Abstract

How does the overall structure of network interactions in a community affect the spread of new products, norms, attitudes, or behaviors? This research project answers this question by examining whether peer (spillover) effects are facilitated or hindered by a pervasive feature of real-life social networks, namely, network cohesiveness measured by a clustering coefficient. We theoretically and empirically explore the impact of the clustering coefficient on the spread of phenomena of economic interest in networks. Although sociologists and economists agree that local cohesion influences such spread of phenomena, no consensus exists regarding the direction of such an effect probably due to the lack of a systematic theoretical and/or empirical analysis of the role of network cohesiveness in diffusion processes.

The contribution of this research is threefold. First, we develop theoretical hypotheses regarding the role of the clustering coefficient by applying standard approaches from strategic interaction models. We predict that the effect of the clustering coefficient depends on the nature of externalities for the diffused phenomenon; that is, network cohesion facilitates diffusion under positive externalities and prevents it under negative externalities.

Second, we propose an empirical approach under the framework of spatial autoregressive (SAR) models to capture the effect of network cohesion on social interactions. The proposed strategy extends the traditional SAR model with a random coefficient specification on the endogenous peer-effect parameter. To support this extension, we address two relevant econometric issues, namely, (i) theoretical bounds on endogenous peer effects and (ii) the endogeneity problem from network link selections.

Lastly, we apply our approach to four large-scale network datasets embedded in different socio-economic contexts. We observe the outcome of the diffusion of various phenomena ranging from smoking, schooling outcomes, knowledge of other households' economic well-being to adoption of microfinance or new mobile phones. The important common features of these data are the presence of many independent networked groups that allow for cross-network comparisons of global network properties and the behavior in each community being subject to strong peer effects. Moreover, the diversity of the diffused phenomena across the datasets allows us to analyze the predicted interaction between the clustering coefficient and the sign of the externalities.

Long-Term Impact

This projects aims to provide an integral theoretical and empirical analysis of the clustering coefficient, which is a measure of network cohesion and a fundamental feature of real-life networks. With regard to academic impact, first, we can inform scholars about the role of this network property in how different economic agents influence one another and what determines this process. This contribution is important to both sociology and economics, because no theoretical consensus exists regarding the role of cohesive neighborhoods (see the classic debate in sociology between Granovetter (1985) or Coleman (1988) vs. Burt (1992) regarding social closure vs. structural holes) and because empirical evidence suggests that social cohesion may stimulate the spread of particular behaviors in several contexts while hindering the spread of others in other contexts (e.g., Podolny and Baron (1997); Gargiulo and Benassi (2000); Burt (2005)). This research can thus settle a long-term debate concerning the role of social cohesion.

Second, we provide a methodological contribution to the future analysis of *heterogeneous* network effects. Specifically, we extend the current workhorse model for social interaction studies on networks, the spatial autoregressive (SAR) model, with a random coefficient specification on peer effects. This extension allows researchers to identify and estimate the impact of the specific features of each community on the magnitude of endogenous peer effects. Different features reflect different network properties in our applications, but they can also reflect other variable of interest. We hope that our methodology will stimulate research on how different cultural aspects (e.g., different market conditions, social norms, or consumer habits across societies) shape diffusion. Such a cross-cultural analysis in the framework of network effects is virtually inexistent and our methodology may pave the way for a new research line.

With regard to practical impact, network data analysis and active targeting of particular individuals in social, firm, interbank, or even state networks have became common in business and politics due to the emergence of "big data." Analysis and regulation of the structure of financial markets have become routine, and the role of production chain networks in systemic stability and price setting has been acknowledged. Many research agencies and research-oriented firms actively promote new research ties to exploit network spillovers, and many firms and marketing practitioners target specific market actors, particularly the key players, to maximize the spread of products (see Banerjee et al. (2013); Acemoglu et al. (2012); Elliott et al. (2014); Galasso and Ravallion (2005); Alderman and Haque (2006) among many others). This research provides new policy recommendations for all these actors with regard to the improved design of socio-economic systems and enhanced prevention or promotion of behaviors; it may even guide existing public health policies that deal with the spread of diseases or desirable habits.

Objectives

This proposal has the following objectives.

1. Propose a theoretical model that explains the conditions under which network clustering affects the strength of network influences.
2. Construct an empirical approach based on a spatial autoregressive-type model. The network effects of clustering are explicitly considered in this approach through random coefficient specification.
3. Provide a thorough empirical analysis of the effect of clustering on the diffusion of educational outcomes, smoking behavior, adoption of new products (mobile phones and microfinance), and information by using four large-scale network datasets.
4. Provide a strategic guide for optimal diffusion policies for policy makers or marketing practitioners.

1 Background of the Research

Sociologists and economists agree that economic agents do not act in isolation. Instead, most individuals, firms, and countries are embedded in complex networks of relationships with other actors and are thus subject to peer effects and network externalities (Granovetter, 1985; Coleman, 1988; Burt, 1992; Karlan et al., 2009; Glaeser et al., 2003). Literature has documented considerable network effects on educational outcomes (e.g., Calvó-Armengol et al., 2009; Lin, 2010), labor markets (Topa, 2001), technology adoption (Conley and Udry, 2010), product adoption (Banerjee et al., 2013), and adoption of attitudes, values, or behaviors (Jackson and Yariv, 2007; Centola, 2010, 2011) among many other examples. Moreover, the recent financial crisis has revealed the presence of important network influences on financial and other markets and has indicated that the financial problems of partners or competitors can diffuse through the financial system and production chains (see Acemoglu et al. (2012) and Elliott et al. (2014) for two recent examples in economics).

Given the abundant evidence on peer and network effects in most contexts of economic interest, new developments in network theory, and increasing availability of network data, economists have begun to analyze diffusion and network effects extensively. Many new questions have emerged. This proposal addresses the question of how the underlying network structure enhances or prevents the diffusion of norms, behaviors, and attitudes. Theoretically, related literature has proposed several models that highlight the role of network features, such as average connectivity, variance in connectivity distribution, assortativity, short distances, and spectral properties of the adjacency matrix (Jackson and Yariv, 2007; Bollobás et al., 2010; Bramoullé et al., 2014; Ballester et al., 2006). Nevertheless, this line of literature provides little understanding of the role of local cohesion.

The objective of this research is to perform an exhaustive theoretical and empirical analysis of the role of social cohesion within the neighborhoods of agents in network effects. The standard measure of such local cohesion is the *clustering coefficient*, which is the ratio between the number of friends of an individual who are mutual friends themselves and the number of all possible friends in the individual's neighborhood. The higher the clustering coefficient is, the denser the connectivity within one's circle of friends is. The denser the connectivity is, the more cohesive the neighborhood is. The *average clustering coefficient* is the average of this measure across all the members of the network and reflects the overall local density of relationships in the network.

Two important reasons explain why the clustering coefficient deserves attention within the framework of diffusion and peer effects. First, a high clustering coefficient is one of the most pervasive features of real-life social networks (Jackson and Rogers, 2007; Jackson, 2008). Moreover, as argued for decades in sociology, the clustering coefficient may play a key role in the diffusion of socio-economic phenomena. Nevertheless, no consensus exists regarding the direction of its effect. On the one hand, Coleman (1988) and his colleagues argued in favor of cohesive neighborhoods as promoters of trust; thus, such neighborhoods promote the adoption of novel norms and behaviors. On the other hand, Burt (1992) and others advocated that network brokerage and, consequently, low neighborhood density, facilitates the flow of information and the adoption of new behaviors, products, and norms; high neighborhood density inhibits the inflow of new information and may cause old norms to persist. In sum, scholars have hypothesized that the clustering coefficient of a network exerts an effect on diffusiveness in a society, but they disagreed on how this effect occurs. A possible reason is that, to the best of our knowledge, no formal model provides a hypothesis for this direction, and little empirical work has focused on the role of this feature in the context of network effects.

Several authors have suggested that the effect of network cohesion depends on the application because local network density and the clustering coefficient appear to favor the spread of behaviors in several contexts while preventing them in others (Podolny and Baron, 1997; Gargiulo and Benassi, 2000; Burt, 2005). This observation provides additional motivation for this project. Are these differences in literature caused by the conditions of the context or by the different network structures across the analyzed contexts? Do phenomena exhibit positive externalities (e.g., schooling outcomes or diffusion of microfinance) generate the same network effects as those exhibiting negative externalities (e.g., smoking or misbehavior)? The difficulty of answering these questions may originate from several empirical issues. First, observing more than one network or the diffusion of two different phenomena in the same network is difficult. For example, Centola (2010) performed an experiment on two artificially designed networks and argued that the clustering coefficient enhances product adoption. However, observing only two networks forced him to confound the effect of the clustering coefficient with the shortening of network distances (similar to Watts and Strogatz (1998)).

The second reason why the clustering coefficient deserves attention within the framework of diffusion and peer effects is the estimation of network effects. Identification and estimation of peer effects from social interaction models have elicited increasing attention, and applications range from group interactions using the linear-in-means model (e.g., Hoxby, 2000; Gaviria and Raphael, 2001) to network interactions using spatial autoregressive (SAR) models (e.g., Calvó-Armengol et al., 2009; Bramoullé et al., 2009; Lin, 2010). The advantage of the latter is that the "reflection problem" (Manski, 1993) inherited from the linear-in-means model can be bypassed. Considering that the reflection problem prevents researchers from separating the endogenous peer effect from contextual exogenous effects, we adopt the SAR modeling approach in this project. Recently, apart from the identification of a unified endogenous peer effect, researchers have also begun to explore the *heterogenous* nature of the endogenous effect (Masten, 2014; Patacchini et al., 2017; Hsieh and Lin, 2017; Hsieh and van Kippersluis, 2017). Our study contributes to this emerging research trend by (1) proposing that heterogeneous network effects across people may originate from different local network environments and (2) providing a formal estimation framework that enables the identification of such heterogeneity.

Lastly, only a few of studies have analyzed the *global* network structural effects on individual outcomes (see Banerjee et al., 2013; Hu et al., 2014; Alatas et al., 2016). A possible reason is the lack of empirical data that provide sufficient network communities. Another reason is the inability to capture global network structural effects in the modeling approaches mentioned above. Our study aims to overcome both limitations.

2 Research Plan and Methodology

The current research simultaneously tackles the theoretical and empirical questions regarding the role of clustering in network diffusion processes. Our strategy involves three steps. First, we develop a theoretical framework based on standard strategic interaction models that permit the analysis of local network density and the nature of externalities embedded in the diffused phenomenon. Second, we extend the SAR model with a random coefficient specification on the endogenous peer-effect parameter. The objective is to estimate heterogeneity in network effects across networks in the function of their clustering. Third, we apply this extended model to study four network datasets (i) that contain a large number of network communities (thus enabling a cross-network comparison of global network structural effects), (ii) that cover different contexts and economic outcomes (including behaviors generating positive and negative externalities; see below for details), and (ii) in which the behavior within each community is subject to strong peer influences. The combination of the different datasets and the use of the proposed methodology overcome the empirical challenges in the studies discussed above.

2.1 Theoretical Hypotheses

This subsection briefly introduces our theoretical model. The approach follows the recent literature on strategic network interactions (e.g., Ballester et al., 2006; Bramoullé et al., 2014). We are forced to leave most details unexplained and only provide a brief sketch of the main hypotheses and the underlying intuitions due to space limitations. All details will be studied during the project execution and discussed in the end-of-project report.

Given a set of m_g individuals in network g, each individual i can opt either to adopt or not to adopt an action, i.e., $a_i \in \{0, 1\}$.[1] We denote $a_{-i} = (a_j)_{j \neq i}$ as the vector of the actions of others. The incurred action cost is denoted by $c > 0$. The payoff is determined by an individual's action and the actions of his/her friends. We make standard assumptions regarding the utility function. Additionally, we assume that people care about their friends, $j \in N_i(g)$, where $N_i(g)$ denotes the set of neighbors (friends) of individual i in network g. We denote $u_i(.)$ and $U_i(.)$ as the monetary payoff of individual i and his/her utility function, respectively. Formally, $\frac{\partial U_i(.)}{\partial u_j(.)} > 0$ for $j \in \{i\} \cup N_i(g)$; $\frac{\partial U_i(.)}{\partial u_j(.)} = 0$ otherwise.[2] As a result of this assumption, in equilibrium individuals consider the effect of their actions on their friends. Let γ measure the externality of an action on the payoff of friends; $\gamma > 0$ (< 0) if the externalities are positive (negative).[3] An individual takes an action when $U_i(a_i = 1, a_{-i}; \gamma, c, g) \geq U_i(a_i = 0, a_{-i}; \gamma, c, g)$.[4]

The goal of the theoretical analysis is to provide a full characterization of the set of Nash equilibria. Nevertheless, literature on network games indicates that these games typically exhibit large multiplicity of equilibria. Two common solutions are available. Scholars either add additional assumptions regarding the game structure to ensure unicity (e.g., Galeotti et al., 2010) or provide additional restrictions on the equilibrium concept (Bramoullé and Kranton, 2007). The existence of this problem depends, among other things, on the network structure at least for several classes of games (Ballester et al., 2006, Bramoullé et al., 2014). Therefore, given the empirical motivation of our model, the imposition of additional restrictions will be evaluated during the theoretical analysis and, in part, in the function of the network architectures observed in the data.

Previous studies have indicated that the connectivity of an individual affects his/her decision when externalities are present because having many or few friends directly affects

[1] The discrete strategy space is assumed for simplicity. We plan to extend the analysis to continuous spaces during the project execution.

[2] Examples of such non-selfish preferences include altruism, efficiency seeking, or some forms of inequity aversion among others.

[3] Two types of externalities exist: (i) positive externalities determined by the network because people care about their friends, independent of the behavior, and (ii) a second type (arising from the behavior) that represents the idea of externalities in the standard strategic interaction framework.

[4] See Bramoullé et al., (2014) for examples of particular games fitting into this framework.

the incentives for taking an action. Nevertheless, minimal attention has been given to the role of clustering. Our preliminary analysis suggests that there exist an equilibrium, in which the clustering coefficient plays the following role: fully rational agents consider both the *direct* effect of their decision on their neighbors, and the *indirect* effect of their decision on neighbors through their common friends.[5] As a result, individuals with (the same network degree but) different cohesiveness of their neighborhoods are expected to differ systematically in the adoption rates of norms, behaviors, or products if they care about the well-being of their friends, but their behaviors generate an externality on them. Whether the overall effect is positive or negative and whether the effect depends on the nature of the diffused phenomenon are empirical questions, but we can extend existing theoretical approaches by the following conjecture: for agent i with $|N_i(g)| = |N_i(g')|$, the probability of adopting the action is, *ceteris paribus*, higher if an individual has higher clustering coefficient under positive externalities. Formally, $P(a_i = 1; a_{-i}, \gamma, c, g) > P(a_i = 1; a_{-i}, \gamma, c, g')$ if $\gamma > 0$ and $|N_i(g) \cap N_{-i}(g)| > |N_i(g') \cap N_{-i}(g')|$. The intuition behind this statement is the following: under positive externalities, an individual with the same degree in two networks g and g' is likely to adopt an action in equilibrium if his/her clustering is high, because he/she directly considers the effect he/she has on his/her neighbors through his/her action and indirectly through common friends. If $\gamma < 0$, the contrary applies: an individual tends to adopt an action less likely if he/she has high clustering. At the network level, we thus hypothesize that a more clustered network, *ceteris paribus*, creates a higher local strategic interdependence and, consequently, mutual reinforcement of behaviors. This condition ultimately leads us to state the following testable hypothesis for our empirical applications:

Hypothesis: *The estimated network effect, measured by λ, increases (decreases) in the average clustering coefficient of a network when $\gamma > 0$ (< 0).*

In practice, we do not measure γ directly, but we can test whether more clustered networks lead to higher estimated λ in the diffusion of, for example, schooling outcomes compared with less clustered ones, and whether the reverse occurs for, say, smoking or criminal behavior.[6]

2.2 Empirical Study

2.2.1 Social Interaction Model with Random Endogenous Peer Effect

This subsection presents an empirical framework designed to test the hypothesis above. We first introduce the spatial autoregressive (SAR) model, which is used in this project as the workhorse to study the interactions between individuals within a network community:

$$Y_g = \lambda W_g Y_g + X_g \beta_1 + W_g X_g \beta_2 + \ell_g \alpha_g + \epsilon_g, \quad \epsilon_g \sim \mathcal{N}_{m_g}(0, \sigma_\epsilon^2 I_{m_g}), \tag{1}$$

where $g = 1, \cdots, G$ denotes network communities; $Y_g = (y_{1,g}, y_{2,g}, ..., y_{m_g,g})'$ is the vector capturing the dependent variable of m_g individuals in group g, W_g is the spatial weight (network adjacency) matrix, where $w_{ij,g} = 1$ if individual i links to individual j and zero otherwise; X_g is the exogenous characteristics; ℓ_g is the m_g-dimensional vector of ones; α_g is the unobserved group effect; and ϵ_g is the error term with a multivariate normal distribution of dimension m_g. Following Manski's (1993) terminology, coefficient

[5] The latter can be understood as a network multiplier effect.

[6] We are aware that finding "identical" networks that only differ in clustering is difficult. The empirical approach deals with this problem by controlling for other network features.

λ captures the endogenous peer effect; β_1 and β_2 capture own and contextual exogenous effects, respectively; and α_g captures the correlated effect.

A key innovation of this project is to extend the SAR model of Eq. (1) to capture global network structural effects. Motivated by our theoretical hypotheses, we model network structural effects through the random coefficient specification on the endogenous effect.[7] We let S_g denote an R_1-dimensional measure of network structures (we mainly consider the clustering coefficient, but it can also include the average degree, eigenvalues, and others) with respect to group g. We denote F_g as an R_2-dimensional vector of group specific variables other than network structural measures (e.g., size, location, economic conditions, etc.) that help in explaining heterogeneity across groups. We change λ in Eq. (1) to λ_g, which varies with S_g and F_g through a linear model with normal disturbance, as follows:

$$\lambda_g = \zeta_1 + S_g\zeta_2 + F_g\zeta_3 + e_g, \quad e_g \sim \mathscr{N}(0, \omega^2), \quad g = 1, \ldots, G. \tag{2}$$

We also consider alternatives of the generalized linear model (GLM) specification with different link functions.

$$\lambda_g = \mathscr{G}^{-1}(\zeta_1 + S_g\zeta_2 + F_g\zeta_3) + e_g, \quad e_g \sim \mathscr{N}(0, \omega^2), \quad g = 1, \ldots, G \tag{3}$$

where $\mathscr{G}(\cdot)$ is a transformation that maps from $(-1, 1)$ to the real line. The reason for the GLM specification is the fact that the effective domain of the peer effect in the SAR model lies within $(-1, 1)$. Thus, the link function prevents the linear predictor $\zeta_1 + S_g\zeta_2 + F_g\zeta_3$ from exceeding bounds. We consider the tangent function, $\tan\left(\frac{\pi x}{2}\right)$, and the inverse hyperbolic tangent function, $\frac{1}{2}\log\left(\frac{1+x}{1-x}\right)$, for link function $\mathscr{G}(\cdot)$.

We have two remarks. First, this model extension creates heterogeneity in endogenous peer effects across network communities. We perform a likelihood ratio test under the hypothesis that all λ_g's are the same in the original SAR model of Eq. (1). By rejecting this hypothesis, we can claim that this extension provides improved model specification. Second, our random coefficient specification on the endogenous peer effect is related to the random coefficient model in the simultaneous equation system (Kelejian, 1974; Hsiao and Pesaran, 2008). Individual-specific random coefficients are not pointly identified in the simultaneous equation system. However, our model considers group-specific random coefficients that are at the group level and thus identified under regularity conditions.

2.2.2 Endogeneity of Network Links

In the SAR model of Eq. (1), although the group-level correlated effect α_g can handle the confounding factors at the group level, studies have shown that the estimate of the endogenous peer effect still suffers from endogeneity due to uncontrolled individual correlated effects (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Jackson, 2014). Therefore, in this project, we adopt the selection correction approach based on Hsieh and Lee (2016), in which the outcome equation (1) is jointly estimated with the

[7] One can potentially specify network structural measures through correlated effects in Eq. (1). However, the cost of doing so is that α_g can only be interpreted as a random effect instead of a fixed effect. In social interaction studies (e.g., Clark and Loheac, 2007; Fletcher, 2010), the "fixed" group effect is preferred because it helps control other unobserved environmental factors that are not included in X_g, as well as the endogeneity of group formation. A random effect assumption is inappropriate because it does not allow α_g to be correlated with other regressors in the model.

link formation equation. With information on network links, we can explore unobserved individual heterogeneity through latent factors and use them to correct the potential endogeneity of the network. Specifically, the error term ϵ_g in (1) is further decomposed into the latent factor part and the pure error. Thus, our outcome equation can be rewritten as

$$Y_g = \lambda_g W_g Y_g + X_g \beta_1 + W_g X_g \beta_2 + \ell_g \alpha_g + Z_g \delta_1 + W_g Z_g \delta_2 + u_g, \tag{4}$$

where $u_g \sim \mathcal{N}_{m_g}(0, \sigma_u^2 I_{m_g})$, $Z_g = (z'_{1,g}, ..., z'_{m_g,g})'$ denotes the collection of unobserved variables (latent factors) in group g, $z_{i,g} = (z_{i1,g}, ..., z_{i\bar{d},g})'$ is the $\bar{d}$-dimensional vector of latent factors of individual i, and u_g and Z_g are assumed to be independent. Eq. (4) generalizes the model in Hsieh and Lee (2016) in which ϵ_g and Z_g are assumed to be jointly normal, and u_g is defined as the disturbance of ϵ_g conditioned on Z_g. Under the specification in (4), we replace the strong normality assumption with the conditional mean of ϵ_g given Z_g being linear, i.e., $\mathrm{E}(\epsilon_g|Z_g) = Z_g\delta_1 + W_g Z_g \delta_2$, where Z_g captures the unobserved individual correlated effect and $W_g Z_g$ captures the unobserved contextual effect (Fruehwirth, 2014).

In Eq. (4), identification of Z_g originates from the following link formation model,

$$\begin{aligned} P(w_{ij,g}|c_{ij,g}, z_{i,g}, z_{j,g}) &= \left(\frac{\exp(\psi_{ij,g})}{1+\exp(\psi_{ij,g})}\right)^{w_{ij,g}} \left(\frac{1}{1+\exp(\psi_{ij,g})}\right)^{1-w_{ij,g}}, \\ \psi_{ij,g} &= c_{ij,g}\gamma_0 + \eta_1|z_{i1,g} - z_{j1,g}| + \cdots + \eta_{\bar{d}}|z_{i\bar{d},g} - z_{j\bar{d},g}|, \end{aligned} \tag{5}$$

where $c_{ij,g}$ is the $\bar{q}$-dimensional vector of observed dyad-specific variables, and $|z_{id,g} - z_{jd,g}|$'s are used to capture how the distances of unobservable characteristics between individuals i and j affect their friendship decisions. The latent factor $z_{i,g}$ influences the friendship formation (5) and outcome equation (4), and is assumed to follow a normal distribution, namely,

$$p(z_{i,g}|\mu_{z,g}, \sigma^2_{z,g}) \sim \mathcal{N}_{\bar{d}}(\mu_{z,g}, \sigma^2_{z,g} I_{\bar{d}}), \tag{6}$$

where $\mu_{z,g} = (\mu_{z1,g}, \ldots, \mu_{z\bar{d},g})'$ and $\sigma^2_{z,g}$ is a scalar. Therefore, Eqs (4), (5), and (6), together with the random coefficient specification of Eq. (2) (or Eq. (3)), complete the construction of our empirical model.

2.2.3 Estimation

We apply the Bayesian approach to estimate our model. The Bayesian approach has long been considered a useful alternative to the frequentist approach for estimating spatial models. As pointed out by Anselin (1988, pp. 90), the likelihood part of the posterior density *swamps* the influence of the diffuse prior, and the Bayesian estimate of SAR is comparable to the maximum likelihood estimate (see also Berger (1985) for some theoretical justification). Moreover, in contrast to the asymptotic maximum likelihood properties, the Bayesian approach in SAR allows for a precise assessment of the finite sample probability (see Anselin (1982)). With regard to the selection-correction approach in network endogeneity, Hsieh and Lee (2016) showed the usefulness of the Bayesian method in handling complex latent space structures. In this project, the random endogenous peer effect brings further complexities to SAR models. We consider the Markov Chain Monte Carlo (MCMC) algorithm to facilitate the posterior inferences of the proposed model.

2.2.4 Network Data Sets

AddHealth The Add Health longitudinal survey (conducted between 1994 and 2008) provides a nationally representative sample of adolescents in grade 7 through 12 from 132 schools (Harris, 2009). AddHealth represents detailed information on respondents' demographic backgrounds, academic performance, health-related behaviors, and most importantly, friendship networks constructed from the respondents' nominations. We regard each school as a natural boundary for forming a network community (total of 132 networks) and study how friendship network structures in school affect students' academic performance and other behaviors, such as smoking, club participation, and misbehavior.

Indian rural villages The Indian rural village microfinance survey was conducted in 2006, under the microfinance program in rural southern Karnataka, India (Banerjee et al., 2013). The data consist of responses to a baseline survey from a list of 75 villages and the microfinance participation outcome from 43 villages on the list. Information on households (e.g., house conditions such as roof type, number of rooms, latrine type, electricity provider, and whether anyone in the household is a microfinance client) and individuals (e.g., age, caste, religion, language, and occupation) is available. We study household participation behaviors through social connections between households, which exist as long as household members report a relationship with members from another household.

Indonesian villages These data were obtained from a randomized field experiment on targeted cash transfer programs in Indonesia (Alatas et al., 2016). The experiment took place between November 2008 and March 2009. The purpose of the experiment was to compare the accuracy of different targeting methods in identifying beneficiaries for transfer programs. The data contained information from 631 independent sampled hamlets on households' social networks and their perceived rank on the poverty level of other households in the hamlet. The network connections represent familiar relationships and common memberships in social groups within the hamlet. The perceived rank is used for comparison with the "actual" rank (based on actual per-capita expenditure) to form the error rate. On the basis of these data, we can study how the information regarding households' economic status is moderated by the network structure within the hamlet.

China Telecom China telecommunication network data are obtained from a major Chinese mobile carrier, which provides us access to its entire customer base of 1.36 million users in two medium-sized cities in Sichuan, China (Hu et al., 2014). We know the age, gender, location, phone model, phone usage (including number and duration of phone calls, messaging activity, and Internet access) of each customer. The Call Detailed Record allows us to construct links between users (a link is placed between two individuals when they have called or texted each other within the same month). The individual dependent variable that we wish to study is the adoption of a new phone model. We chose the sample period from November 2012 to May 2013, which matched the release of a new phone model (Samsung Note II). To construct network samples, we first obtain a sample of 26,000 customers by a snowball sampling initiated from Samsung Note II adopters. Then we follow the T-CLAP network partitioning algorithm used in de Matos et al. (2014) to obtain a total of 243 network communities with 100 customers in each community.

References

Acemoglu, Daron, Vasco M Carvalho, , A Ozdaglar, A Tahbaz-Salehi. 2012. The network origins of aggregate fluctuations. *Econometrica* **80**(5) 1977–2016.

Alatas, Vivi, Abhijit Banerjee, Arun G Chandrasekhar, Rema Hanna, Benjamin A Olken. 2016. Network structure and the aggregation of information: Theory and evidence from indonesia. *The American Economic Review* **106**(7) 1663–1704.

Alderman, Harold, Trina Haque. 2006. Countercyclical safety nets for the poor and vulnerable. *Food Policy* **31**(4) 372–383.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. NATO Asi Series. Series E, Applied Sciences, Springer.

Anselin, Luc. 1982. A note on small sample properties of estimators in a first-order spatial autoregressive model. *Environment and Planning A* **14**(8) 1023–30.

Ballester, Coralio, Antoni Calvó-Armengol, Yves Zenou. 2006. Who's who in networks. wanted: The key player. *Econometrica* **74**(5) 1403–1417.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, Matthew O Jackson. 2013. The diffusion of microfinance. *Science* **341**(6144) 1236498.

Berger, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics, Springer.

Bollobás, Béla, Christian Borgs, Jennifer Chayes, Oliver Riordan. 2010. Percolation on dense graph sequences. *The Annals of Probability* **38**(1) 150–183.

Bramoullé, Y, R Kranton. 2007. Public goods in networks. *Journal of Economic Theory* **135**(1) 478–494.

Bramoullé, Yann, Habiba Djebbari, Bernard Fortin. 2009. Identification of peer effects through social networks. *Journal of Econometrics* **150**(1) 41–55.

Bramoullé, Yann, Rachel Kranton, Martin D'amours. 2014. Strategic interaction and networks. *The American Economic Review* **104**(3) 898–930.

Burt, Ronald S. 1992. *Structural holes: The social structure of competition*. Harvard university press.

Burt, Ronald S. 2005. *Brokerage and closure: An introduction to social capital*. OUP Oxford.

Calvó-Armengol, Antoni, Eleonora Patacchini, Yves Zenou. 2009. Peer effects and social networks in education. *The Review of Economic Studies* **76**(4) 1239–1267.

Centola, Damon. 2010. The spread of behavior in an online social network experiment. *Science* **329**(5996) 1194–1197.

Centola, Damon. 2011. An experimental study of homophily in the adoption of health behavior. *Science* **334**(6060) 1269–1272.

Clark, Andrew E, Youenn Loheac. 2007. "it wasn't me, it was them!" social influence in risky behavior by adolescents. *Journal of Health Economics* **26**(4) 763–784.

Coleman, James S. 1988. Social capital in the creation of human capital. *American Journal of Sociology* S95–S120.

Conley, Timothy G, Christopher R Udry. 2010. Learning about a new technology: Pineapple in ghana. *The American Economic Review* **100**(1) 35–69.

de Matos, Miguel Godinho, Pedro Ferreira, David Krackhardt. 2014. Peer influence in the diffusion of iphone 3g over a large social nework. *MIS Quarterly* **38**(4) 1103–1133.

Elliott, Matthew, Benjamin Golub, Matthew O Jackson. 2014. Financial networks and contagion. *The American Economic Review* **104**(10) 3115–3153.

Fletcher, J. M. 2010. Social interactions and smoking: evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics* **19** 466–484.

Fruehwirth, Jane Cooley. 2014. Can achievement peer effect estimates inform policy? a view from inside the black box. *Review of Economics and Statistics* **96**(3) 514–523.

Galasso, Emanuela, Martin Ravallion. 2005. Decentralized targeting of an antipoverty program. *Journal of Public Economics* **89**(4) 705–727.

Galeotti, A., S. Goyal, M.O. Jackson, F. Vega-Redondo, L. Yariv. 2010. Network games. *Review of Economic Studies* **77**(1) 218–244.

Gargiulo, Martin, Mario Benassi. 2000. Trapped in your own net? network cohesion, structural holes, and the adaptation of social capital. *Organization Science* **11**(2) 183–196.

Gaviria, Alejandro, Steven Raphael. 2001. School-based peer effects and juvenile behavior. *Review of Economics and Statistics* **83**(2) 257–268.

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

Goldsmith-Pinkham, Paul, Guido W Imbens. 2013. Social networks and the identification of peer effects. *Journal of Business & Economic Statistics* **31**(3) 253–264.

Granovetter, Mark. 1985. Economic action and social structure: the problem of embeddedness. *American Journal of Sociology* 481–510.

Harris, Kathleen Mullan. 2009. The national longitudinal study of adolescent to adult health (add health), waves i & ii, 1994–1996; wave iii, 2001–2002; wave iv, 2007-2009 [machine-readable data file and documentation] .

Hoxby, Caroline. 2000. Peer effects in the classroom: Learning from gender and race variation. Tech. rep., National Bureau of Economic Research.

Hsiao, Cheng, M Hashem Pesaran. 2008. Random coefficient models. *The Econometrics of Panel Data*. Springer, 185–213.

Hsieh, Chih-Sheng, Lung Fei Lee. 2016. A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics* **31**(2) 301–319.

Hsieh, Chih-Sheng, Xu Lin. 2017. Gender and racial peer effects with endogenous network formation. *Regional Science and Urban Economics* forthcoming.

Hsieh, Chih-Sheng, Hans van Kippersluis. 2017. Smoking initiation: Peers and personality. *Quantative Economics* forthcoming.

Hu, Mandy, Chih-Sheng Hsieh, Jianmin Jamie Jia. 2014. Predicting social influence based on dynamic network structures. *Available at SSRN 2542032* .

Jackson, Matthew O. 2008. Average distance, diameter, and clustering in social networks with homophily. *International Workshop on Internet and Network Economics*. Springer, 4–11.

Jackson, Matthew O. 2014. Networks in the understanding of economic behaviors. *The Journal of Economic Perspectives* 3–22.

Jackson, Matthew O, Brian W Rogers. 2007. Meeting strangers and friends of friends: How random are social networks? *American Economic Review* **97**(3) 890–915.

Jackson, Matthew O, Leeat Yariv. 2007. Diffusion of behavior and equilibrium properties in network games. *The American Economic Review* **97**(2) 92–98.

Karlan, Dean, Markus Mobius, Tanya Rosenblat, Adam Szeidl. 2009. Trust and social collateral. *The Quarterly Journal of Economics* **124**(3) 1307–1361.

Kelejian, Harry H. 1974. Random parameters in a simultaneous equation framework: identification and estimation. *Econometrica: Journal of the Econometric Society* 517–527.

Lin, Xu. 2010. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics* **28**(4) 825–860.

Manski, Charles F. 1993. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* **60**(3) 531–542.

Masten, Matthew. 2014. Random coefficients on endogenous variables in simultaneous equations models. Tech. rep., cemmap working paper, Centre for Microdata Methods and Practice.

Patacchini, Eleonora, Edoardo Rainone, Yves Zenou. 2017. Heterogeneous peer effects in education. *Journal of Economic Behavior & Organization* **134** 190–227.

Podolny, Joel M, James N Baron. 1997. Resources and relationships: Social networks and mobility in the workplace. *American Sociological Review* 673–693.

Topa, Giorgio. 2001. Social interactions, local spillovers and unemployment. *The Review of Economic Studies* **68**(2) 261–295.

Watts, Duncan J, Steven H Strogatz. 1998. Collective dynamics of'small-world networks. *Nature* **393**(6684) 440.