary choice variable:

$$y_{i,v} = \begin{cases} 1 \text{ if } y_{i,v}^* \geq 0 \\ 0 \text{ otherwise,} \end{cases}$$

which means that an individual will take up the microfinance program if her latent utility is greater than zero, and her choice variable is coded as one. On the other hand, if her latent utility is negative, she will not participate, and her choice variable is zero.

The rational expectation on the unobserved outcomes is formed upon complete information on the network structure and the individual characteristics.

Assumption 2.2. *(The best-response) The latent outcome variable, $y_{i,v}^*$, is a maximizer of the expected utility of each individual.*

Denote the choice probability, $P[y_{i,v} = 1|G_v, X_v]$, as $p_{i,v}$. Then the expected outcome, $\mathbb{E}[y_{i,v}|G_v, X_v]$, is equal to $p_{i,v}$ and computed as follows:

$$
\begin{aligned}
p_{i,v} &= P_{i,v}\Bigg(\sum_{q=1}^{Q} \rho_q w_{iq,v}(\gamma)\mathbb{E}(\mathbf{y_v}|G_v, X_v) + x_{i,v}\beta + \mathbb{I}_{(v>1)}\tau_v - \varepsilon_i > 0\Bigg) \\
&= F\Bigg(\sum_{q=1}^{Q} \rho_q w_{iq,v}(\gamma)\mathbf{p}_v + x_{i,v}\beta + \mathbb{I}_{(v>1)}\tau_v\Bigg),
\end{aligned}
$$

where $\mathbf{p}_v = (p_{1,v}, \ldots, p_{N_v,v})'$. For simplicity, denote $\eta_{i,v}(\theta, \mathbf{p}_v) = \sum_{q=1}^{Q} \rho_q w_{iq,v}(\gamma)\mathbf{p}_v + x_{i,v}\beta + \mathbb{I}_{(v>1)}\tau_v$.

The unique equilibrium of the game is a village-wise solution of the following nonlinear system of N_v equations, which is a function of θ. Denote such solution as $\mathbf{p}_v^*(\theta)$. Therefore, the solution for each i in village v, $p_{i,v}^*$, is:

$$
p_{i,v}^*(\theta) = F(\eta_{i,v}(\theta, \mathbf{p}_v^*(\theta))). \tag{2.1}
$$

This is a well-defined nonlinear operator defined on a Banach space (Yang and Lee 2017). Although the existence of equilibrium is guaranteed by the Brouwer fixed-point theorem, there may be multiple equilibria. To focus only on the case of unique equilibrium, the stability assumption that corresponds to Assumption 1.1 is required.

Assumption 2.3. *(Uniqueness)*

$$
\sum_{q=1}^{Q} |\rho_q| ||W_q(\gamma)|| \max_{i,v} f_{i,v}(\eta_{i,v}) < 1, \ \forall v
$$

Then, a unique equilibrium exists for each village by the contraction mapping theorem.

Identification of the parameters is shown by a similar way to the continuous outcome model. The following assumptions are parallel to those given in the previous chapter for continuous outcome variables.

Assumption 2.4. *(i) $F_i(\cdot)$ is a strictly increasing distribution function with unity variance that is known to the econometrician. (ii) Let $Z = [A_1\mathbf{p}, \ldots, A_Q\mathbf{p}, -D_1\mathbf{p}, \ldots, -D_Q\mathbf{p}, X, \mathbb{I}]$. $plim_{N\to\infty} N^{-1}Z'Z$ exists and is a finite positive definite matrix of full rank.*

Proposition 2.1. *Under Assumption 2.4, the set of parameters, $\theta = (\rho', \gamma, \beta', \tau')'$, is identified.*

Proof. See Appendix B. □

The assumption above suggests that each network is required to be linearly independent of each other to identify ρ_q. As previously mentioned, since the average overlapping rate of each network is 50%, this condition is satisfied.

The estimate of θ is obtained by maximizing the following log-likelihood function, $\mathcal{L}$, with respect to θ:

$$\mathcal{L}(\theta) = \sum_{v=1}^{V} \sum_{i=1}^{N_v} \big[y_{iv} \log F(\eta^*_{iv}) + (1 - y_{iv}) \log(1 - F(\eta^*_{iv})) \big]$$

where $\eta^*_{iv} = \eta^*_{iv}(\theta, \mathbf{p}^*_v(\theta)) = \sum_{q=1}^{Q} \rho_q w_{iq,v}(\gamma) \mathbf{p}^*_v(\theta) + x_{iv}\beta + \mathbb{I}_{(v>1)}\tau_v$.

2.4 Results

2.4.1 Point Estimates

As a benchmark, estimation without the peer effect is conducted with the ordinary logit model (Model 1).[13] In Table 2.3, it appears that the leader groups have a higher tendency to join the program. It is not unexpected, considering that they were initially encouraged to participate.

The additional characteristics that are not included in the original research (castes and religion) show significant correlations with the take-up rate. Because the other backward castes are the major group in most of the villages, they are set as the base for the caste variables. The underprivileged castes, such as scheduled caste and scheduled tribe, appear to be more likely to participate in the program. In contrast, the general caste, which is a relatively more favored group, shows a negative estimate. Also, Islam, a minor religion, has a significantly higher effect compared to the base religion variable, Hinduism. Among the income proxies, only having no in-house latrine and no access to electricity show a significantly positive correlation.

13. The estimates from this model are used as the initial point for the estimation of the rest of the models.

The standard SAR estimation is carried out using the union network to verify the existence of the peer effect (Model 2). Note that using the standard model implies the pure complementarity peer effect because the conformity parameter is fixed as zero. The result shows evidence of a significant peer effect.[14] Under the peer effect, it appears that the coefficient of the "leader group" variable is reduced by the largest margin. This suggests that the leaders were also affected by the complementary benefit from their peers. Therefore, without the peer effect, the direct effect of being designated as leaders is overestimated.

Table 2.3: Ordinary logit and SAR with the union network

	Model 1		Model 2	
	Coefficient	SE	Coefficient	SE
ρ				
Union			0.125***	0.018
β				
Constant	−1.376***	0.289	−1.864***	0.223
Roof material: tile	0.105	0.076	0.099	0.076
No. of rooms	−0.026	0.032	−0.053*	0.032
No. of beds	−0.018	0.034	−0.025	0.034
No latrine in house	0.353***	0.085	0.401***	0.085
House not owned	0.020	0.103	0.031	0.101
No access to electricity	0.190***	0.073	0.212***	0.072
Leader group	1.098***	0.167	0.572***	0.087
General caste	−0.312**	0.134	−0.254***	0.125
Minority castes	−0.045	0.235	0.228	0.239
Scheduled caste	0.493***	0.083	0.395***	0.076
Scheduled tribe	0.406***	0.146	0.372***	0.138
Islam	0.627***	0.087	0.935***	0.147
Multiplier			1.40	
Likelihood	−3451.06		−3431.58	
AIC	6928.12		6891.16	
BIC	7018.82		6988.84	

Significant at *10%, **5%, ***1%.

Table 2.4 reports a set of estimates with the higher-order SAR model: the pure complementarity model with $\gamma = 0$ (Model 3), the pure conformity model with $\gamma = 1$ (Model 4), and the

14. This result highlights the methodological difference between this paper and Banerjee et al. (2013). For a further discussion, see Appendix C.

GSAR model with $\gamma \in [0, 1]$ (Model 5). Model 3 reveals that only three of the networks, Relatives, Temple, and Visit, transmit significant peer effects. Note that the first four networks highly correlated with the income proxies turn out to be insignificant, which suggests that they are properly controlled by the covariates. In Model 4, no significant peer effect is found, and the coefficients of the household characteristics are similar to Model 1.

The main result of this paper is Model 5. In addition to the peer effects, the conformity parameter, γ, is estimated as 0.284 and significant at 95% confidence level. Although the number is closer to zero than one, this does not imply that complementarity is the dominant motive. The social multiplier computed from the estimates is 0.009, which means that the complementary benefit is not high enough to overwhelm the deviation cost for the villagers. This becomes more apparent with the marginal effects that will be presented later.

Despite the existence of a significant conformity motive, the pure conformity model does not capture such peer effect from the data. One of the possible explanations is the higher requirement of the pure complementarity model. Due to the smaller variance in the difference of choice probabilities, $p_j - p_i$, the pure conformity model needs more samples to achieve the same level of performance as the pure complementary model. The other reason is attributed to the misspecified model. It can be the case that not only do the villagers care more about conformity, but they also benefit from complementarity to some degree. For such instances, using the two pure models and conducting model selection will not provide an accurate picture of the peer effect.

As an extension of Model 5, the conformity parameters are separately estimated for each network dimension of interest (Model 6). Then, the weight matrix for q-th dimension is constructed as $W_{q,v}(\gamma_q) = A_{q,v} - \gamma_q D_{q,v}$. Because of the higher computational burden from estimating Q additional parameters, only the two networks that appear to be relevant from Model 5 are chosen for this extension, which are "Relatives" and "Visit."[15] The estimation result is shown in Table 2.5. The conformity parameter for the "Relatives" network appears to be 0.644, which is even higher than that of Model 5. On the other hand, the "Visit" network shows 0.091, which is almost

15. Temple company is excluded due to the insufficient number of samples for the increased estimation complexity.

Table 2.4: Higher-order SAR models

	Model 3: $\gamma = 0$		Model 4: $\gamma = 1$		Model 5: $\gamma \in [0, 1]$	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
ρ						
Money	−0.034	0.079	0.034	0.347	−0.058	0.115
Advice	−0.065	0.058	0.050	0.343	−0.109	0.080
Kerosine	0.042	0.068	0.047	0.345	0.047	0.097
Medical	0.001	0.079	0.035	0.394	0.101	0.113
Non-relatives	−0.006	0.063	0.000	0.254	−0.044	0.090
Relatives	0.141**	0.071	0.056	0.389	0.188*	0.103
Temple	0.371***	0.125	0.059	0.977	0.539***	0.198
Visit	0.234***	0.061	0.042	0.292	0.390***	0.108
γ					0.284**	0.123
β						
Constant	−1.783***	0.223	−1.381***	0.303	−1.671***	0.223
Roof material: tile	0.126*	0.074	0.114	0.088	0.147*	0.080
No. of rooms	−0.065**	0.031	−0.031	0.036	−0.086**	0.034
No. of beds	−0.019	0.033	−0.021	0.039	−0.019	0.036
No latrine in house	0.402***	0.083	0.374***	0.106	0.420***	0.092
House not owned	0.051	0.099	0.021	0.115	0.054	0.106
No access to electricity	0.216***	0.070	0.193**	0.086	0.226***	0.077
Leader group	0.768***	0.134	1.125***	0.192	0.696***	0.137
General caste	−0.217*	0.117	−0.328**	0.145	−0.226*	0.120
Minority castes	0.369	0.234	−0.070	0.269	0.458*	0.249
Scheduled caste	0.328***	0.070	0.495***	0.090	0.305***	0.071
Scheduled tribe	0.325**	0.128	0.413**	0.164	0.346**	0.135
Islam	0.606***	0.086	0.690***	0.139	0.657***	0.101
Multiplier	3.53		0.00		0.009	
Likelihood	−3414.81		−3446.18		−3409.37	
AIC	6871.62		6934.36		6862.74	
BIC	7018.14		7080.88		7016.23	

Significant at *10%, **5%, ***1%.

zero. This result confirms the well-known fact that diverse aspects of one's social network convey different types of peer effects through them.

From the perspective of model selection, Model 6 appears to be the best model according to both AIC and BIC.

Table 2.5: Network specific conformity parameters

	Model 6	
	Coefficient	SE
ρ		
Relatives	0.410**	0.179
Visit	0.250**	0.121
γ		
Relatives	0.644**	0.256
Visit	0.091	0.367
β		
Constant	−1.769***	0.229
Roof material: tile	0.137*	0.082
No. of rooms	−0.079**	0.034
No. of beds	−0.027	0.037
No latrine in house	0.413***	0.094
House not owned	0.058	0.107
No access to electricity	0.230***	0.078
Leader group	0.836***	0.143
General caste	−0.234*	0.121
Minority castes	0.308	0.257
Scheduled caste	0.335***	0.074
Scheduled tribe	0.340**	0.135
Islam	0.667***	0.108
Multiplier	0.007	
Likelihood	−3411.4	
AIC	6856.80	
BIC	6975.41	

Significant at *10%, **5%, ***1%.

2.4.2 Marginal Effects

The impact of the peer effect on the outcome variables is measured by the overall change caused by manipulating each explanatory variable. For the SAR models, it can be computed in two ways: the "naive" and the "sophisticated" approach. The naive marginal effect is obtained from $f(\eta_{iv}^*)\hat{\theta}$ for continuous variables("No. of rooms" and "No. of beds") and $P[y_{iv} = 1|x_{ik,v} = 1] - P[y_{iv} = 1|x_{ik,v} = 0]$ for dummy variables. This approach ignores the multiplier effect by ignoring the change in the probability caused by the change in explanatory variables. That is, it can be obtained by taking a derivative on the choice probability, $p_{iv}(\hat{\theta}) = F(\sum_{q=1}^{Q} \hat{\rho}_q W_{q,v}(\hat{\gamma})\mathbf{p}_v^*(\hat{\theta}) + x_{iv}\hat{\beta})$, with respect to $(w_{i1,v}(\hat{\gamma})\mathbf{p}_v^*(\hat{\theta}), \ldots, w_{iQ,v}(\hat{\gamma})\mathbf{p}_v^*(\hat{\theta}), x_{i1,v}, \ldots, x_{iK,v})$. In other words, the naive approach regards the choice probabilities of one's neighborhood as given constants and include them in a regression as independent variables.

Meanwhile, the sophisticated approach computes the marginal effect of an explanatory variable considering its effect on the equilibrium together to incorporate the multiplier effect. In this process, a change in i's characteristic changes not only i's own choice probability but the others' as well. Denote the aggregated weight matrix of v-th village as $W_v^{(A)} = \sum_{q=1}^{Q} \hat{\rho}_q W_{q,v}(\hat{\gamma})$ and its i-th row as $w_{i,v}^{(A)}$. The direct and indirect effects for the continuous variables ("No. of rooms" and "No. of beds" variables) are

$$\text{Direct effect}_{i,v} = \frac{\partial p_{i,v}^*}{\partial x_{ik,v}} = f(w_{i,v}^{(A)}\mathbf{p}_v^* + x_{i,v}\hat{\beta})\left(\hat{\beta}_k + w_{i,v}^{(A)}\frac{\partial M_v}{\partial x_{ik,v}}\right)$$
$$\text{Indirect effect}_{i,v} = \frac{1}{N_v - 1}\sum_{j\neq i}\frac{\partial p_{j,v}^*}{\partial x_{ik,v}} = \frac{1}{N_v - 1}\sum_{j\neq i} f(w_{i,v}^{(A)}\mathbf{p}_v^* + x_{i,v}\hat{\beta})\left(w_{j,v}^{(A)}\frac{\partial M_v}{\partial x_{ik,v}}\right),$$

where

$$\frac{\partial M_v}{\partial x_{ik,v}} = \left(I_{N_v} - f_{diag}(W_v^{(A)}\mathbf{p}_v^* + X_v\hat{\beta})W_v^{(A)}\right)^{-1} f_{diag}(W_v^{(A)}\mathbf{p}_v^* + X_v\hat{\beta})\left(\frac{\partial X_v}{\partial x_{ik,v}}\right),$$

and $\frac{\partial X_v}{\partial x_{ik,v}}$ is the $N \times K$ matrix of which (i,k) entry is 1 and 0 elsewhere. For the dummy variables, the marginal effect of k-th covariate is computed as $P[y_{i,v} = 1 | x_{ik,v} = 1, M_v(x_{ik,v=1})] - P[y_{i,v} = 1 | x_{ik,v} = 0, M_v(x_{ik,v=0})]$. In the following tables, I report both direct and indirect effect averaged over all population in every village.[16]

The results for all of the specifications are shown in Table 2.6 and Table 2.7, in percentage points. In Model 1, the magnitudes are smaller than in any other model due to the lack of the multiplier effect. In other words, any changes in the explanatory variables of the benchmark model do not bring any additional impacts through the equilibrium. The main focus for the rest of the models is the difference between the naive and the sophisticated effects. In Model 2 and 3, the direct effects are larger than the naive effects in general due to the complementarity peer effect. This

16. For the treatment of the marginal effects of the SAR model, see Chomsisengphet, Kiefer, and Liu (2018).

Table 2.6: Marginal effects: Model 1-3

	Model 1	Model 2			Model 3		
		Naive	Direct	Indirect	Naive	Direct	Indirect
Roof material: tile	1.43	1.34	1.34	0.002	1.69	1.72	0.004
No. of rooms	−0.35	−0.71	−0.71	−0.001	−0.86	−0.88	−0.002
No. of beds	−0.14	−0.35	−0.35	−0.001	−0.26	−0.27	−0.001
No latrine in houses	2.72	5.13	5.16	0.008	5.09	5.17	0.012
House not owned	0.16	0.42	0.42	0.001	0.69	0.70	0.002
No access to electricity	1.49	2.91	2.92	0.005	2.94	2.99	0.007
Leader group	4.83	8.68	8.73	0.013	9.14	9.31	0.021
General caste	−2.40	−3.24	−3.26	−0.005	−2.77	−2.81	−0.007
Minority castes	−0.35	3.27	3.29	0.005	5.44	5.54	0.013
Scheduled caste	3.89	5.59	5.61	0.009	4.55	4.63	0.011
Scheduled tribe	3.13	5.50	5.52	0.009	4.70	4.79	0.011
Islam	7.89	15.64	15.73	0.024	12.30	12.54	0.028
Union		1.69					
Money					−0.45		
Advice					−0.87		
Kerosine					0.56		
Medical					0.01		
Non-relatives					−0.09		
Relatives					1.89		
Temple					4.96		
Visit					3.13		

The numbers are in percentage points.

result confirms the positive multiplier effect computed from the parameter estimates. Policymakers might be interested in the marginal effect of the leaders. With the complementarity peer effect, their direct effect tends to be higher than the naive effect, which suggests the positive feedback caused by the positive multiplier. Meanwhile, in the rest of the models where conformity appears to be dominant, the direct effects are lower than the naive effects. It is as expected because the individuals will move toward the social norm, and such a motive will attenuate the consequences of any increase in their attributes. It is noteworthy that the same pattern is observed from Model 5 as well. Despite the estimated conformity parameter being lower than one, this is an evidence of a significant conformity motive.

Table 2.7: Marginal effects: Model 4-6

	Model 4			Model 5			Model 6		
	Naive	Direct	Indirect	Naive	Direct	Indirect	Naive	Direct	Indirect
Roof material: tile	1.83	1.55	0.002	1.99	1.85	0.007	1.86	1.71	0.005
No. of rooms	−0.65	−0.55	−0.001	−1.14	−1.06	−0.004	−1.07	−0.99	−0.003
No. of beds	−0.28	−0.23	0.000	−0.25	−0.23	−0.001	−0.36	−0.33	−0.001
No latrine in houses	5.52	4.72	0.005	5.28	4.93	0.019	5.30	4.89	0.016
House not owned	0.28	0.24	0.000	0.73	0.67	0.002	0.80	0.74	0.002
No access to electricity	2.66	2.24	0.002	3.08	2.85	0.011	3.17	2.90	0.009
Leader group	14.31	11.66	0.012	9.96	9.14	0.033	10.30	9.34	0.029
General caste	−2.73	−4.09	−0.004	−2.97	−2.77	−0.010	−3.01	−2.79	−0.009
Minority castes	−2.01	−1.72	−0.002	6.80	6.24	0.023	4.53	4.12	0.013
Scheduled caste	7.11	5.94	0.006	4.22	3.91	0.014	4.72	4.32	0.014
Scheduled tribe	6.55	5.39	0.006	5.12	4.72	0.017	5.00	4.55	0.014
Islam	21.49	17.05	0.018	10.92	9.97	0.036	13.73	12.35	0.038
Money	0.50			−0.79					
Advice	1.62			−1.47					
Kerosine	1.19			0.65					
Medical	0.96			1.33					
Non-relatives	−1.26			−0.56					
Relatives	2.46			2.43			5.56		
Temple	3.43			7.06					
Visit	0.41			5.15			3.39		

The numbers are in percentage points.

As mentioned earlier, this implies that the pure conformity model with $\gamma = 1$ is not necessary to demonstrate the conformity motive. Due to its strong restriction on γ, using the pure conformity model can lead a researcher to the wrong conclusion that the conformity motive does not exist at all. Instead, a social multiplier less than one and a sufficiently high conformity parameter will be more relevant indicators of the conformity motive.

An interesting observation is that the indirect effects of Model 5 and 6 are higher than that of Model 3. Again, consider the leader group as an example. Even though those who are appointed as leaders might be less willing to join the program compared to those under complementarity, the other people's take-up rate can be increased even more by the same conformity motive. This result shows that a strong conformity motive can be utilized to raise the participation rate of the general population other than the target group.

2.4.3 Effective Networks

The effective networks are constructed using only the ones found to convey significant peer effects. As a result, a total of three networks, Relatives household, Temple company, and Visit, are selected, and an aggregated effective network is constructed as a union of them. Its graph statistics are presented along with the previously shown statistics table.

Table 2.8: Comparison with the relevant networks

	Aggregated		Union		Selected		Individual	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Degree	6.36	1.27	9.22	1.69	2.36	0.16	3.00	0.24
Density	0.04	0.02	0.06	0.04	0.02	0.01	0.03	0.01
Spectral Radius	9.78	2.35	14.07	3.20	4.12	0.54	5.22	0.66

Note: the selected networks are Relatives, Temple and Visit. The statistics are averaged over the three networks.

The result shows that the union network is noticeably overspecified compared to the aggregated network. The outcome density is 4%, which is lower than that of the union network, 6%. More importantly, the spectral radius, which is a more accurate measure of a network's connectivity, is 9.78. Compared to 14.07 of the union network, this is a significantly lower number. Network researchers are often concerned about network misspecification, especially with missing links. However, this result suggests that the other type of misspecification, or overspecification, can also be a problem.

2.5 Counterfactual Analysis

In the original case of BCDJ, BSS has designated a group of individuals to introduce the microfinance program and boost its take-up rate. These people, who are referred to as "leaders," were chosen based on their socioeconomic status that was believed to give the leaders more influence over the other villagers. In this section, I conduct a series of counterfactual analyses where such leaders are selected by alternative criteria.

The leaders, ceteris paribus, are more willing to take up the program since they were the first of the villagers whom BSS initially invited to the meeting. According to the marginal effect analysis in Section 3, it increases an individual's own latent utility by 9.96% and her peers' by 0.033% on average. The basic idea of utilizing peer effect is using these changes as leverage to boost the aggregate outcome of entire villages.

However, not every leader is equal. Depending on their location in a social network, the overall impact on the aggregate outcome differs. In the framework of the stable equilibrium, not only one's position on a network but the characteristics of one's peers matter. For example, suppose that there are two candidates for a leader whose centrality measures are identical. One of them, however, has no peers who are potentially willing to join the program, and the other has a lot. The centrality measures based solely on graph statistics do not capture the entire impact on the aggregate outcome when individuals in a network are heterogeneous in their attributes. Therefore, the main issue is identifying those who will increase it by the largest margin.

The analysis begins with a baseline case where no leader exists. It is presented in Table 2.9 as "No injection point." This is computed with the estimates previously obtained by alternating only the leader status variable. Without leaders, it is expected that 1,262 villagers will take up the program.

A potential improvement made by introducing a group of leaders is rooted in three different sources: the number of leaders, their relative position on a network and the characteristics of their peers. In this analysis, I keep the number of leaders the same as that of the original case of BSS so that the effect from the first factor remains constant. That exact effect is presented as "Random selection" in the same table. It appears that simply adding some leaders increases the overall participation rate by 10.58%. Compared to the BSS leaders, it is only 1.55% lower. This suggests that there is no additional benefit from choosing injection points based on their socioeconomic status as BSS did.

Instead of choosing leaders based on their status, my alternative approach is adopted from Ballester, Calvó-Armengol, and Zenou (2006). In their paper, each individual is removed one by one, and how much each removal reduces the aggregate outcome is evaluated. The one who reduces the most is called a "key player," and the rest of the individuals are ranked accordingly.[17] Following their procedure, I compute counterfactual outcomes using the estimates obtained by the GSAR model and evaluate each individual's potential impact. Then, I assign the top players as alternative leaders, with the same number as BSS.

The result is summarized in Table 2.9. First, the key player approach enhances the aggregate participation rate by 21.73%. Also, compared to the BSS leaders, it is 9.48% higher. This additional improvement can be explained as a consequence of considering each leader's position and their peers' attributes simultaneously.

Compared to the counterfactuals derived from the graph centrality measures, the key players perform better. This shows that the centrality measures alone are not enough to identify the central players when the individuals are heterogeneous. Still, all of them result in higher outcomes than the BSS leaders do.

The gap between these outcomes suggests multiple implications. First, the leader group by BSS was not the best choice to catalyze peer effect, whether its motive is complementarity or conformity. For instance, the BSS leaders are more privileged than most of the potential users of microfinance, which means that they have no higher incentive to take up the program.[18] As a result, in spite of the initial effort of BSS, the take-up rate of the leaders is 0.24, which is not much higher than that of non-leaders, 0.17. This is more pronounced when it is compared to the counterfactual leaders, whose take-up rate is 0.30.[19]

In contrast, the counterfactual leaders are among the more underprivileged, which means they are more likely to be actual consumers of microfinance.[20] Moreover, their positions in networks are more favorable. For example, their average degree is 11.14, which is higher than that of the BSS leaders, 8.36.

In summary, the leaders chosen by BSS were not the best group to utilize peer effect. Indeed,

17. Note that the links are unchanged by the removal. This can be justified when the measurement is based on a short-term change.
18. For more detailed descriptive statistics on the BBS leaders, see Table B.3
19. This is from the original data, which means that it could have been higher if they were assigned as leaders.
20. See Table B.4 for descriptive statistics of the counterfactual leaders.

only 12.26% of the counterfactual leaders overlap with the BSS leaders.

Table 2.9: Counterfactual aggregate take-up rate

	Mean	S.D.	Min.	Max.	Med.	Percentages over No injection	BSS
Benchmarks							
No injection point	1261.86						
Random selection*	1395.31	2.37	1386.12	1402.61	1395.28	10.58	-1.55
Chosen by BSS	1403.00					11.19	-
Counterfactuals							
Key Player	1536.00					21.73	9.48
Eigenvalue	1470.54					16.54	4.81
Katz-Bonacich	1461.37					15.81	4.16
Degree	1468.69					16.39	4.68

*: 1,000 replications with randomly chosen leaders

A theoretical limitation is the fact that the role of the leaders as informers is overlooked. In BCDJ, the authors focus on the villagers who did not take up the program but only passed through the information. If the BSS leaders were more capable of spreading the information, although not joining the program themselves, their designation as leaders may be justified. However, their average degree on the network is 8.36, which is not significantly higher than that of non-leaders, 6.36. It is more pronounced when compared to the counterfactual leaders, whose average degree is 11.14.[21]

Therefore, if those who are more likely to join the program themselves and benefit more from it were selected as leaders, the overall peer effect would have been more potent.

2.6 Conclusion

In this chapter, I revisit the case of Banerjee et al. (2013) and apply the GSAR model to reveal the peer effect among individuals. The GSAR model is modified to incorporate binary outcome variables and multi-dimensional networks. By the GSAR estimation, I find a significant conformity peer effect, contrasting the original study. This result is derived from the stable equilibrium assump-

21. See Appendix B for more details.

tion, which is a fundamental difference from the diffusion model. Instead of considering diffusion patterns over time, the GSAR model assumes that every individual has reached an equilibrium.

Additionally, I demonstrate that only a few dimensions of one's social network channel peer effect and the underlying motives of peer effect can vary depending on each dimension. In the case of the microfinance program, they appear to be under the influence of their relatives, those with whom they go to temples and those with whom they spend free time together. Also, The driving motivations are conformity for the relatives network and complementarity for the free time network.

For policy implications, I conduct a counterfactual analysis and find that the BSS leaders were not the most effective choices. The result shows that the leaders based on the "key player" approach by Ballester, Calvó-Armengol, and Zenou (2006) perform better than the leaders appointed by BSS by 10%.

One of the limitations of this study is that the role of the leaders as information spreaders is overlooked. This is due to the assumption of stable equilibrium, which is a characteristic of the SAR model that distinguishes it from the diffusion models. The idea of measuring the motives of peer effect, however, can be extended to the information diffusion models. I leave this as a topic for future research.

Chapter 3

Estimation of Networks with Binary Outcome Variables and Panel Data

A major obstacle to studying social networks is data availability. Random sampling, which ensures the consistency of an estimator for typical cases, does not guarantee the same for the SAR models. The reason is that the SAR model has the outcome variables of one's peers as explanatory variables to describe the interdependency of individuals, which brings endogeneity to a regression. Thus, researchers who want to study peer effect must acquire more densely sampled data, ideally the entire population.

If such data is available, a researcher can choose from a variety of SAR models depending on contexts. The core issue of the SAR model is addressing the endogeneity. Similar to the typical cases, using instrumental variables is one of the solutions. Kelejian and Prucha (1998b) proposes using the peers positioned more than two walks away on a network as an IV. Their methodology is refined by Lung-fei Lee (2003) for higher efficiency. Along the same line, Bramoullé, Djebbari, and Fortin (2009) shows that the reflection problem of Manski (1993) is addressed when network data is more specific than the linear-in-means model.

For discrete choice models, Lee, Li, and Lin (2014) provides the fundamental relying upon rational expectation. In this framework, individuals in a group form rational expectation on others' actions based on the information of networks and the characteristics of their peers. The same technique is extended to incomplete information, where individuals are not fully aware of their peers' characteristics (Yang and Lee 2017; Yang, Lee, and Qu 2018).

In a case where such data are not available, the option is much limited. It is possible to construct an artificial network rooted in homophily, but this heavily depends on the homophily assumption itself. Another alternative is modeling the formation process of a network. In their working paper, Boucher and Houndetoungan (2020) show that if a possibility of establishing a link between people follows a distribution function, the whole structure can be estimated. In a similar way, Lin, Tang, and Yu (2020) exploits a part of regressors that are relevant to the network. Although this does not directly obtain the structure, the peer effect is ultimately estimated.

The other line of the literature that I follow in this chapter is estimating the entire network from observed data. One of the examples is Lam and Souza (2019). They estimate an unknown network based on preliminary knowledge using the shrinkage estimator. The preliminary knowledge is represented by multiple weight matrices that serve as foundations of the network of interest. Taking a step further, de Paula, Rasul, and Souza (forthcoming) imposes only sparsity on the network structure. Sparsity is a common feature in many social networks: individuals in a network are connected with only a small fraction of the entire population. By implementing the adaptive elastic net GMM estimator, a type of the shrinkage estimator, they estimate the individual links in a network using panel data. They also show that the data requirement is not prohibitively high. However, their model is based on continuous outcome variables and, therefore, cannot be directly applied to binary choice variables.

In this chapter, I extend the methodology of de Paula, Rasul, and Souza (forthcoming) in two directions. First, I extend the estimation technique to the SAR model of binary outcome variables. Second, I implement the binary particle swarm optimizer for the estimation of network links.

The binary outcome SAR model is based on Lee, Li, and Lin (2014). As outlined in the previous chapter, it draws upon the rational expectation formed by individuals on the behavior of their peers. This property allows a researcher to incorporate the case of binary outcome into the estimation methodology of de Paula, Rasul, and Souza (forthcoming).

The binary particle swarm optimizer (BPSO) is a variant of the particle swarm optimizer, implemented particularly for discrete search spaces (Kennedy and Eberhart 1997; Vieira et al. 2013; Li, Xue, and Zhang 2021). In cases where weights are not the central matter, it is beneficial to reduce a model to an unweighted graph. This is even more useful for the binary outcome model, where the variation on the outcome variable is smaller, so it conveys less information.

Another advantage of the BPSO is that it can utilize the particles more efficiently by exploring a narrower search space. Instead of determining a weight between zero and one, the BPSO only decides if a link exists or not. As a result, one can use a smaller number of particles and significantly reduce computation time.

The BPSO itself does not favor a sparse model by nature. However, sparsity is a necessary property for estimation, as the number of links is higher than that of samples in most cases. To impose it, a shrinkage estimator is combined with the optimizer. The Lasso estimator, proposed by Tibshirani (1996), shrinks each coefficient to obtain exact zero estimates. Although this is not an essential feature of the BPSO, the penalized estimation can be combined with it so that sparser models are favored.

By combining these two lines of methodologies, I propose an algorithm where particles of the BPSO explore the space of weight matrices. In other words, each of them will test a different specification of a weight matrix with binary elements. With this given network, they evaluate the likelihood function based on the Lasso estimator and compare the value with the other particles. The particles will eventually head to the global optimum, and consequently, a true specification of the network can be captured.

With a set of Monte Carlo simulations, I show that individual network links can be estimated with high accuracy using the proposed method. For example, for a circular network where individuals are connected with two peers besides, the entire network is captured with 100% of accuracy with five individuals and a hundred time periods. For a random network with 15 individuals and 100 periods, the accuracy is 97%.

The advantage of using the BPSO is the lower number of particles required to achieve a certain level of accuracy. In the simulation with 10 individuals and 25 periods, there is no difference in performance between using 10,000 and 30,000 particles. Considering that the size of the search space is 2^{45}, it demonstrates that the required computation power is not prohibitively high for the proposed algorithm.

The rest of this chapter consists of four sections. In Section 3.1, the binary choice model and main assumptions are introduced. In Section 3.2, I will overview my proposed estimation procedure and the link selection estimator used. Next, in Section 3.3, preliminary results from Monte Carlo simulations will be presented. Section 3.4 finishes the last chapter by providing a concluding remark.

3.1 Model

I consider N individuals in a non-stochastic network observed over a total of T time periods. The network at time t, unknown to an econometrician, is an undirected and unweighted graph consisting of a set of vertices, V, and edges, E. Following the typical setting of the SAR literature, self-loops are not allowed. I denote such a network structure as $G_{N_t} = G_{N_t}(V, E)$.[22]

Assumption 3.1. *The network structure, G_{N_t}, and the total population, N_t, are invariant over time. That is, $G_{N_t} = G$.*

From the graph, a corresponding adjacency matrix is formed, defined as follows:

Definition 3.1.

$$A = \begin{cases} [a_{ij}], & \textit{if } i \neq j \\ 0, & \textit{otherwise} \end{cases} .$$

The binary elements of the adjacency matrix, a_{ij}, represent links between individuals with $a_{ij} = 1$ if a link exists and $a_{ij} = 0$ if not.

Based on the adjacency matrix above, a weight matrix is derived for the SAR model. For example, if a peer effect is modeled as an average influence from one's peers, the weight matrix will be a row-sum normalized adjacency matrix. Denote it as W, its i-th row as w_i and i, j element as w_{ij}.

22. The graph can be relaxed to be directed if large enough data are provided.

Then, the latent utility model with peer effect is:

$$y^*_{i,t} = \lambda w_i E(d_t|G, X_t) + w_i x_{i,t}\delta + x_{i,t}\beta + \varepsilon_{i,t}, \tag{3.1}$$

where d_t is a binary choice vector equal to $(d_{1,t}, \ldots, d_{N,t})'$. Each individual, i, chooses an action, $d_{i,t}$, in a choice set, $\{0, 1\}$, at time t according to her latent utility, $y^*_{i,t}$. The latent utility denotes the relative utility derived from her action, and she will choose one only if the utility is positive. Thus, the observed choice variable of i at t is,

$$d_{i,t} = \begin{cases} 1 \text{ if } y^*_{i,t} \geq 0 \\ 0 \text{ otherwise.} \end{cases}$$

The error term, $\varepsilon_{i,t}$, follows an i.i.d. logistic distribution with the scale factors of $(0, 1)$.

In regression (3.1), the two peer effect parameters, λ and δ, measure the "endogenous effect" and the "contextual effect," respectively (Manski 1993).

This model is built upon the theoretical foundation of Lee, Li, and Lin (2014). That is, I assume that an individual does not directly observe the others' actions or latent utility. Instead, they take actions simultaneously or without observing the others' actions. Thus, they must form expectations on the decision vector of the entire population, d_t.

Also, I assume that their rational expectation is independent between each period.

Assumption 3.2. $E[d_t|d_{t-1}, d_{t-2}, \ldots, d_1, G, X_t, X_{t-1}, \ldots, X_1] = E[d_t|G, X_t]$

The rational expectation on the unobserved outcomes is formed upon complete information on the network structure and the individual characteristics. Note that the network structure is unknown only to an econometrician.

Assumption 3.3. *(Complete information) The network structure, G, and the characteristics of one's peers, X_t, are public information for all t.*

With rational expectation, individuals choose their actions to maximize their latent utility.

Assumption 3.4. *(Best response) The choice variable of each player i at time t, $d^*_{i,t}$, is a maximizer of her latent utility at t.*

Then, this is a simultaneous-move game where every individual plays her best response. The following assumption ensures a unique equilibrium of the game.

Assumption 3.5. *(Uniqueness)*

$$|\lambda|||W||\max_i f_i(\eta_i) < 1$$

When a weight matrix is row-sum normalized, its spectral radius is unity. Therefore, the domain of the endogenous effect coefficient is $|\lambda| < 4$ for this model, where the error term follows the logistic distribution.

The regression (3.1) can be written in a matrix form:

$$y_t^* = \lambda W E(d_t|G, X_t) + W X_t \delta + x_t \beta + \varepsilon_t, \tag{3.2}$$

where, $y_t^* = (y_{1,t}^*, \ldots, y_{N,t}^*)'$. In this model, W is unknown to an econometrician, and its elements are subjects of estimation along with the SAR coefficients.

Since the graph is symmetric, only the upper (or lower) triangle needs to be estimated. Denote the vector of the upper-triangle elements as $\omega = (w_{1,2}, \ldots, w_{1,N}, w_{2,3}, \ldots, w_{2,N}, \ldots, w_{N-1,N})$ with a dimension of $(N^2 - N)/2$.

The set of SAR coefficients is denoted as $\theta = (\lambda, \delta', \beta')$, with a dimension of $1 + 2K$.

Define $p_{i,t}(\omega, \theta)$ as the probability of choosing 1 for i at t with given ω and θ. Since the choice set is $\{0, 1\}$, it is equivalent to $P(d_{i,t} = 1) = E(d_{i,t}|G, X_t)$. Denote an $N \times 1$ vector of $p_{i,t}$ as p_t. By the cdf of the logistic distribution, the equilibrium fixed point is obtained,

$$p_{i,t}^*(\omega, \theta) = P(\lambda w_i E(d_t|G, X_t) + w_i x_{i,t} \delta + x_{i,t} \beta + \varepsilon_{i,t} > 0) \tag{3.3}$$

$$= P(\lambda w_i p_t^* + w_i x_{i,t} \delta + x_{i,t} \beta + \varepsilon_{i,t} > 0) \tag{3.4}$$

$$= 1 - F(-\lambda w_i p_t^* - w_i x_{i,t} \delta - x_{i,t} \beta) \tag{3.5}$$

$$= \frac{1}{\exp(-\lambda w_i p_t^* - w_i x_{i,t} \delta - x_{i,t} \beta))}, \tag{3.6}$$

where w_i is i-th row of W. Under Assumption 3.5, a unique solution exists for (3.6) by the contraction mapping theorem.

3.2 Estimation

The elements of the weight matrix are estimated along with the regression coefficients by minimizing the following likelihood function built upon the fixed point discussed in the previous section:

$$L(\omega, \theta, p^*(\omega, \theta)) = \sum_{t=1}^{T} \sum_{i=1}^{N} \left[y_{i,t} \log(p^*_{i,t}(\omega, \theta)) + (1 - y_{i,t})(\log(1 - p^*_{i,t}(\omega, \theta)) \right].$$

I propose an algorithm for the minimization problem that consists of an inner and an outer likelihood function.

For each weight matrix, corresponding estimates of the coefficients can be found by the nested fixed point algorithm of Rust (1987, 1988) or the nested pseudo-likelihood algorithm (NPL) of Aguirregabiria and Mira (2002, 2007)[23] That is, the estimated vector of coefficients is a function of a weight matrix. Denote it as $\hat{\theta}(\omega)$. Then, the likelihood function can be concentrated to a function of ω.

$$L_c(\omega) = \sum_{t=1}^{T} \sum_{i=1}^{N} \left[y_{i,t} \log(p^*_{i,t}(\omega, \hat{\theta}(\omega))) + (1 - y_{i,t})(\log(1 - p^*_{i,t}(\omega, \hat{\theta}(\omega))) \right].$$

Despite the difficulty of showing the convexity of the concentrated function, the binary particle swarm optimization (BPSO) can handle such maximization problems. In the optimization process, each particle explores a search space, which is every possible combination of a binary vector, ω. Then, it evaluates the value of the likelihood function, $L_c(\omega)$, and compares it with the values the other particles have found. After the communication, the "swarm" will move toward the latest global optimum and evaluate the likelihood for each ω it faces on the way.

Considering the high-dimensional nature of the weight matrix, the number of swarms must be increased accordingly as the number of vertices in a network increases. Specifically, the size of the search space is $2^{(N^2-N)/2}$, which increases exponentially as a network grows. The number of particles, however, can be increased only linearly due to the limitation of computational power. I

23. With modern computers, the computational times of the two algorithms do not differ dramatically. However, since thousands of estimations must be performed for the algorithm proposed in this chapter, even a small saving in computational time is valuable. For this reason, I adopt the NPL for my algorithm.

show that, despite such hindrance, the BPSO handles the high-dimensional problem relatively well with its randomized search algorithm.

A necessary assumption is that the true value of ω uniquely maximizes the likelihood function.

Assumption 3.6. *$L_c(\omega) < L_c(\omega_0)$ for any $\omega \neq \omega_0$, where ω_0 is the true elements of a network.*

Although the sparsity of a network is an unnecessary property for the SAR model, it is required for the estimation of network links because of its high dimensionality. As the BPSO itself does not inherently favor sparse models, I combine the Lasso estimator in a similar way to de Paula, Rasul, and Souza (forthcoming).

Then, the estimate of the weight matrix, $\hat{\omega}$, is obtained by maximizing the following penalized likelihood function:

$$L_p(\omega) = L_c(\omega) - \phi||\omega||,$$

where ϕ is a scalar penalty term and $||\omega||$ is the number of nonzero estimates. Then, the coefficients are obtained correspondingly by $\hat{\theta}(\hat{\omega})$.

I choose the penalty term, ϕ_1, based on the BIC criterion:

$$BIC = -2L(\theta) + \frac{\log T}{T}||\omega||,$$

Note that only the number of links is the subject of penalization. This is because the contextual and direct effect coefficients, δ and β, are not high-dimensional in general. I evaluate the BIC for each penalty term and choose the one that yields the lowest value, following de Paula, Rasul, and Souza (forthcoming).[24] [25]

The whole process is summarized in Algorithm 1. The goal of the algorithm is to find the weight matrix that leads to the global maximum of the penalized likelihood function by moving particles, denoted as ω^k.

24. As pointed out in de Paula, Rasul, and Souza (forthcoming), using the BIC criterion in this way has no theoretical foundation. Instead, other model selection methods, such as cross-validation, are generally used.

25. I evaluated with six different penalties, $\{.01, .025, .05, .1, .15, .3\}$.

Algorithm 1 Link detection with the BPSO

```
Initialize $L^*_{p,n-1}$, $L^*_{p,n}$ and $\omega^k$
$n \leftarrow 1$
while $L^*_{p,n-1}(W,\theta) \neq L^*_{p,n}(W,\theta)$ do
    $L^*_{p,n-1} \leftarrow L^*_{p,n}$
    Move $\omega^k$ according to the PSO parameters
    Construct $W^k$ from each $\omega^k$
    Obtain $\hat{\theta}^k(W^k)$
    Compute $L(W^k, \hat{\theta}_p(W^k))$ for each $\omega^k$
    $W_g \leftarrow \arg\max_{W^k} L(W^k, \hat{\theta}_p(W^k))$
    $L^*_n \leftarrow L(W_g, \hat{\theta}(W_g))$
end while
$\hat{W} \leftarrow W_g$ and $\hat{\theta} \leftarrow \theta(W_g)$
```

3.3 Monte Carlo Simulations

This section presents the estimator's performance with Monte Carlo simulations. I test different networks and sample sizes that differ in the number of time periods and individuals.

For the simulations, I emulate social networks with two stylized graphs.

(1) Circular network: individuals are placed on a ring where each of them is linked with two people beside her. The degree is two for everyone, and, consequently, density decreases as the size of a network increases.

(2) Random network: a link is generated by a uniform distribution. In this section, the probability is set so that the density of a network is about 5%. Thus, the average degree increases with the size of a network.

For each network, the sums of the rows are normalized to one. In the case of the random network, I assign a link randomly if any empty row is generated.

With the simulated networks, panel data are generated by,

$$y_t^* = \lambda W E(y_t|X_t, G_t) + W X_t \delta + X_t \beta + \varepsilon_t.$$

Each covariate, X_t, is generated by the normal distribution with a mean of zero and a standard deviation of 2. The idiosyncratic error, ε_t, follows the logistic distribution with $(0, 2)$.

The number of particles is an integral part of the PSO because it directly affects its computational time and performance. I use particles varying from 100 to 3,000 according to the size of the networks. These are relatively smaller than the number of potential links, $2^{(N^2-N)/2}$, which means that evaluating all possible combinations by brute force is not feasible.

3.3.1 Results

First, I present three performance measures for assessing simulation results. Successful detection of existing links is separated into four cases: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The overall measure of accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}.$$

It approaches unity as the portion of correctly estimated links increases.

The first measure, however, is disproportionately influenced by zero estimates due to sparsity. One may be interested more in whether existing links are correctly found. In this case, the following measure can be used:

$$\text{Sensitivity} = \frac{TP}{TP + FN}.$$

It represents the portion of true estimates among the links estimated as nonzero.

The portion of correctly detected nonzero links can be over-represented if there are many noises, e.g., too many links are estimated as nonzero. The following measure is useful for capturing such cases:

$$\text{Specificity} = \frac{TN}{FP + TN}.$$

Results from the simulations with each network structure are summarized in Table 3.1. For each scenario, a thousand replications are performed. The true values of the coefficients are $\lambda_0 = 1$, $\delta_{10} = 2$ and $\beta_{10} = 3$. To assess the estimation performance of the coefficients, I compare the estimates to the references that are obtained by the standard SAR model with true networks. Since the estimates with estimated networks cannot be better than the reference estimates, any inaccuracy caused by the noise on the network should be assessed by the discrepancy from the reference, not the true values. If the reference is not close to the true value enough, it is due to the lack of large enough samples.

It appears that the performance of link estimation improves as the time period increases. For each N, the overall accuracy approaches 100%. The existing links are captured by 100% for most cases (sensitivity), as the circular network is very sparse.

As shown in Table 3.2, the result from the random networks shows a similar pattern but with a lower accuracy due to its higher density. It also provides a useful implication for smaller samples.

Table 3.1: Circular network

	$N = 5$			$N = 10$			$N = 15$		
	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
Link Detection									
Accuracy	0.968	0.997	1.000	0.935	0.987	0.998	0.901	0.965	0.987
Sensitivity	0.984	1.000	1.000	0.982	1.000	1.000	0.935	0.990	1.000
Specificity	0.957	0.995	1.000	0.923	0.983	0.997	0.895	0.961	0.986
Coefficients									
$\lambda_0 = 1$	1.320	1.094	1.046	1.091	1.005	0.992	0.861	0.859	0.912
	(2.407)	(0.754)	(0.436)	(0.984)	(0.459)	(0.299)	(0.601)	(0.359)	(0.240)
$\delta_{10} = 2$	3.071	2.248	2.091	2.583	2.102	2.026	1.929	1.814	1.866
	(4.387)	(0.640)	(0.317)	(0.741)	(0.337)	(0.209)	(0.365)	(0.273)	(0.179)
$\beta_{10} = 3$	4.519	3.350	3.127	3.529	3.106	3.030	2.507	2.601	2.762
	(6.492)	(0.914)	(0.439)	(0.962)	(0.472)	(0.293)	(0.459)	(0.391)	(0.265)
Reference									
$\lambda_0 = 1$	1.415	1.097	1.047	1.103	1.029	1.002	1.089	1.026	1.011
$\delta_{10} = 2$	2.703	2.236	2.091	2.212	2.092	2.040	2.098	2.068	2.028
$\beta_{10} = 3$	4.052	3.342	3.127	3.317	3.138	3.055	3.168	3.093	3.040
Swarm size	100			1000			3000		

Standard deviations in parentheses
1,000 iterations for each simulation

Table 3.2: Random network

	$N = 5$			$N = 10$			$N = 15$		
	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$	$T = 25$	$T = 50$	$T = 100$
Link Detection									
Accuracy	0.983	0.998	1.000	0.946	0.982	0.994	0.897	0.948	0.970
Sensitivity	0.988	0.999	1.000	0.963	0.986	0.995	0.866	0.926	0.951
Specificity	0.981	0.998	1.000	0.942	0.981	0.993	0.902	0.952	0.973
Coefficients									
$\lambda_0 = 1$	1.336	1.072	1.037	1.060	0.981	0.966	0.750	0.754	0.814
	(3.123)	(0.682)	(0.444)	(0.836)	(0.454)	(0.292)	(0.519)	(0.317)	(0.221)
$\delta_{10} = 2$	2.721	2.196	2.112	2.303	2.007	1.981	1.625	1.505	1.550
	(2.784)	(0.631)	(0.345)	(0.667)	(0.346)	(0.251)	(0.315)	(0.240)	(0.251)
$\beta_{10} = 3$	4.007	3.228	3.160	3.271	2.989	2.964	2.176	2.194	2.334
	(3.723)	(0.868)	(0.480)	(0.940)	(0.496)	(0.365)	(0.408)	(0.341)	(0.360)
Reference									
$\lambda_0 = 1$	1.328	1.071	1.038	1.105	1.037	0.999	1.054	1.007	1.014
$\delta_{10} = 2$	2.660	2.191	2.112	2.222	2.099	2.047	2.123	2.073	2.028
$\beta_{10} = 3$	3.938	3.283	3.160	3.338	3.149	3.065	3.183	3.090	3.042
Swarm size	100			1000			3000		

Standard deviations in parentheses
1,000 iterations for each simulation

For $N = 15$, observe that the magnitudes of estimated β's are lower than the references. This bias is caused by the noise in the estimated weight matrices.

To clarify the direction of the bias, I present a set of Monte Carlo simulations with noisy networks in Appendix C. It shows that even small noises in an estimated network lead to downward biased coefficients. A consistent downward bias is observed from different levels of noise.

As the BPSO does not rely upon a guaranteed global optimum, it should be clarified if the inaccurate estimates from the low-sized samples are attributed to the optimizer or the size of the samples. To verify this, I test with different sizes of swarms.

As the swarm size increases, the link detection improves, and the coefficients accordingly. Considering the total number of possible combinations (2^{45}), even 30,000 particles are insufficient to cover the search space completely. However, it is confirmed that the link detection performance increases by using more computing power.

Table 3.3: Test with different swarm sizezs

	Reference	1,000	10,000	30,000
Link Detection				
Accuracy		0.965	0.967	0.967
Sensitivity		0.978	0.974	0.974
Specificity		0.962	0.965	0.965
Likelihood	28.743	28.941	26.201	26.054

$N = 10$ and $T = 25$
Standard deviations in parentheses
100 iterations for each simulation
Average true likelihood: 30.113

3.4 Conclusion

In this last chapter, I present an estimation algorithm for a binary choice SAR model with unknown network structures by extending the methodology proposed by de Paula, Rasul, and Souza (forthcoming). The links between individuals in a network are estimated by the binary particle swarm optimization process, combined with the Lasso estimator, which favors sparse models.

The algorithm evaluates each particle's likelihood function with a given weight matrix. The particles compare their values, and the entire swarm eventually finds the global optimum.

I perform Monte Carlo simulations with two stylized networks: a circle and a random network. The result shows that the proposed algorithm performs with respectable accuracy. Moreover, the number of required particles is not prohibitively high for modern computers.

Bibliography

Aguirregabiria, Victor, and Pedro Mira. 2002. “Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models.” **Econometrica** 70 (4): 1519–1543.

———. 2007. “Sequential Estimation of Dynamic Discrete Games.” **Econometrica** 75 (1): 1–53.

Anselin, Luc. 1988. **Spatial Econometrics: Methods and Models.** Kluwer Academic Publishers.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou. 2006. “Who’s Who in Networks. Wanted: The Key Player.” **Econometrica** 74 (5): 1403–1417.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2013. “The Diffusion of Microfinance.” **Science** 341 (6144): 1236498–1236498.

Blommestein, Hans J. 1983. “Specification and Estimation of Spatial Econometric Models: A Discussion of Alternative Strategies for Spatial Economic Modelling.” **Regional Science and Urban Economics** 13 (2): 251–270.

Boucher, Vincent. 2016. “Conformism and Self-selection in Social Networks.” **Journal of Public Economics** 136:30–44.

Boucher, Vincent, and Bernard Fortin. 2016. “The Oxford Handbook of the Economics of Networks,” 276–302.

Boucher, Vincent, and Aristide Houndetoungan. 2020. "Estimating Peer Effects Using Partial Network Data." **Working Paper, Centre de recherche sur les risques.**

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of Peer Effects Through Social Networks." **Journal of Econometrics** 150 (1): 41–55.

Bramoullé, Yann, Rachel Kranton, and Martin D'Amours. 2014. "Strategic Interaction and Networks." **American Economic Review** 104 (3): 898–930.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman. 2014. "Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions." **Econometrica** 82 (4): 1273–1301.

Chandrasekhar, Arun G., and Randall Lewis. 2011. "Econometrics of Sampled Networks." **Working Paper.**

Chomsisengphet, Souphala, Hua Kiefer, and Xiaodong Liu. 2018. "Spillover Effects in Home Mortgage Defaults: Identifying the Power Neighbor." **Regional Science and Urban Economics** 73:68–82.

Debarsy, Nicolas, and James LeSage. 2018. "Flexible Dependence Modeling Using Convex Combinations of Different Types of Connectivity Structures." **Regional Science and Urban Economics** 69:48–68.

Debarsy, Nicolas, and James P. LeSage. 2022. "Bayesian Model Averaging for Spatial Autoregressive Models Based on Convex Combinations of Different Types of Connectivity Matrices." **Journal of Business & Economic Statistics** 40 (2): 547–558.

De Paula, Aureo, Imran Rasul, and Pedro Souza. Forthcoming. "Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition." **Review of Economic Studies.**

Giorgi, Giacomo De, Anders Frederiksen, and Luigi Pistaferri. 2019. "Consumption Network Effects." **The Review of Economic Studies** 87 (1): 130–163.

Glaeser, Edward L., Bruce I. Sacerdote, and Jose A. Scheinkman. 2003. "The Social Multiplier." **Journal of the European Economic Association** 1 (2-3): 345–353.

Griffith, Alan. 2022. "Name Your Friends, but Only Five? The Importance of Censoring in Peer Effects Estimates Using Social Network Data." **Journal of Labor Economics** 40 (4): 779–805.

Higham, Nicholas J. 2008. **Functions of Matrices.** Society for Industrial / Applied Mathematics.

Hsieh, Chih-Sheng, and Xu Lin. 2017. "Gender and Racial Peer Effects with Endogenous Network Formation." **Regional Science and Urban Economics** 67 (Ann. Stat. 18 1990): 135–147.

Huang, Jun S. 1984. "The Autoregressive Moving Average Model for Spatial Analysis." **Australian Journal of Statistics** 26 (2): 169–178.

Ipsen, Ilse C F, and Dean J Lee. 2011. "Determinant Approximations." **arXiv.**

Kelejian, Harry H., and Ingmar R. Prucha. 1998a. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." **The Journal of Real Estate Finance and Economics** 17 (1): 99–121.

———. 1998b. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." **The Journal of Real Estate Finance and Economics** 17 (1): 99–121.

———. 2010. "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances." **Journal of Econometrics** 157 (1): 53–67.

Kennedy, J., and R.C. Eberhart. 1997. "A Discrete Binary Version of the Particle Swarm Algorithm." **1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation** 5:4104–4108 vol.5.

Kitts, James A., and Diego F. Leal. 2021. "What is(n't) a Friend? Dimensions of the Friendship Concept among Adolescents." **Social Networks** 66:161–170.

Lam, Clifford, and Pedro C.L. Souza. 2019. "Estimation and Selection of Spatial Weight Matrix in a Spatial Lag Model." **Journal of Business & Economic Statistics,** 1–41.

Lee, Lung-fei. 2007. "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models." **Journal of Econometrics** 137 (2): 489–514.

Lee, Lung-fei, Ji Li, and Xu Lin. 2014. "Binary Choice Models with Social Network under Heterogeneous Rational Expectations." **Review of Economics and Statistics** 96 (3): 402–417.

Lee, Lung-fei, and Xiaodong Liu. 2010. "Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances." **Econometric Theory** 26 (1): 187–230.

Lee, Lung-fei. 2003. "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances." **Econometric Reviews** 22 (4): 307–335.

Li, An-Da, Bing Xue, and Mengjie Zhang. 2021. "Improved Binary Particle Swarm Optimization for Feature Selection with New Initialization and Search Space Reduction Strategies." **Applied Soft Computing** 106:107302.

Lin, Zhongjian, Xun Tang, and Ning Neil Yu. 2020. "Uncovering Heterogeneous Social Effects in Binary Choices." **Journal of Econometrics.**

Liu, Xiaodong, Eleonora Patacchini, and Edoardo Rainone. 2017. “Peer Effects in Bedtime Decisions among Adolescents: A Social Network Model with Sampled Data.” **The Econometrics Journal** 20 (3): S103–S125.

Liu, Xiaodong, Eleonora Patacchini, and Yves Zenou. 2014. “Endogenous Peer Effects: Local Aggregate or Local Average?” **Journal of Economic Behavior & Organization** 103:39–59.

Liu, Xiaodong, and Jiannan Zhou. 2017. “A Social Interaction Model with Ordered Choices.” **Economics Letters** 161:86–89.

Manski, Charles F. 1993. “Identification of Endogenous Social Effects: The Reflection Problem.” **The Review of Economic Studies** 60 (3): 531.

Patacchini, Eleonora, and Yves Zenou. 2012. “Juvenile Delinquency and Conformism.” **The Journal of Law, Economics, and Organization** 28 (1): 1–31.

Rust, John. 1987. “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher.” **Econometrica** 55 (5): 999.

———. 1988. “Maximum Likelihood Estimation of Discrete Control Processes.” **SIAM Journal on Control and Optimization** 26 (5): 1006–1024.

Saberi, Majid, Reza Khosrowabadi, Ali Khatibi, Bratislav Misic, and Gholamreza Jafari. 2021. “Topological Impact of Negative Links on the Stability of Resting-state Brain Network.” **Scientific Reports** 11 (1): 2176.

Tibshirani, Robert. 1996. “Regression Shrinkage and Selection Via the Lasso.” **Journal of the Royal Statistical Society: Series B (Methodological)** 58 (1): 267–288.

Ushchev, Philip, and Yves Zenou. 2020. “Social Norms in Networks.” **Journal of Economic Theory** 185:104969.

Vieira, Susana M., Luís F. Mendonça, Gonçalo J. Farinha, and João M.C. Sousa. 2013. "Modified Binary PSO for Feature Selection using SVM Applied to Mortality Prediction of Septic Patients." **Applied Soft Computing** 13 (8): 3494–3504.

Yang, Chao, and Lung-fei Lee. 2017. "Social Interactions under Incomplete Information with Heterogeneous Expectations." **Journal of Econometrics** 198 (1): 65–83.

Yang, Chao, Lung-fei Lee, and Xi Qu. 2018. "Tobit Models with Social Interactions: Complete vs Incomplete Information." **Regional Science and Urban Economics** 73:30–50.

Appendix A

Appendix for Chapter 1

A.1 Proofs

Proof of Proposition 1.1. The estimate of the GSAR model is obtained by minimizing the sum of squared residuals:

$$SSR(\theta) = \Big(y - \big(I - \rho W(\gamma)\big)^{-1} X\beta\Big)' \Big(y - \big(I - \rho W(\gamma)\big)^{-1} X\beta\Big).$$

Define T_θ as follows:

$$T(\rho, \gamma, \beta) = T_\theta = \begin{bmatrix} -2X'\big(I - \rho W(\gamma)\big)^{-1} \\ -2\beta' X'\big(I - \rho W(\gamma)\big)^{-1'} W(\gamma)'\big(I - \rho W(\gamma)\big)^{-1'} \\ 2\beta' X'\big(I - \rho W(\gamma)\big)^{-1'} \rho D\big(I - \rho W(\gamma)\big)^{-1'} \end{bmatrix}.$$

Then, from the exogeneity of the covariates and the network structure, the following equation holds under the true parameters:

$$E[T_{\theta_0}\big(I - \rho_0 W(\gamma_0)\big)^{-1}\varepsilon)] = 0.$$

Therefore, the solution for the first-order condition exists.

Next, I show the uniqueness of the solution.

$$\begin{aligned}
\alpha(\theta) =& T_\theta\Big(y - \big(I - \rho W(\gamma)\big)^{-1} X\beta\Big) \\
=& T_\theta\Big(\big(I - \rho_0 W(\gamma_0)\big)^{-1} X\beta_0 - \big(I - \rho W(\gamma)\big)^{-1} X\beta\Big) \\
=& T_\theta\big(I - \rho W(\gamma)\big)^{-1}\Big(\big(I - \rho W(\gamma)\big)\big(I - \rho_0 W(\gamma_0)\big)^{-1} X\beta_0 - X\beta\Big) \\
=& T_\theta\big(I - \rho W(\gamma)\big)^{-1}\Big(\big(I - \rho_0 W(\gamma_0) + \rho_0 W(\gamma_0) - \rho W(\gamma)\big)\big(I - \rho_0 W(\gamma_0)\big)^{-1} X\beta_0 - X\beta\Big) \\
=& T_\theta\big(I - \rho W(\gamma)\big)^{-1}\Big(\big(S + \rho_0(A - \gamma_0 D) - \rho(A - \gamma D)\big)S^{-1} X\beta_0 - X\beta\Big) \\
=& T_\theta\big(I - \rho W(\gamma)\big)^{-1}\Big(\big(S + (\rho_0 - \rho)A - (\rho_0\gamma_0 - \rho\gamma)D\big)S^{-1} X\beta_0 - X\beta\Big) \\
=& T_\theta\big(I - \rho W(\gamma)\big)^{-1}\Big(X(\beta_0 - \beta) + AS^{-1}X\beta_0(\rho_0 - \rho) - DS^{-1}X\beta_0(\rho_0\gamma_0 - \rho\gamma)\Big) \\
=& T_\theta\big(I - \rho W(\gamma)\big)^{-1}[X, AS^{-1}X\beta_0, -DS^{-1}X\beta_0]\begin{bmatrix}\beta_0 - \beta \\ \rho_0 - \rho \\ \rho_0\gamma_0 - \rho\gamma\end{bmatrix},
\end{aligned}$$

where $S = I - \rho_0 W(\gamma_0)$.

The weight matrix, $\rho W(\gamma) = \rho A - \rho\gamma D$ is uniquely determined by ρ and γ because the adjacency matrix has nonzero elements except on its main diagonal, while the degree matrix is diagonal.

Therefore, if $T_\theta\big(I - \rho W(\gamma)\big)^{-1}[X, AS^{-1}X\beta_0, -DS^{-1}X\beta_0]$ has full rank, for any feasible values of ρ and γ, which is assumed by Assumption 1.2, the model is identified. □

Proof of Proposition 1.3. The threshold for the social multiplier effect is obtained by solving the following equation with respect to ρ:

$$\det(I - \rho W) = \det(I - \rho(A - \gamma D)) = 1.$$

Denote the solution as $\bar{\rho}(\gamma)$. Then, apply the sparse matrix approximation (Higham 2008; Ipsen and Lee 2011) and obtain

$$\det(I - \rho(A - \gamma D)) \approx \exp(\mathrm{tr}(\log(I - \rho(A - \gamma D)))) = 1.$$

By taking a logarithm on both-hand sides, the problem reduces to the following equation.

$$\mathrm{tr}(\log(I - \rho(A - \gamma D))) = 0$$

For the logarithm in the trace, use the power series expansion of matrix logarithm, which is available due to Assumption 1.1.

$$\log(I - \rho W) = \frac{(-\rho(A - \gamma D))^1}{1} - \frac{(-\rho(A - \gamma D))^2}{2} + \frac{(-\rho(A - \gamma D))^3}{3} + \cdots \tag{A.1a}$$

$$= -\rho(A - \gamma D) - \frac{\rho^2(A - \gamma D)^2}{2} - \frac{\rho^3(A - \gamma D)^3}{3} - \cdots \tag{A.1b}$$

The expansion (A.1b) is truncated at the second term, and each term is separated by the property of trace.

$$\mathrm{tr}\left(-\rho(A - \gamma D) - \frac{\rho^2(A - \gamma D)^2}{2}\right) = 0 \tag{A.2a}$$

The first term of (A.2a) is represented as a sum of degrees:

$$\mathrm{tr}(-\rho(A - \gamma D)) = -\rho(\mathrm{tr}(A) - \gamma \mathrm{tr}(D)) = \rho\gamma \sum_i d_i,$$

and the second term is:

$$\begin{aligned}
\mathrm{tr}\left(-\frac{\rho^2(A - \gamma D)^2}{2}\right) &= \mathrm{tr}(-\frac{\rho^2}{2}(A^2 - \gamma AD - \gamma DA + \gamma^2 D^2)) \\
&= -\frac{\rho^2}{2}\sum_{i=1}^{N}\sum_{j \neq i} a_{ij}a_{ji} - \frac{\rho^2\gamma^2}{2}\sum_i d_i^2 \\
&= -\frac{\rho^2}{2}\sum_{i=1}^{N} d_i - \frac{\rho^2\gamma^2}{2}\sum_i d_i^2. \quad \text{(by the symmetric adjacency matrix)}
\end{aligned}$$

Then, the equation (8a) can be rewritten by adding the two terms:

$$\rho\gamma \sum_{i=1}^{N} d_i - \frac{\rho^2}{2}\sum_{i=1}^{N} d_i - \frac{\rho^2\gamma^2}{2}\sum_i d_i^2 = 0$$

$$\gamma \sum_{i=1}^{N} d_i - \rho\frac{1}{2}\sum_{i=1}^{N} d_i - \rho\frac{\gamma^2}{2}\sum_i d_i^2 = 0$$

The equation is a polynomial equation of degree 1, and its solution is

$$\bar{\rho}(\gamma) = \frac{2\gamma \sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} d_i + \gamma^2 \sum_{i=1}^{N} d_i^2} = \frac{2\gamma}{1 + \gamma^2 \sum_{i=1}^{N} d_i^2 / \sum_{i=1}^{N} d_i}.$$

This completes the proof. □

Appendix B

Appendix for Chapter 2

B.1 Proof

The proof below follows the argument of Liu and Zhou (2017).

Proof of Proposition 2.1. Consider two observationally equivalent sets of parameters, $\theta = (\rho, \gamma, \beta)$ and $\tilde{\theta} = (\tilde{\rho}, \tilde{\gamma}, \tilde{\beta})$. Then, for each set, there exists an equilibrium that satisfies

$$p_i^*(\theta) = F_i\Bigg(\sum_{q=1}^{Q} \rho_q a_{iq} \mathbf{p}^*(\theta) - \sum_{q=1}^{Q} \rho_q \gamma d_{iq} \mathbf{p}^*(\theta) + x_i\beta + \mathbb{I}_{(v>1)}\tau\Bigg)$$
$$\tilde{p}_i^*(\tilde{\theta}) = F_i\Bigg(\sum_{q=1}^{Q} \tilde{\rho}_q a_{iq} \mathbf{p}^*(\tilde{\theta}) - \sum_{q=1}^{Q} \rho_q \tilde{\gamma} d_{iq} \mathbf{p}^*(\tilde{\theta}) + x_i\tilde{\beta} + \mathbb{I}_{(v>1)}\tilde{\tau}\Bigg),$$

where a_{iq} and d_{iq} are i-th row of A_q and D_q. The probability vector is uniquely determined with θ by the fixed point theorem, and denote it as $\mathbf{p}^*(\theta) = (p^*(\theta)_i, \ldots, p^*(\theta)_N)$.

Since $p_i^*(\theta)$ and $\tilde{p}_i^*(\tilde{\theta})$ are observationally equivalent, $p_i^*(\theta) = \tilde{p_i}^*(\tilde{\theta}) = p_i^*$. Then, $F_i(\rho a_i \mathbf{p}^* - \rho\gamma d_i \mathbf{p}^* + x_i\beta + \mathbb{I}_{(v>1)}\tau) = F_i(\tilde{\rho} a_i \mathbf{p}^* - \rho\tilde{\gamma} d_i \mathbf{p}^* + x_i\tilde{\beta} + \mathbb{I}_{(v>1)}\tilde{\tau})$. By the assumption of strictly increasing CDF (Assumption 2.4.(i)), the model parameters are identified if the following equation has a unique solution.

$$[A_1\mathbf{p}^*, \ldots, A_Q\mathbf{p}^*, D_1\mathbf{p}^*, \ldots, D_Q\mathbf{p}^*, X, \mathbb{I}] \begin{bmatrix} \rho_1 - \tilde{\rho}_1 \\ \vdots \\ \rho_Q - \tilde{\rho}_Q \\ \rho_1\gamma - \tilde{\rho}_1\tilde{\gamma} \\ \vdots \\ \rho_Q\gamma - \tilde{\rho}_Q\tilde{\gamma} \\ \beta - \tilde{\beta} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{B.1}$$

Under Assumption 2.4.(ii), $[A_1\mathbf{p}^*, \ldots, A_Q\mathbf{p}^*, D_1\mathbf{p}^*, \ldots, D_Q\mathbf{p}^*, X, \mathbb{I}]$ has full rank and θ is uniquely deteremined.

□

B.2 Tables

Table B.1: Full descriptive statistics

			Whole samples		Participating samples		Participating samples w/ caste data	
			N=14,890		N=9,591		N=7,919	
			Number of villages = 75		Number of villages = 43		Number of villages = 43	
	Min	Max	Mean	S.D.	Mean	S.D.	Mean	S.D.
Outcome variable								
Microfinance take-up	0	1			0.18	0.39	0.18	0.38
Income proxies								
Roof material: tile	0	1	0.32	0.47	0.34	0.47	0.37	0.48
Number of rooms	0	20	2.36	1.32	2.34	1.31	2.41	1.30
Number of beds	0	50	0.85	1.33	0.86	1.42	0.82	1.23
Without Latrine	0	1	0.73	0.44	0.73	0.45	0.73	0.45
House not owned	0	1	0.10	0.30	0.10	0.30	0.10	0.30
Without electricity	0	1	0.38	0.49	0.38	0.48	0.38	0.48
Social statuses								
Leader group	0	1	0.12	0.33	0.12	0.33	0.13	0.33
Other backward castes							0.53	0.50
General caste							0.12	0.33
Minority castes							0.2	0.13
Scheduled caste							0.28	0.45
Scheduled tribe							0.05	0.22
Religions								
Hinduism	0	1	0.96	0.20	0.95	0.21	0.95	0.21
Islam	0	1	0.04	0.20	0.05	0.21	0.05	0.21
Village								
Village Population					223.21	56.17	184.16	86.08
Graph statistics								
Union								
Average degree					9.20	1.65	9.22	1.69
Density					0.04	0.01	0.06	0.04
Spectral radius					15.08	2.56	14.07	3.20
Individual networks								
Average degree					2.73	0.30	3.00	0.24
Density					0.01	0.001	0.03	0.01
Spectral radius					5.76	0.39	5.22	0.66

Table B.2: OLS estimation on the network degree

	Coefficents	S.E.	t-stats	P-value
Regressors				
Microfinance take-up	-0.001	0.004	-0.347	0.729
Roof material: tile	-0.001	0.002	-0.736	0.463
Number of rooms	0.098	0.015	6.493	0.000
Number of beds	0.062	0.009	6.840	0.000
Without Latrine	-0.021	0.004	-5.606	0.000
House not owned	-0.005	0.003	-1.775	0.078
Without electricity	-0.016	0.004	-3.657	0.000
Leader group	0.005	0.003	1.904	0.059
Other backward castes	0.009	0.004	2.213	0.028
General caste	-0.001	0.001	-1.392	0.166
Scheduled caste	-0.006	0.004	-1.562	0.120
Scheduled tribe	-0.002	0.002	-0.955	0.341
Islam	-0.002	0.002	-1.314	0.191

The regressand is the degree of individuals.

Table B.3: BSS leader statistics

Variables	Leaders Mean	S.D.	Min	Max	Non-leaders Mean	S.D.	Min	Max
Outcome variable								
Microfinance take-up	0.24	0.42	0	1	0.17	0.38	0	1
Income proxies								
Roof material: tile	0.31	0.46	0	1	0.38	0.48	0	1
Number of rooms	2.82	1.61	0	18	2.35	1.24	0	19
Number of beds	1.10	1.44	0	12	0.78	1.20	0	24
No latrine in house	0.60	0.49	0	1	0.74	0.44	0	1
House not owned	0.07	0.26	0	1	0.10	0.30	0	1
No access to electricity	0.27	0.44	0	1	0.39	0.49	0	1
Social statuses								
Other backward castes	0.53	0.50	0	1	0.53	0.50	0	1
General caste	0.20	0.40	0	1	0.11	0.32	0	1
Minority castes	0.02	0.12	0	1	0.02	0.13	0	1
Scheduled caste	0.21	0.41	0	1	0.29	0.45	0	1
Scheduled tribe	0.04	0.19	0	1	0.05	0.22	0	1
Religions								
Hinduism	0.96	0.21	0	1	0.95	0.21	0	1
Islam	0.04	0.21	0	1	0.05	0.21	0	1
Graph Statistics								
Degree	8.36	5.75	0	46	6.36	4.89	0	58
Eigenvector centrality								
Katz-Bonacich centrality								
Sample size	995				6,924			

The degrees are obtained from the union of Relatives, Temple and Visit networks.

Table B.4: Counterfactual leader statistics

Variables	Leaders				Non-leaders			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Outcome variable								
Microfinance take-up	0.30	0.46	0	1	0.16	0.37	0	1
Income proxies								
Roof material: tile	0.50	0.50	0	1	0.35	0.48	0	1
Number of rooms	2.00	0.96	0	10	2.47	1.33	0	19
Number of beds	0.49	0.86	0	10	0.86	1.27	0	24
No latrine in house	0.92	0.27	0	1	0.70	0.46	0	1
House not owned	0.13	0.33	0	1	0.09	0.29	0	1
No access to electricity	0.67	0.47	0	1	0.33	0.47	0	1
Social statuses								
Other backward castes	0.23	0.42	0	1	0.57	0.49	0	1
General caste	0.01	0.07	0	1	0.14	0.35	0	1
Minority castes	0.03	0.16	0	1	0.02	0.13	0	1
Scheduled caste	0.61	0.49	0	1	0.23	0.42	0	1
Scheduled tribe	0.12	0.33	0	1	0.04	0.19	0	1
Religions								
Hinduism	0.88	0.32	0	1	0.96	0.19	0	1
Islam	0.12	0.32	0	1	0.04	0.19	0	1
Graph Statistics								
Degree	11.14	5.28	0	58	5.96	4.67	0	40
Eigenvector centrality								
Katz-Bonacich centrality								
Sample size	995				6,924			

The degrees are obtained from the union of Relatives, Temple and Visit networks.
The outcome variable is from the original data, not counterfactual.

B.3 Peer Effect on Homophily Networks

In this section, I present estimation results with alternative specifications using hypothetical networks based on homophily. Homophily refers to the tendency to form social networks with people who share similar characteristics.

My estimation method does not address endogeneity on networks. Therefore, potential sources of endogeneity should be checked to ensure unbiasedness.

As mentioned in Section 2.1.1, there is a tendency to report those with higher socioeconomic status as peers when the villagers are asked to reveal their peers through the questions related to economic status, i.e., from whom they want to borrow money. Since none of those networks turned out to be significant, no additional treatment was needed.

Still, social status, which is no less significant than economic status, was not asked to reveal any peers connected through it. For example, a question such as "How many of your friends belong to the same caste as yours?" could have been asked.

To examine this issue, I construct hypothetical networks based on homophily and use them for the GSAR estimation in this section. Thanks to the higher-order feature of the model, any number of hypothetical networks may be included. I build hypothetical networks derived from each of the six social status variables. Each dimension represents the similarity of a pair of individuals regarding each attribute. For example, if individual i and j share the same religion, they will be considered as peers in "Religion" network.[26]

In Table B.5, it is shown that none of the homophily networks are positively significant. Two of them, "general caste" and "scheduled caste," show negative estimates at the significance level of 1%. However, the magnitudes of those estimates are very small. Despite their significance, this does not necessarily mean that those coefficients should be interpreted as strongly negative peer effects.

A noticeable difference is that the two of the surveyed networks that are significant in the

26. When constructing the homophily networks, "other backward castes" are excluded because they are the majority in most villages, which may lead to a complete network.

original estimation, “Relatives” and “Temple,” are now insignificant. Obviously, these two are very closely correlated with social status. A possible explanation of the result, therefore, can be that its effect on the network is dispersed to the other newly included networks.

Observe that the explanatory power of the social status variables does not move to the homophily networks. Also, due to the low intensities of the homophily networks, there is no remarkable change in the rest of the estimates, including marginal effects.

Of course, there may be more unobserved factors that drive network formation. However, considering that socioeconomic status is the most substantial factor in most cases, the major source of endogeneity is examined by this hypothetical network analysis.

B.4 Discussion on Banerjee et al. (2013)

This section discusses the methodological differences between the GSAR model and the diffusion model of Banerjee et al. (2013), BCDJ henceforth. The main finding of BCDJ is the role of non-participants in the information diffusion process. Although the non-participants are less likely to pass information to others, they play a larger role in the overall take-up rate due to their larger number. Meanwhile, the authors do not find any evidence of the “endorsement effect,” which corresponds to the peer effect of this paper, in their study. In other words, not only does the probability of participation not increase with the number of already participating peers, but it even slightly decreases. Their paper does not fully explain this negative result, but some hints are suggested by Chandrasekhar and Lewis (2011). In that paper, the insignificant peer effect is attributed to the missing links in sampled networks. As mentioned in Chapter 4, the organizers surveyed only 50% of the villagers, so the issue of misspecification is not inconceivable.

Conversely, the peer effect is estimated as significantly positive in this paper with the same surveyed networks. This result may be viewed from a different angle because the missing links can only decrease the estimate of peer effect, not increase. For the estimation, BCDJ uses a two-step procedure where the coefficients for the independent variables are estimated without networks on the leader group first, and those first-step estimates are used to estimate the probability of

Table B.5: Estimation results with homophily networks

	GSAR		Marginal effects		
	Coefficient	SE	Naive	Direct	Indirect
ρ					
Surveyed networks					
Money	0.033	0.098	0.44		
Advice	−0.036	0.060	−0.48		
Kerosine	−0.009	0.084	−0.12		
Medical	0.114	0.089	1.52		
Non-relatives	0.028	0.079	0.38		
Relatives	0.060	0.091	0.80		
Temple	0.252	0.193	3.37		
Visit	0.223***	0.086	3.11		
Homophily networks					
Leader group	−0.024	0.017	−0.32		
General caste	−0.033***	0.005	−0.44		
Minority castes	0.020	0.120	0.27		
Scheduled caste	−0.017***	0.003	−0.22		
Scheduled tribe	0.004	0.124	0.05		
Islam	0.011	0.031	0.15		
γ	0.371***	0.123			
β					
Constant	−1.481***	0.240			
Roof material: tile	0.076	0.071	1.02	1.06	0.002
No. of rooms	−0.083***	0.030	−1.10	−1.16	−0.002
No. of beds	−0.024	0.032	−0.32	−0.33	−0.001
No latrine in house	0.375***	0.086	4.76	4.89	0.009
House not owned	0.015	0.096	0.20	0.22	0.000
No access to electricity	0.166**	0.068	2.24	2.36	0.004
Leader group	0.586***	0.092	8.79	9.23	0.015
General caste	−0.298**	0.127	−3.72	−3.95	−0.007
Minority castes	−0.042	0.292	−0.55	−0.58	−0.001
Scheduled caste	0.397***	0.077	5.55	5.71	0.010
Scheduled tribe	0.379***	0.144	5.54	5.95	0.009
Islam	1.026***	0.165	17.19	19.52	0.027
Multiplier					
Likelihood	−3409.98				
AIC	6875.96				
BIC	7071.32				

Significant at *10%, **5%, ***1%.

passing information and the endorsement effect. To describe the diffusion process, they adopt the SIR (Susceptible-Infected-Recovered) model. In that model, the individuals are not aware of the microfinance program until they are reached by those who have already acquired that information (susceptible). Once they learn about the program, a decision must be made at the moment of learning (infected), and there is no second chance thereafter (recovered). At each stage, the decision is made by the following logistic function:

$$\log \frac{p_{it}}{1-p_{it}} = X_i\hat{\beta} + \lambda F_{it},$$

where F_{it} is the number of participants in one's peer group divided by the number of the peers informed about the program. The subscript t denotes the choice probability obtained for each time period. Note that $\hat{\beta}$ is not estimated along with λ. It is a first-step estimate from the leader group, and the other parameters, including λ, is estimated later based on $\hat{\beta}$. Other than this, the specification is very similar to the local-average model, and a guess can be made that the idea of the endorsement effect is based on the positive social multiplier. From this, a hypothetical SAR estimation can be conducted to simulate their methodology. The peer effect is separately estimated with the coefficients for the household characteristics estimated from the leader groups. The two-step procedure is: 1) estimate the coefficients only with the leader groups, and 2) estimate the peer effect with the entire sample using the estimated coefficients. The main assumption implicitly made here is that the coefficients do not vary across the rest of the individuals outside of the leader group. As the leaders are chosen by the organization based on their characteristics, it is reasonable to assume that there is no significant social link between them. Therefore, the estimated coefficients will be asymptotically unbiased if there are no unobservable differences. Unfortunately, a regression including the leader dummy as a covariate shows that the leaders have a significantly higher tendency to participate.[27] Therefore, one can expect that the coefficients for the non-leader individuals, which should be lower than those of the leaders, will be overestimated and compensated by the negative peer effect. This conjecture is supported by the estimation result shown in B.6.

27. See Table B.7.

The explanatory variables are based on the dataset and codes published by the authors. It shows

Table B.6: The two-step and the SAR estimation

	Two-step estimation	SAR
	N=995	N=7919
Constant	-.871***	-1.668***
	(.239)	(.107)
Number of rooms	-.155**	-.086***
	(.066)	(.029)
Number of beds	-.089	-.056*
	(.075)	(.030)
No private latrine	.260	.129*
	(.177)	(.076)
House not owned	.127	.335***
	(.178)	(.062)
	N=7919	N=7919
Peer effect	-.122***	.082***
	(.015)	(.012)

that the estimated peer effect is significantly negative, which is parallel to the estimate reported by BCDJ.[28] As expected, the standard SAR regression on the entire sample, including the leader group, shows a lower constant term and a positive peer effect. The lower estimate for the constant suggests that the entire group has a much lower baseline preference for microfinance.

Another possible explanation is the difference in the timeline. First, the SIR diffusion model does not allow late decisions. The spread pattern over time assumes that the newcomers are only those who are newly informed about the program. However, it is also possible that they initially declined at the time of learning but opted in later.[29] . In that case, the peer effect on the equilibrium of the SAR model will be more accurate.

That said, these results do not imply that these two approaches are exclusive. The SAR model has a limitation: it relies upon the strong assumption of the rational expectation and does not reflect the change of the take-up rate over time. Instead, estimation is easier to perform and does not require panel data. Meanwhile, the diffusion model explains the diffusion pattern over

28. See Table B.7 for the two-step estimates with a richer set of covariates. The result is similar.

29. This is not verifiable with the published data because the panel data is only available as overall participation rates over time and the participants are not identifiable.

time better, but the probability of passing information must be estimated as well. The additional computational burden and theoretical assumptions will be the cost that must be incurred.

Table B.7: The two-step and the SAR estimation with additional covariates

	Two-step estimation	SAR
	N=995	N=7919
Constant	-1.186***	-2.050***
	(.260)	(.116)
Roof material	.251	.143**
	(.166)	(.060)
Number of rooms	-.131**	-.063**
	(.067)	(.030)
Number of beds	-.063	-.032
	(.076)	(.030)
Islam	1.290***	1.046***
	(.404)	(.138)
No private latrine	.337*	.200**
	(.187)	(.080)
House not owned	-.159	.057
	(.308)	(.096)
House not owned	-.159	.057
	(.308)	(.096)
No electricity	.022	.214***
	(.190)	(.067)
General caste	-.153	-.409***
	(.221)	(.113)
Minority caste	.559	.153
	(.659)	(.231)
Scheduled caste	.385*	.396***
	(.198)	(.068)
Scheduled tribe	-.390	.223*
	(.448)	(.131)
Leader		.564***
		(.085)
	N=7919	N=7919
Peer effect	-.122***	.082***
	(.015)	(.012)

Appendix C

Appendix for Chapter 3

C.1 Biases from Noisy Networks

In Chapter 3, it was noted that the estimates are downward biased by noisy estimated networks. Due to the nature of the estimation process, some noise is inevitable even with large enough samples. In this section, I delineate the pattern of the bias through Monte Carlo simulations.

As mentioned earlier, the noise can be categorized into false positives and false negatives. To experiment with both types of false signals, I build noisy networks according to the procedure introduced below.

I add false links into a true network, represented by an adjacency matrix, A, for the false positives and generate a noisy network, A_{noise}. The level of a "positive noise" is defined as:

$$noise_{pos} = \frac{||A_{noise} - A||}{||A||},$$

where $||A||$ is the number of existing links.

For the false negatives, I subtract true links from a true adjacency matrix. Similarly, the level of a "negative noise" is:

$$noise_{neg} = \frac{||A - A_{noise}||}{||A||}.$$

I experiment with three levels of noise between 1% to 10%.

Toy data are generated by the same process as the previous simulations. The weight matrix is derived by row-sum normalizing the noisy adjacency matrix, A_{noise}. I use $N = 15$ and $T = 50$

with the circular and random networks for the following cases. As a benchmark case, I run ordinary logit estimation for both cases, ignoring the peer effect and networks, and obtain $\hat{\beta}_{circle} = 1.096$ and $\hat{\beta}_{random} = 0.986$, respectively.

The results are summarized in Table C.1 and Table C.2. Observe that the bias becomes more pronounced as the noise level increases.[30]

Despite the bias, note that the bias from the estimated network is much less than that from the ordinary logit model. Thus, when large enough data are available, and a researcher believes that a significant peer effect exists, estimating a network, even with a lot of noise, yields far better results than completely ignoring it.

Table C.1: SAR estimations with positive noise

	Circular network			Random network		
	1%	5%	10%	1%	5%	10%
$\lambda_0 = 1$	0.993	0.941	0.898	0.950	0.898	0.849
	(.336)	(.328)	(.320)	(.330)	(.318)	(.309)
$\delta_0 = 2$	1.938	1.828	1.737	1.782	1.583	1.398
	(.226)	(.209)	(.196)	(.196)	(.164)	(.139)
$\beta_0 = 3$	2.901	2.738	2.603	2.808	2.578	2.405
	(.307)	(.282)	(.260)	(.295)	(.253)	(.227)

Standard deviations in parentheses
1,000 iterations

Table C.2: SAR estimations with negative noise

	Circular network			Random network		
	1%	5%	10%	1%	5%	10%
$\lambda_0 = 1$	0.861	0.754	0.681	0.923	0.843	0.789
	(.321)	(.304)	(.290)	(.327)	(.314)	(.302)
$\delta_0 = 2$	1.639	1.378	1.184	1.782	1.583	1.398
	(.167)	(.133)	(.114)	(.196)	(.164)	(.139)
$\beta_0 = 3$	2.617	2.334	2.137	2.763	2.539	2.330
	(.256)	(.217)	(.195)	(.288)	(.253)	(.222)

Standard deviations in parentheses
1,000 iterations

30. In a circular network, each individual has two links. Accordingly, the total number of links is 30. Then, one false positive link will be equal to $1/30 = 0.033$, or about 3%. Likewise, three false positive links will be 10% level. Also, removing three links (false negative) will result in 10% noise level.