Estimating Local Network Externalities

Margherita Comola*

February 18, 2016

Abstract

This paper illustrates a procedure to estimate externalities from indirect connections (so-called network externalities) using network data. This structural approach is suitable for static games of endogenous network formation with partial information, and relies on the equilibrium conditions of pairwise stability (Jackson and Wolinsky, 1996). Operationally, it consists in a two-step estimator which addresses omitted variable endogeneity. When the estimation protocol is applied to risk-sharing data within a Tanzanian village, results indicate that indirect connections matter. Network externalities are found to be negative, which can be interpreted as competition over scarce resources.

Keywords: Network Externalities; Pairwise Stability; Risk-sharing

JEL codes: C45; D85; O12

*Paris School of Economics and Université Paris 1 Panthéon-Sorbonne. Email: margherita.comola@psemail.edu

1 Introduction

From its very first steps network theory has claimed that the formation and severance of links depend strategically on the entire community graph (Jackson and Wolisky, 1996; Bala and Goyal, 2000; Jackson, 2004). However, in recent years testing whether network architecture predicts link formation has proved to be a formidable task (Miyauchi, 2014; Sheng, 2014; De Paula Richards-Shubik and Tamer, 2015; Mele, 2015). This paper proposes a simple estimation protocol for static models of endogenous network formation with externalities from indirect connections (so-called network externalities). It assumes that the observed network is in a pairwise stable equilibrium (Jackson and Wolinsky, 1996) to derive testable predictions about agent's expected utility from indirect connections.

The validity of the estimation procedure crucially relies on two features of the underlying network formation game: partial information, and pairwise stability. On the one hand, agents play a static game with partial information where they form links simultaneously on the basis of their beliefs. Assuming that these beliefs are built on observable characteristics, the estimation strategy boils down to a simple two-step estimator where the first stage fits a conditional probability representing agents' beliefs about the emerging network, and in the second stage the predicted network replaces the observed network in the computation of the externalities. As long as there is a covariate which serves as a valid instrument, this two-step approach addresses omitted variable endogeneity.[1] On the other hand, the equilibrium concept of pairwise stability proves crucial for estimation as it allows to circumvent the issue of dependence between linking decisions due to strategic interactions. Intuitively, if agents are playing a myopic best response strategy in the context of pairwise stability, one can estimate preferences over many small 'local' perturbations of the existing network architecture which represent unilateral deviations from the *status quo*. This paper's contribution to applied network economics is twofold. First, it shows how partial observability (Poirier, 1980) can be used to estimate a large class of bilateral link formation games. Second, it proposes a empirical tool for applied economists who want to identify preferences over network externalities. A large theoretical effort has been devoted to modeling externalities in different network formation contexts (Jackson, 2008). However, evidence based on experimental and observational data lags behind, and little information is available on the value of indirect connections and their decay rate for most real-life situations.[2]

[1] A two-step approach is also taken by Konig, Liu and Zenou (2014) and Leung (2015). Kelejian and Piras (2014) propose a similar solution in the context of spatial regressions.

[2] The study of cross-firm collaborative networks suggests that information flows are insignificant for indirect neighbors (Breschi and Lissoni, 2005; Singh, 2005). On the other hand, experimental evidence with dictator games shows that further-away connections are relevant and decay follows an inverse distance law (Goree *at al.*, 2010).

This paper contributes to filling this gap by explaining how local perturbations of the equilibrium can be exploited to recover network externalities.

The econometrics of network formation is at the frontier of the applied network literature (see Advani and Malde, 2014; Chandrasekhar, 2015; Graham, 2015 for recent reviews). This paper deals with the common, yet problematic setting where the researcher observes one single network at one single period, and wants to include network covariates in the objective function of agents. In this scenario the structural equation can have multiple solutions (Bresnahan and Reiss, 1991; Tamer, 2003), and the calculation may become intractable due to the combinatory complexity of networks. One solution is provided by the exponential random graph models where a dynamic meeting protocol acts as equilibrium selection mechanism (Christakis *et al.*, 2010; Badev, 2013; Chandrasekhar and Jackson, 2014; Hsieh an Lee, 2015; Koenig, 2015; Mele, 2015). These models provide a solid micro-foundation for link formation, but need to be solved with Markov Chain Monte Carlo sampling techniques which are computationally difficult and often inconsistent.[3] An alternative solution is to condition on certain classes of models that replicate some empirical features of the network, or to limit the degree to which other players can affect one's utility.[4] This paper neither tries to elicit global preferences (*i.e.* preferences over the entire network topology) as exponential random graph models do, nor it attempts to replicate observed topological patterns. The estimation strategy proposed here only extrapolates information from local perturbations of the existing equilibrium in order to gather information on the utility from indirect connections. To do so I focus on static network formation with network externalities, and use the solution concept of pairwise stability. Other recent studies rely on the same framework but assume full information and achieve set identification (De Paula Richards-Shubik and Tamer, 2014; Miyauchi, 2014; Sheng, 2014). The closest contribution to this paper is Leung (2015), who assumes partial information and achieves point identification. He models a simultaneous game where agents take linking decisions on the basis of their beliefs on the structure of the emerging network. The two-step estimator he proposes is robust to the presence of multiple equilibria, and circumvents the dependence between actions due to strategic interaction by conditioning on observables (De Paula and Tang, 2012). In this paper I follow the same inferential approach but I differ from Leung (2015) in two important aspects: first, I consider undirected links and use the equilibrium notion of pairwise stability (while he uses directed links and

[3]See Bhamidi Bresler and Sly (2008), Shalizi and Rinaldo (2012), Chatterjee and Diaconis (2013), Chandrasekhar ad Jackson (2014).

[4]Some papers identify structural parameters by the rate at which various sub-graphs are observed in the overall network (Chandrasekhar and Jackson, 2015) or by aggregating individuals into 'types' and assuming that agents have preferences only over the type of their partners (De Paula, Richards-Shubik and Tamer, 2014). On a similar line, Boucher and Mourifie (2013) study a scenario where individual preferences display weak homophyly.

a Bayesian solution concept). Second, I focus on a procedure to consistently estimate the magnitude and decay of externalities from indirect connections.

An illustration on risk-sharing data is presented. Lacking access to formal insurance, most households in developing countries are forced to rely on risk-sharing arrangements in face of idiosyncratic shocks such as health-related expenses, funerals and court trials. These arrangements are based on pre-existing interpersonal links and often take the form of informal loans or gifts.[5] The data I use contain information on self-declared risk-sharing links in a Tanzanian village named Nyakatoke. During 2000, all adult individuals of Nyakatoke were asked *"Can you give a list of people [...], who you can personally rely on for help and/or that can rely on you for help in cash, kind or labor?"*. This piece of information is used to draw the village network, and to investigate the role of local community architecture. Specifically, I test whether agents choose risk-sharing partners on the basis of their individual characteristics only, or also the characteristics of indirect connections (*e.g.* friends of friends) matter. Risk-sharing arrangements are an intriguing example that may combines positive and negative network externalities: friends of friends are beneficial if they broaden social interactions, but detrimental if there is competition for resources. Results suggest that Nyakatoke villagers do evaluate potential partners' connections, and that the negative component seems to prevail, which can be interpreted in light of competition for scarce economic and/or social resources.

The paper is organized as follows. Section 2 introduces the theoretical setting. Section 3 illustrates the estimation strategy. In Section 4 the consistency of the estimated parameters is demonstrated with a simulation exercise. Sections 5 and 6 illustrate the data and the variables for the empirical application respectively. Section 7 presents the results, while Section 8 concludes. The Appendix describes the computational steps to replicate the estimation protocol of Section 3.

2 Theoretical setting

The game-theoretical literature on strategic network formation has flourished in the last two decades (Jackson, 2008). While some economists have approached network formation from a non-cooperative perspective (Bala and Goyal, 2000; Galeotti, Goyal and Kamphorst, 2006), the majority of them have focused on stable networks, where links are formed at the discretion of self-interested agents whose utility depends on the overall network architecture. This paper builds on the latter literature, and it is inspired by the two popular models of network formation discussed by Jackson and Wolinsky (1996): the connection model and the coauthor model. In the connection

[5]See Townsed (1994), Udry (1994), Fafchamps and Lund (2003), De Weerdt and Dercon (2006).

model links represent social relationships and network externalities are positive, while in the co-author model agents are researchers who collaborate in common projects and network externalities are negative. However, while for the sake of theoretical tractability Jackson and Wolinsky (1996) solve these models for homogeneous agents, my empirical approach allows for full heterogeneity.[6]

Let $N(g) = \{1, ..., N\}$ be a set of players connected in a given network g. The network is denoted by the adjacency matrix $g = [g_{ij}]$: for any unique pair of players (dyad) $\{ij\}$ such that $i, j \in N(g)$ and $i \neq j$, $g_{ij} = 1$ indicates that i and j are linked under the network g, and $g_{ij} = 0$ that they are not. The network matrix is symmetric ($g_{ij} = g_{ji}$). By a standard abuse of notation, let $g + ij$ denote the network g with the link g_{ij}, that is, with $g_{ij} = 1$ and, analogously, let $g - ij$ denote the network g without the link g_{ij}, that is, with $g_{ij} = 0$.

For a dyad $\{ij\}$, let the symmetric m-dimensional vector X_{ij} define the relational attributes of the two agents. Let t_{ij_g} define the geodesic distance in the network g, that is, the number of links in the shortest path between i and j. Geodesic distance is integer and symmetric ($t_{ij_g} = t_{ji_g}$) and is set to infinity if there is no path between i and j. The matrix $\tau_g = [t_{ij_g}]$ denotes the distance matrix induced by g. The utility of agent i from network g is given by

$$u_i(g) = \sum_{j \in N(g)} a(t_{ij_g})\, y_j + \sum_{j : t_{ij_g} = 1} \beta X_{ij}, \tag{1}$$

where y_j is j's attribute representing his desirability as a partner (say wealth), $\beta \in R^m$ are the weights attached to the relational attributes of direct connections only, and $a(.)$ are the weights attached to the wealth of all agents by their geodesic distance: these weights can be positive or negative in sign, and for $t_{ij_g} > 1$ represent network externalities in the dimension of wealth. Extending network externalities to other dimensions is straightforward.

In this paper I adopt the equilibrium notion of pairwise stability (Jackson and Wolinsky, 1996), which states that link formation requires the consent of both parties involved while severance can be done unilaterally.[7] We say that network g is pairwise stable if

[6] Vandenbossche and Demuynck (2012) provide a theoretical characterization of the unique stable equilibrium in a model with heterogeneous agents, but they rule out network externalities from indirect connections.

[7] Pairwise stability was the first equilibrium concept proposed for networks - see Dutta and Mutuswami (1997) and Jackson and Van Den Nouweland (2005) for early refinements.

$$
\begin{aligned}
&(i) \quad \forall g_{ij}=1,\ u_i(g+ij) \geq u_i(g-ij)\ \&\ u_j(g+ij) \geq u_j(g-ij) \\
&(ii) \quad \forall g_{ij}=0,\ u_i(g+ij) > u_i(g-ij) \Rightarrow u_j(g-ij) > u_j(g+ij)\,.
\end{aligned} \tag{2}
$$

That is, a network is pairwise stable if all links which are *ceteris paribus* profitable are formed, and no player would benefit from severing an existing link. Pairwise stable networks do not allow for multiple or simultaneous deviations, which is crucial for the estimation strategy illustrated in the next section.

3 Estimation strategy

In what follows I introduce the building blocks of the estimation strategy. Section 3.1 shows how to decompose the agent's utility in order to disentangle network externalities. Section 3.2 explains how the equilibrium conditions of pairwise stability prove useful for identification. Section 3.3 shows how common beliefs about the emerging network configuration can be modeled *via* a two-step estimator analogous to two-stage least squares (2SLS), with an additional element of stochastic variability that needs to be taken into account for inference purposes. The computational details are left to the Appendix.

3.1 Network externalities

For a given network configuration g, from Equation (1) we can write

$$
\begin{aligned}
u_i(g+ij) - u_i(g-ij) &= \sum_{k \in N(g+ij)} a(t_{ik_{g+ij}}) y_k - \sum_{k \in N(g-ij)} a(t_{ik_{g-ij}}) y_k + \beta X_{ij} && (3) \\
&= a(1) y_j + a(2) \Lambda_{2_{ij}} + a(3) \Lambda_{3_{ij}} + ... + \beta X_{ij} && (4) \\
&= \sum_{\phi=1}^{N-2} a(\phi) \Lambda_{\phi_{ij}} + \beta X_{ij}\,, && (5)
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda_{\phi_{ij}} &= \lambda_{i,\phi}^{+ij} - \lambda_{i,\phi}^{-ij} \\
\lambda_{i,\phi}^{+ij} &= \sum_k y_{\substack{k \in N(g+ij) \\ t_{ik_{g+ij}}=\phi}} \\
\lambda_{i,\phi}^{-ij} &= \sum_k y_{\substack{k \in N(g-ij) \\ t_{ik_{g-ij}}=\phi}}\,.
\end{aligned}
$$

Equation (5) shows how the total utility in terms of direct and indirect partners' discounted wealth can be decomposed by geodesic distance. $\Lambda_{\phi_{ij}}$ represents the net gain in terms of wealth of ϕ-step-away agents that i gets if he links with j and the rest of g remains the same. Note that $\Lambda_{1_{ij}}$ boils down to the partner's wealth y_j, and that ϕ is capped strictly below $N-1$ which is the maximum finite distance attainable in a network of N players.[8] In what follows the Λ_ϕ terms are referred to as 'structural' regressors (as opposite to 'standard' regressors) because they are statistics for the architecture of the network g. Note that the patterns and sign of the structural regressors cannot be predicted beforehand, since the formation of one link changes the entire network architecture.[9] The associated coefficients $a(\phi)$ for $\phi > 1$ capture the network externalities in the dimension of wealth and can be interpreted as revealed preferences over partners' network position. The procedure to compute these structural regressors is described in the Appendix, Section A.1.

3.2 Equilibrium conditions

The local equilibrium conditions of pairwise stability prove crucial for identification. For a given network configuration g, we can define two latent binary response variables w_{ij} and w_{ji} which represent the willingness to link of i and j respectively. In this context link formation under pairwise stability reduces to estimating the two-equation model

$$\begin{aligned} P(g_{ij}=1) &= P\left(w_{ij}=1 \;\&\; w_{ji}=1\right) \\ P(g_{ij}=0) &= 1-P\left(w_{ij}=1 \;\&\; w_{ji}=1\right), \end{aligned} \tag{6}$$

[8]Since a new link does not change the set of reachable partners, from the individual perspective each wealth gain at a certain distance is reflected in a loss at another distance. If all players are indirectly connected with finite geodesic distance, it is sufficient to cap ϕ strictly below the maximum geodesic distance Φ because all Λ_ϕ terms are perfectly collinear by construction: $\sum_{\phi=1}^{\Phi} \Lambda_{\phi_{ij}} = 0$. Even if the network has multiple components or isolated agents, a finite distance can never exceed $N-1$. The issue has very little practical relevance, because if the network is sufficiently dense the Λ_ϕ terms become highly collinear after a few steps anyway.

[9]For example, if $t_{ij_{g-ij}} = 2$ and the link ij is formed, from i's perspective the direct gain of y_j is reflected in a two-step-away loss, so that $\Lambda_{2_{ij}}$ is *ceteris paribus* negative $(-y_j)$, but the formation of the link is likely to induce other changes in the overall architecture which may counterbalance the loss (*e.g.* a third player who used to be further away from i is now two steps away).

where

$$w_{ij} = \begin{cases} 1 & \sum_{\phi=1}^{N-2} a(\phi)\Lambda_{\phi_{ij}} + \beta X_{ij} > \epsilon_{ij} \\ 0 & \text{otherwise} \end{cases}$$

$$w_{ji} \equiv \begin{cases} 1 & \sum_{\phi=1}^{N-2} a(\phi)\Lambda_{\phi_{ji}} + \beta X_{ji} > \epsilon_{ji} \\ 0 & \text{otherwise,} \end{cases}$$

and the coefficients $a(.)$ and β are constrained to be the same across the two equations. This can be estimated as a bivariate probit with partial observability (Poirier, 1980; Abowd and Farber, 1982; Farber, 1983). Partial observability occurs when a positive outcome for one response variable is only observed if the other response variable is also positive. That is, when $g_{ij} = 1$ we know that $w_{ij} = w_{ji} = 1$. *Viceversa*, when $g_{ij} = 0$ we cannot distinguish between the three following cases: $w_{ij} = 0$ & $w_{ji} = 1$; $w_{ij} = 1$ & $w_{ji} = 0$; $w_{ij} = w_{ji} = 0$. By observing the equilibrium outcome we can disentangle the coefficients for w_{ij} and w_{ji}, and estimates of network externalities will be accurate as long as pairwise stability conditions generate sufficient variation in the structural regressors.[10] This approach overcomes reduced-form dyadic regressions (De Weerdt, 2004; Fafchamps and Gubert, 2007) by modeling agent-level incentives. However, like dyadic regressions it still assumes conditional independence between links: as long as agents are playing a myopic best response strategy in the context of pairwise stability, one can estimate preferences over many small 'local' perturbations away from *status quo*, which ultimately allow to recover the coefficients of network externalities. Thus, results should be interpreted as estimates of preferences over local deviations from pairwise equilibrium conditions.

3.3 Two-step estimator

Following Leung (2015), I rely on two identification assumptions: first, I assume a static game of link formation with partial information. In this context agents form links simultaneously on the basis of their common beliefs, which are built on observable characteristics. Secondly, agents can coordinate on a symmetric equilibrium where observationally-equivalent agents choose the same *ex-ante* strategies. In this context the estimation strategy boils down to a simple two-step procedure which I describe in what follows.

To avoid confusion in the notation, let us call g^e the equilibrium network that we observe in data. The first stage of the procedure fits a conditional probability representing agents' beliefs about the emerging network. To do so, I run a dyadic regression

[10] Note that the bivariate probit model with partial observability also suits link formation with no network externalities, as illustrated by Comola and Fafchamps (2014).

of the links in g^e on the relational characteristics of the dyads X_{ij}, plus (at least) one dyadic covariate Z_{ij} which serves as excluded instrument. This generates a probability distribution $\widehat{p}$, where $\widehat{p}_{ij}$ is the fitted probability that $g^e_{ij} = 1$. Note that $\widehat{p}$ represents the common beliefs by incorporating all the *ex-ante* information available to agents, including the predictable components of network externalities.

In the second stage of the procedure I estimate Equation (6) by computing the structural regressors Λ_ϕ on the basis of $\widehat{p}$. Since $\widehat{p}$ is a distribution rather than a fitted variable, it needs first to be mapped into a topological object representing the expected network architecture. For this purpose I adopt the stochastic network formation rule known as Poisson model, and draw a network realization assuming that all links are created independently according to the probabilities given by $\widehat{p}$.[11] By expressing the expected link outcomes as a reduced-form function of the dyads' relational observables, I circumvent two major problems. First, standard regression techniques may suffer from omitted variable endogeneity of the following form: if two agents are observationally equivalent but they have different network positions, this may be imputed to network externalities as well as to unobserved characteristics. As long as there is a valid instrument, the fitting procedure rules out unobserved heterogeneity and all network realizations drawn from $\widehat{p}$ are purged from omitted variable endogeneity in the usual 2SLS way. Second, the estimator is robust to the presence of multiple equilibria as long as one assumes that agents can coordinate on one symmetric equilibrium through a public signal (Leung, 2015), which allows the estimation of the parameters of interest with only one observed network realization.[12]

As a last issue, note that K different random draws from $\widehat{p}$ will produce K different network realizations $g^1, ..., g^K$. If I run the second-stage regression of Equation (6) using one arbitrary network realization k, the estimated coefficients $a(\phi)_k$ are unbiased but standard inference is inconsistent. In analogy with 2SLS bootstrap (Freedman, 1984), in the empirical illustration I correct the confidence intervals by using the quadratic assignment procedure (QAP), a re-sampling technique specifically designed for dyadic data which has been widely used in social network analysis (Hubert and Schultz, 1976; Krackhardt, 1987; Nyblom *et al.,* 2003; Eagle, Pentlan and Lazer, 2009). This consists in a appropriately designed permutation test which accounts for the two sources of variation, namely the correlation among dyadic observations and the variability in the

[11]The Poisson model is the workhorse of static random graphs, and has been shown to exhibit some of the key features of large-scale social networks (Jackson, 2008). Under the standard Poisson model all links are formed independently with identical probability, while I use the dyad-specific probabilities given by $\widehat{p}$.

[12]For simultaneous equation models with discrete outcomes, the existence of multiple equilibria in the underlying game raises problem of logical inconsistency in the associated likelihood function (Heckman, 1978; Gourieroux, Lafont and Monfort 1980; Maddala 1983; Tamer, 2003).

generated networks. A detailed description of the two-step procedure and the QAP permutation test is reported in the Appendix, Section A.2.

4 Simulations

In this section the consistency of the estimated parameters for network externalities is demonstrated with a simulation exercise. I proceed in the following way:

1. I posit a data generating process of the form:

$$u_i(g+ij) - u_i(g-ij) = x_j + 0.5\Lambda_{2_{ij}} + d_{ij}$$

   where $u_i(g+ij) - u_i(g-ij)$ represents i's utility from linking with j, x_j is an individual attribute representing wealth, and d_{ij} is a dyadic attribute representing distance ($d_{ij} = d_{ji}$). The term $\Lambda_{2_{ij}}$ represents network externalities from 2-step away contacts in the dimension of wealth. According to network theory externalities deteriorate with distance, therefore the coefficient for $\Lambda_{2_{ij}}$is assumed to be smaller than the one for x_j (Jackson and Wolinsky, 1996; Bala and Goyal, 2000). After normalizing the coefficients of x_j and d_{ij} to one, I have chosen to set the coefficient of $\Lambda_{2_{ij}}$ to 0.5 as this value increases significantly the number of links: in the stochastic process described in Step 2 below, *ceteris paribus* one gets about 30% of existing links if he sets the coefficient of $\Lambda_{2_{ij}}$to 0 versus about 60% of existing links if he sets it to 0.5.

2. I generate multiple pairwise stable networks following this data generation process. To the best of my knowledge no previous study has proposed a procedure to generate stable network architectures with externalities, and the task has proven to be computationally challenging. I program an algorithm which operates as follows:

   (a) The algorithm starts from 30 individual observations and generates the corresponding dyadic sample (870 directed *vs.* 435 undirected dyads) by randomly drawing the parameters $x_j \sim N(0, 0.1)$ and $d_{ij} \sim N(0, 0.1)$;

   (b) It starts by building a pairwise stable network with no externalities, that is, it sets $g_{ij} = 1$ if $u_i(g+ij)-u_i(g-ij) = x_j+d_{ij} > 0$ and $u_j(g+ij)-u_j(g-ij) = x_i + d_{ij} > 0$;

   (c) It loops across all dyads in random order for multiple times (rounds) introducing network externalities. That is, for each dyad ij such that $g_{ij} = 0$ and for each round r, the algorithm first computes the terms $\Lambda_{2_{ij}}$ and $\Lambda_{2_{ji}}$ on the basis of the current matrix of geodesic distances, and then it evaluates the utilities $u_i(g + ij) - u_i(g - ij) = x_j + 0.5\Lambda_{2_{ij}} + d_{ij}$ and

$u_j(g+ij) - u_j(g-ij) = x_i + 0.5\Lambda_{2_{ji}} + d_{ij}$ respectively. If thanks to network externalities the link results now profitable for both parts involved, the algorithm sets $g_{ij} = 1$ and updates the entire distance matrix (by re-computing *all* geodesic distances between *all* indirectly-connected agents). The loop continues until there is no change in the entire network architecture for two consecutive rounds, *i.e.* until the network stabilizes and the link updating process stops. The resulting network is a pairwise stable equilibrium with externalities of the form specified above;

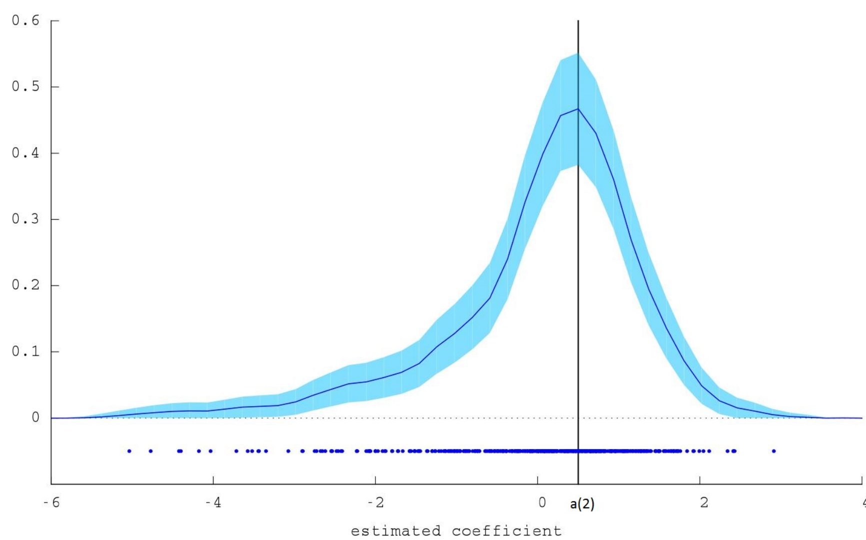
This procedure is repeated 500 times, which produces a dataset of 500 simulated pairwise stable networks. Note that each network is an independent draw from the underlying distribution which is already purged from omitted variable endogeneity;

3. For each of these 500 simulated networks I compute the structural regressors following the procedure P_1 in the Appendix;

4. Finally, for each of these 500 networks I estimate Equation (6) as a bivariate probit with partial observability. This gives me 500 estimated parameters of interest $a(2)$ for indirect connections Λ_2.

Figure 1 plots the non-parametric distribution of the estimated $a(2)$ coefficients using a Kernel Epanechnikov (bandwidth computed with Silverman's optimal rule). The light blue area indicates the 99% confidence interval, while the dark blue dots represent the frequency of observations.

Figure 1: Simulation results: distribution of $a(2)$



The graph clearly shows that the distribution of the estimated coefficients is centered around the true value of the parameter which is 0.5: indeed, by trimming the 1% extreme values, the mode of the $a(2)$'s distribution is 0.48. Overall, this simulation exercise should reassure the reader that the parameters of network externalities can be consistently estimated with the computational procedure I propose.

5 Data

My application builds on a large empirical literature on the formation of risk-sharing arrangements in developing countries (Coate and Ravallion, 1993; Grimard, 1997; Ligon, Thomas and Worrall, 2000; Fafchamps and Lund, 2003; Goldstein, De Janvry and Sadoulet, 2005).[13] Several of these studies estimate dyadic regression models with relational attributes, and they do not recognize the importance of network architecture.[14] I use data from the household survey of Nyakatoke, a small village in the Buboka rural district of Tanzania. All 307 adult inhabitants of Nyakatoke belonging to 119 different households have been interviewed in five regular intervals from February to December 2000. This has produced a rich dataset containing information on households' demographics, wealth, income sources and income shocks, transfers and risk-sharing links. These data have been the object of numerous articles (De Weerdt and Dercon, 2006; De Weerdt and Fafchamps, 2011; Vandenbossche and Demyunck, 2012; Comola and Fafchamps, 2014; Comola and Fafchamps, 2015). In what follows the $(119 \cdot 118)/2 = 7021$ undirected household dyads are taken as units of analysis.

Rural villages are an appropriate setting to study network formation, because the community can be entirely surveyed and because several confounding effects (such as spatial and informational barriers) can be reasonably ruled out. The village of Nyakatoke is isolated (the few unpaved roads leaving the village are hardly passable during rains) and relatively poor (consumption for adult equivalent is less than 2 US$ a week, and average food share in consumption is 77%). Households get most of their income from agricultural activities, especially the cultivation of coffee and banana; other sources of income are rare and off-farming activities are mostly considered supplementary to farming. All households are Muslim, Lutheran or Catholic.[15]

In Nyakatoke risk-sharing is the main strategy to cope with idiosyncratic shocks like

[13] Among the few theoretical contributions on risk-sharing see Genicot and Ray (2003), Bloch Genicot and Ray (2008), Jackson Rodriguez-Barraquer and Tan (2012), Ambrus Mobius and Szeidl (2014).

[14] The exception is Krishnan and Sciubba (2009), who identify the common features of all equilibrium configurations in a model with negative network externalities and test the model's predictions against data on labor exchange arrangements in Ethiopia.

[15] For additional information on Nyakatoke I remand to Mitti and Rweyemamu (2001) and De Weerdt (2002).

sickness, death, crime and court cases, and ceremonies (De Weerdt and Dercon, 2006). During the first survey round all respondents were asked *"Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labour?"*.[16] I interpret these answers as proxies for existing risk-sharing links, which are assumed symmetric (*i.e.* if i is linked to j then j must be linked to i - even tough *ex-post* transfers may flow in one direction only) and bilateral (*i.e.* mutual consent is needed).[17] Therefore I assume that a link between two respondents exists if either of them declares so, and I draw a link between the two households they belong to. This provides me with 490 undirected links among the 7021 households dyads in Nyakatoke.[18] The resulting network is dense, with a mean geodesic distance of 2.5 steps and a maximum geodesic distance of 5 steps. No household is isolated, and the average number of links is 8.2. The network exhibits all the empirical regularities of large social networks.[19] A graphical representation of the network is reported in Figure 2.

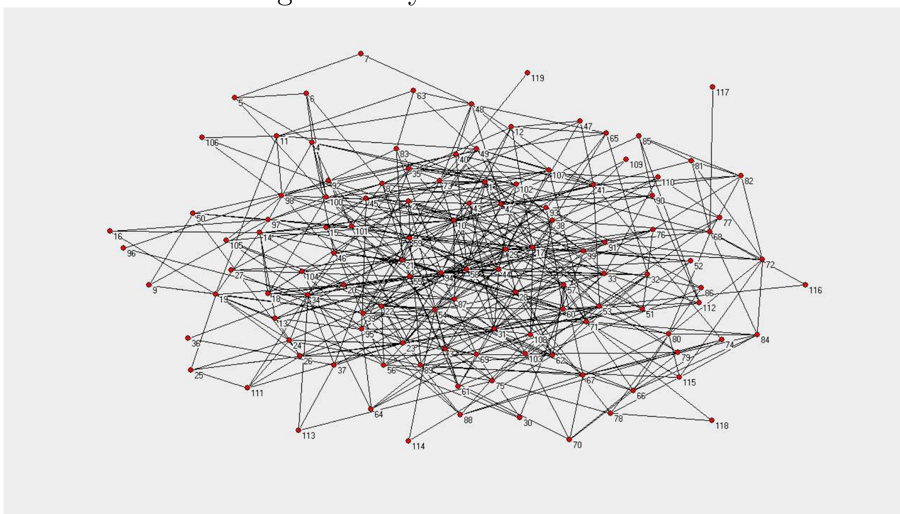
[16]The partners who live out of the village (34% of all mentioned partners) are omitted from the analysis since no information on them is available.

[17]This interpretation is consistent with the phrasing of the question and the way survey respondents have understood it. This risk-sharing question was piloted twice, first in the Philippines (Fafchamps and Lund, 2003) and later in Nyakatoke, and it was adopted in the current form because respondents understood it and were willing to answer it. Other questions were tried, for instance drawing a distinction between those you would help and those you would seek help from, but respondents were confused by the distinction which they perceived as non-existent, and complained they were asked the same question twice.

[18]The share of discordant statements is always very high in self-reported data (Fafchamps and Lund, 2003; De Weerdt and Fafchamps, 2011; Banerjee *et al.*, 2013; Liu *et al.,* 2013). This is also the case in Nyakatoke, where 140 links are reported by both sides and 350 by one side only. If one believes that risk-sharing links are intrinsically bilateral, the treatment of discordant statements turns into an assumption on mis-reporting. In my case, whenever i reports a link and j does not, it is equally legitimate to assume under-reporting by j or over-reporting by i. I choose under-reporting for the sake of convenience, because few positive outcomes exacerbate the convergence difficulties of partial observability models. For a structural approach to mis-reporting using the same data I remand to Comola and Fafchamps (2015).

[19]The Nyakatoke network has an unique component covering the entire population, the diameter is in the order of $ln(n)$ and the clustering coefficient (which measures the tendency of linked nodes to have common neighbors) is 7 times larger than in a randomly generated Poisson network with similar characteristics. For further details on the so-called small world properties see Jackson and Rogers (2007).

Figure 2: Nyakatoke's network



6 Variables definition

The regressors are illustrative of the type of variables to be included in an analysis of this kind. In order to proxy for the relational attributes X_{ij} affecting both the probability of a link and its utility I use religion and blood bonds: the dummy *same religion*$_{ij}$ equals one if the two households profess the same religion, and the dummy *blood bond*$_{ij}$ equals one if a member of household i has a blood bond (child, parents, siblings) with a member of household j. These variables capture homophyly (*i.e.* the tendency to link with similar agents) and have been recognized among the strongest predictors of risk sharing (Fafchamps and Lund, 2003; Fafchamps and Gubert, 2007; De Weerdt and Fafchamps, 2011). They also enter the second-step equation, as one can argue that social proximity decreases the enforcement and monitoring costs of risk-sharing arrangements (Karlan, 2007).

As exclusion restriction Z_{ij} (*i.e.* a variable which affects the probability of a link without affecting its utility) I use the geographical distance between the two houses (in meters). This relies on the assumption that in Nyakatoke there are no informational and infrastructural barriers, and that the enforceability of risk-sharing arrangements does not depend on the households' location within the village area. The assumption seems plausible since the community is rather small, isolated and densely inhabited (the average distance between houses of 523 meters and 90% of households live within a distance of 1 kilometer from each other); however the usual caveats apply.[20]

In the second stage the w_{ij} equation includes: the relational characteristics X_{ij}, the partner's wealth y_j and the structural regressors $\Lambda_{\phi_{ij}}$. The regressors entering the w_{ji} equation are defined analogously. Wealth y_j is the total monetary value of j's land and livestock assets in Tanzanian Shillings (1 unit=100000 *tzs*).[21] Wealth is mainly inherited through patrilineal rules and passed down the clan, which rules out the most pressing inverse causality concerns.[22] All structural regressors are also expressed in Tanzanian Shillings. In Nyakatoke all households are indirectly connected and the maximum geodesic distance is 5, which implies that externalities from step 1 to 5 are perfectly collinear. However, since the network is rather dense (with an average geodesic distance of 2.5), readjustments are reabsorbed quickly, so that $\Lambda_{4_{ij}}$ and $\Lambda_{5_{ij}}$ are perfectly collinear for 98.5% of dyads. Therefore, I cap $a(\phi) < 4$. Descriptive statistics are reported in Table 1.

Table 1: Descriptive statistics

variable	mean	min	max	sd
g_{ij}	0.070			
$same\,religion_{ij}$	0.354			
$blood\,bond_{ij}$	0.016			
$distance_{ij}$	0.522	0.014	1.738	0.303
$wealth_j$	4.143	0	27.970	3.865
$\Lambda_{2_{ij}}$	15.800	-27.970	131.492	13.961
$\Lambda_{3_{ij}}$	-11.628	-109.922	358.834	18.789

Note: $N = 7021$ for standard regressors, and N=7021 $\cdot$ 50 for the structural regressors $\Lambda_{2_{ij}}$ and $\Lambda_{3_{ij}}$ which are computed on 50 network realizations $g^1, ..., g^{50}$.

7 Analysis

Results are computed on the full sample of 7021 household dyads. In the specifications where structural regressors are not included standard probit coefficients are reported,

[20] Since there is no experimental design to be exploited in the Nyakatoke survey, this should be seen as the best suggestive evidence at hand. On the other hand, collecting direct self-reported information on desire to link would not have solved the issue, since such data are plagued by self-censoring (Hitsch, Hortacsu and Ariely, 2010; Belot and Francesconi, 2015).

[21] Data on land were originally in acres, and were transformed in monetary equivalent with a conversion rate of 300000 *tzs* for 1 acre, which reflects average local prices in 2000. For international comparisons, the exchange rate in 2000 was 1 US dollar for 800 *tzs*.

[22] In Nyakatoke customary land tenure laws used to prohibit the selling of land (Mitti and Rweyamamu, 2001).

along with their QAP p-values computed over 250 standard scrambles (see Section A.2 of the Appendix). When the structural regressors enter the specification I drawn 50 network realizations which generates 50 sets of estimated coefficients $a(2)_1, \dots a(2)_{50}$ and $a(3)_1, \dots a(3)_{50}$ and run 5 scrambles each for a total of 250 observations; if this is the case, I report both the coefficients' average and their distribution. As for all permutation tests, the p-value is computed as the percentage of times that a scrambled sample's coefficient lies on the left of the true coefficient of the corresponding network realization. Coefficients and p-values for marginal effects are computed analogously (*i.e.* I locate the observed marginal effect on the distribution of marginal effects under the null hypothesis).

7.1 First-stage results

Table 2 reports the results from the first-stage probit regression, showing that coefficients are strongly significant with the expected sign.

Table 2: First-stage Results

dependent variable: g^e_{ij}		
	(1)	(1)
	coefficient	marginal effect
$same\,religion_{ij}$	0.2104***	0.0249***
	(0.000)	(0.000)
$blood\,bond_{ij}$	1.7149***	0.203***
	(0.004)	(0.000)
$distance_{ij}$	-1.0691***	-0.1266***
	(0.000)	(0.000)

Notes: probit coefficients and marginal effects reported. p-values from 250 standard QAP permutations in parentheses (*** $p < 0.01$; ** $p < 0.05$). Constant term included.

Two exercises described in what follows are designed to illustrate the predictive power of this first-stage regression. First, I compare the network realizations $g^1, ..., g^{50}$ drawn from the fitted probabilities $\widehat{p}$ which I obtain from Table 2 with a uniform-probability Poisson random network. I generate the uniform-probability network g^u such that each link is formed independently with $p = 0.07$ (corresponding to the average expected degree 8.2 of both g^e and $g^1, ..., g^{50}$). As a result, the networks $g^1, ..., g^{50}$:

- are better at predicting existing links: $g^1, ..., g^{50}$ predict on average 17% of the links in g^e, versus 7% of g^u;

- are much better at capturing the homophyly patterns in the data: for instance, for those dyads where $blood\,bond_{ij} = 1$ the links exists in 65% of cases under g^e, in 50% to 60% of cases under $g^1, ..., g^{50}$ and only in 10% of cases under g^u;

- are better at replicating the degree distribution: for instance, the percentage of nodes with degree less or equal to 4 is 19% under g^e, 11% under g^{50} and 6% under g^u.

As a second exercise, I compute the structural regressors on the observed network g^e (that I call Λ_2^e and Λ_3^e) which I report in Table 3. When results of Table 3 are compared with the corresponding statistics Λ_2^1, ..., Λ_2^{50} and Λ_3^1, ..., Λ_3^{50} for the 50 network realizations $g^1, ..., g^{50}$ (reported in Table 1), they suggest that this fitting procedure provides a good approximation of the quantities of interest.

Table 3: Structural regressors for g^e (N=7021)

variable	mean	min	max	sd
$\Lambda^e_{2_{ij}}$	15.290	-21.740	145.621	18.206
$\Lambda^e_{3_{ij}}$	-10.525	-108.732	155.466	20.399

7.2 Second-stage results

Table 4 reports the partial observability bivariate probit results for w_{ij}, that is, the willingness of i to link with j (the coefficients for w_{ji} are not reported because they are constrained to be the same). Column (1) refers to the benchmark case where only the wealth of direct partners affects link formation, while column (2) incorporates network externalities. The corresponding marginal effects are reported in columns (3) and (4). Since we have 50 stochastic realizations of the expected network, the reported coefficients for $\Lambda_{2_{ij}}$ and $\Lambda_{3_{ij}}$ in Table 4 are the average over 50 values. Figures 3 and 4 show that the distribution of these coefficients is well behaved and centered around the mean - which corroborates the evidence in favor of the fitting procedure.

Table 4: Second-stage results

	dependent variable: w_{ij}			
	(1)	(2)	(3)	(4)
	coefficient		marginal effect	
$wealth_j$	0.0108***	0.0116***	0.0032***	0.0029***
	(0.000)	(0.000)	(0.000)	(0.000)
$\Lambda_{2_{ij}}$		-0.0023**		- 0.0006**
		(0.012)		(0.024)
$\Lambda_{3_{ij}}$		-0.0011		-0.0003
		(0.052)		(0.080)
Controls	yes	yes	yes	yes

Notes: bivariate probit coefficients and marginal effects reported. P-values in parentheses (*** $p < 0.01$, ** $p < 0.05$), computed over 250 QAP permuted samples: 250 standard QAP permutations in columns (1) and (3), 5 permutation each for 50 network realizations in columns (2) and (4). Reported coefficients in columns (2) and (4) are averages over 50 values. *same religion*$_{ij}$ and *blood bond*$_{ij}$ included as controls.

Figure 3: Kernel of the estimated coefficients $a(2)$ for $\Lambda_{2_{ij}}$ $(N = 50)$

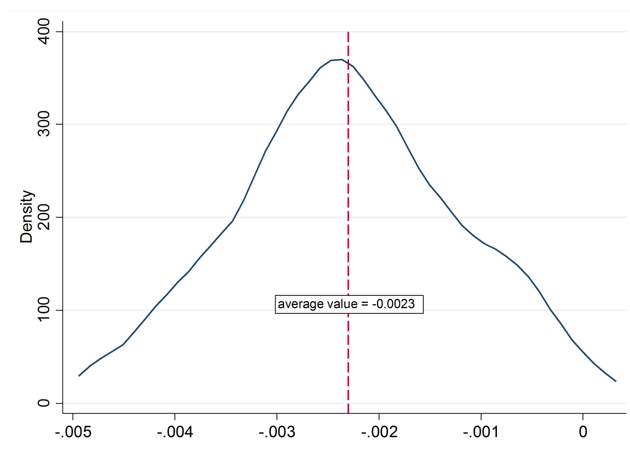
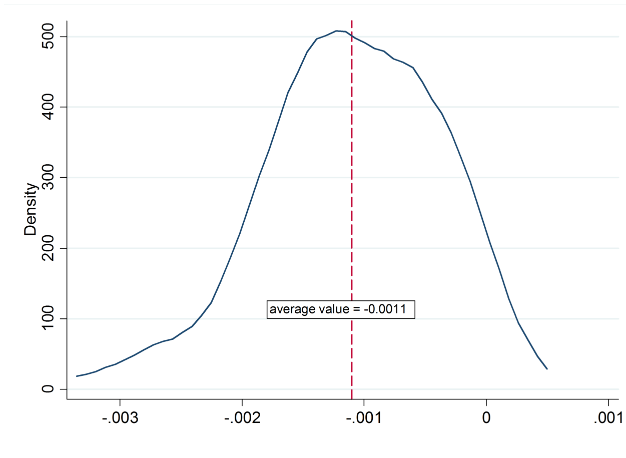


Figure 4: Kernel of the estimated coefficients $a(3)$ for $\Lambda_{3_{ij}}$ $(N = 50)$



Results show that $wealth_j$ is positive and significant across all specifications: as expected, the wealthier a potential partner, the more desirable a link with him. The coefficients for $\Lambda_{2_{ij}}$, which represents the gain in term of two-step-away wealth, is significant and negative. This should be interpreted as revealed preferences over partners' expected position*: "ceteris paribus, I prefer friends whose other friends are few and/or poor"*. On the other hand $\Lambda_{3_{ij}}$ is not significant at standard confidence level, suggesting that agents do not (or are not able to, due to the cognitive complexity of the exercise) take into account the network configuration further away. The absolute magnitude of the effects is decreasing in the geodesic distance, as theory predicts.

The network literature has modeled positive externalities in non-rivalry contexts such as information flows and public goods, and negative externalities in the context of strategic competition and rival goods (Jackson and Wolisky, 1996; Goyal and Joshi, 2006a and 2006b; Morril, 2011). Risk-sharing arrangements in developing countries are an intriguing example which may combine both types of externalities: indirect connections are useful if benefits spill over link, but detrimental if goods are rival. The absence of positive network externalities is not surprising, since in Nyakatoke most economic benefits are tied to agricultural resources which are rival by nature.[23] This can produce a competition mechanism over the number of indirect partners (*i.e.* many connections dilute the effort and the advantage for each of them). Alternatively, this result can be interpreted in terms of strive for social status. From previous work on

[23]See Mitti and Rweyamamu (2001) for anecdotal evidence on free gift of land for cultivation of seasonal crops and casual labor among partners.

patron-client relations we know that asymmetric exchanges, where the wealthy partner provides economic help in exchange of non monetary incentives like esteem and social status, are the norm in agrarian societies (Breman, 1974; Fafchamps, 1992; Platteau, 1995; Platteau and Seki, 2007; Platteau and Sekeris, 2010). In this context villagers may prefer to link with those with high social status, that is, those that are engaged in many asymmetric relationship of the patron-client type with poorer partners. However, both interpretations are speculative and beyond the illustrative scope of this exercise.

8 Conclusions

This paper proposes a structural econometric model to estimate network externalities from indirect connections. The approach is computationally simple, and builds on local pairwise stable conditions in order to estimate individual incentives to deviate from the *status quo*. The proposed two-step estimator addresses omitted variable endogeneity in the context of a simultaneous link announcement game with partial information. When the estimation protocol is applied to risk-sharing data from the Tanzanian village of Nyakatoke, results suggest that the network architecture has indeed an explanatory value. Network externalities in the dimension of wealth are found to be negative, which can be interpreted as a competition mechanism for partners' wealth and/or social status.

While network theory has extensively modeled how local topology provides incentives to link formation, empirical evidence lags behind. This paper is an attempt in this direction, showing that one might still make valid inference about some aspects of preferences over networks, including network externalities, when data on a single network are available.

Appendix: Computational details

This section describes the computational steps to run the estimation protocol of Section 3. In particular, Section A.1 explains how to compute the structural regressors of Section 3.1, while Section A.2 shows how to implement the two-step estimator of Section 3.3 along with an appropriate permutation test to compute confidence intervals.

A.1 Structural regressors

The following procedure explains how to compute $\Lambda_{\phi_{ij}}$ and $\Lambda_{\phi_{ji}}$ of Equation (6) for a given network architecture g.

Procedure P_1

For each dyad $\{ij\}$:

1. generate the two adjacency matrices $g + ij$ and $g - ij$ (where you set $g_{ij} = 1$ and $g_{ij} = 0$ respectively and keep the rest of g as given). By construction one will be the network and the other the dyad-specific counter-factual;

2. use an iterative graph search algorithm (for instance the one by Dijkstra, 1959) to generate the two matrices of geodesic distance τ_{g+ij} and τ_{g-ij};[24]

3. combine τ_{g+ij} with the vector y to obtain $\lambda^{+ij}_{i,\phi}$ and $\lambda^{+ij}_{j,\phi}$ (*i.e.* the sum of wealth of all nodes ϕ-step-away from i and j respectively under $g+ij$) for all $\phi = 2, ..., N-2$;

4. repeat step 3 for τ_{g-ij};

5. combine the outputs from steps 3 and 4 and obtain the $2(N-3)$ structural regressors of interest ($\Lambda_{\phi_{ij}}$ and $\Lambda_{\phi_{ji}}$ for $\phi = 2, ..., N-2$).

A.2 Confidence intervals

As sampled agents are the same across dyads, decisions to link are typically not independent of each other. This invalidates inference unless confidence intervals are computed in an appropriate manner. When data belong to a single population, there are two available solutions: the dyad-specific robust covariance matrix (Fafchamps and Gubert, 2007), and dyad-specific permutation tests. I choose the latter approach and use the quadratic assignment procedure (QAP), a re-sampling technique specifically designed for dyadic data. In standard permutation tests data are repeatedly scrambled in a way which is consistent with the null hypothesis (*i.e.* the hypothesis that the

[24] This can be time-consuming, as it involves computing two distance matrices of dimension $N \times N$ through a graph search routine which must be repeated $N(N-1)/2$ times.

true coefficient of a variable of interest is zero). By constructing the distribution of the coefficient under the null hypothesis and locating the observed coefficient β on this distribution, the researcher is computing the p-value: if at least 95% of the coefficients computed on the permutation samples lie on the left of the observed β one can reject the null hypothesis with 5% significance level. Standard permutation tests yield consistent inference in a large range of situations (Good, 2000). The QAP method is a specific permutation design for dyadic data which has been widely used in social network analysis (Hubert and Schultz, 1976; Krackhardt, 1987; Nyblom *et al.,* 2003; Eagle, Pentlan and Lazer, 2009). In each QAP-permuted sample rows and columns are scrambled the same way: if the k and m rows of the relational variable X are permuted, then the k and m columns are permuted as well. This scramble design preserves the dyadic correlation but eliminates any relationship between variables, allowing the estimation of confidence intervals under the null hypothesis. The QAP test explained in Steps 5-7 below runs on multiple network realizations, in order to take into account the stochastic variability of the data generating process.

Procedure P_2

1. Run the first-stage dyadic probit regression $Pr(g^e_{ij} = 1) = Pr(\gamma + \delta X_{ij} + \vartheta Z_{ij} \geq \upsilon_{ij})$, where the relational attribute Z_{ij} meets the exclusion restriction, and generate the probability distribution $\widehat{p}$ such that $\widehat{p}_{ij}$ is the fitted probability that $g^e_{ij} = 1$;

2. Generate K network realizations $g^1, ..., g^K$ by taking K Poisson draws from $\widehat{p}$;

3. apply Procedure P1 to each of these K network realizations and obtain K sets of structural regressors $\Lambda^1_\phi, ..., \Lambda^K_\phi$;

4. for each $k = 1, ..., K$ estimate Equation (6) using Λ^k_ϕ and obtain the second-stage coefficients $a_k = [a(1)_k, \ ..., \ a(N-2)_k]'$. This gives K sets of unbiased coefficients $a_1, ..., a_K$;

5. for each $k = 1, ..., K$ generate M QAP samples which permute Λ^k_ϕ under the null hypothesis by preserving the order of rows and columns;

6. for each $k = 1, ..., K$ and $m = 1, ..., M$ estimate the permuted coefficients $a^m_k = [a(1)^m_k, \ ..., \ a(N-2)^m_k]'$. This gives $K \cdot M$ sets of permuted coefficients $a^1_1, ..., a^M_K$ for hypothesis testing;

7. compute the p-values for the second-stage coefficients of Step 4 on the basis of the $K \cdot M$ permuted coefficients of Step 6. That is, the p-value of y_j is computed as the percentage of times that $a(1)^m_k < a(1)_k$. Similarly, the p-value of $\Lambda_{2_{ij}}$ is computed as the percentage of times that $a(2)^m_k < a(2)_k$.

References

[1] Abowd, J. and H. Farber (1982): "Job Queues and the Union Status of Workers", *Industrial and Labor Relations Review*, 35(3), 354-367.

[2] Advani A. and B. Malde (2014): "Empirical Methods for Networks Data: Social Effects, Network Formation and Measurement Error", IFS Working Paper W14/34.

[3] Ambrus, A., M. Mobius and A. Szeidl (2014): "Consumption Risk-Sharing in Social Networks", *American Economic Review*, 104(1): 149-82.

[4] Badev, A. (2013): "Discrete Games in Endogenous Networks: Theory and Policy", PSC Working Paper Series, PSC 13-05.

[5] Banerjee, A., A. Chandrasekhar, E. Duflo and M. Jackson (2013): "The Diffusion of Microfinance", *Science*, 341 (6144).

[6] Bala, V. and S. Goyal (2000): "A Noncooperative Model of Network Formation", *Econometrica*, 68(5), 1181-1229.

[7] Belot, M. and M. Francesconi (2015): "Meeting Opportunities and Mate Selection: A field study", *Journal of Human Resources*, forthcoming.

[8] Bhamidi, S. G. Bresler, and A. Sly (2008): "Mixing Time of Exponential Random Graphs", arXiv:0812.2265.

[9] Bloch, F., G. Genicot and D. Ray (2008): "Informal Insurance in Social Networks", *Journal of Economic Theory*, 143(1), 36-58.

[10] Boucher, V. and I. Mourifié (2013): "My friend far far away: Asymptotic properties of pairwise stable networks," SSRN WP n. 2170803.

[11] Breman, J. (1974): "*Beyond Patronage and Exploitation*", University of California Press, Berkeley, CA.

[12] Breschi, S. and F. Lissoni, (2005): "Cross-firm inventors and social networks: Localized knowledge spillovers revisited", *Annales d'Economie et de Statistique*, 189–209.

[13] Bresnahan, T. and P. Reiss (1991): "Empirical models of discrete games", *Journal of Econometrics* 48 (1-2): 57 - 81.

[14] Chandrasekhar, A. (2015): "Econometrics of network formation", *Oxford handbook on the economics of networks* (edited by Yann Bramoullé, Andrea Galeotti and Brian Rogers), forthcoming.

[15] Chandrasekhar, A. and M. Jackson (2014): "Tractable and Consistent Random Graph Models", NBER working paper 20276.

[16] Chandrasekhar, A. and M. Jackson (2015): "A Network Formation Model based on Subgraphs", mimeograph.

[17] Chatterjee, S. and P. Diaconis (2013): "Estimating and understanding exponential random graph models", *The Annals of Statistics*, 41, 2428–2461.

[18] Christakis, N., J. Fowler, G. Imbens and K. Kalyanaraman. (2010): "An empirical model of strategic network formation," NBER Working Paper No. 16039.

[19] Coate, S. and M. Ravallion (1993): "Reciprocity without Commitment: Characterization and Performance of Informal Insurance Arrangements", *Journal of Development Economics*, 40, 1-24.

[20] Comola, M. and M. Fafchamps (2014): "Testing Unilateral and Bilateral Link Formation", *The Economic Journal*, 124: 954-975.

[21] Comola, M. and M. Fafchamps (2015): "The Missing Transfers: Estimating Misreporting in Dyadic Data", *Economic Development and Cultural Change,* forthcoming.

[22] De Paula, A, S. Richards-Shubik and E. Tamer (2015): "Identification of Preferences in Network Formation Games", CeMMAP Working Paper 29/15.

[23] De Paula, A, and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information", *Econometrica*, 80 (1).

[24] De Weerdt, J. (2002): "Community Organizations in Rural Tanzania: A Case Study of the Community of Nyakatoke, Buboka Rural District", Economic Development Initiatives mimeograph.

[25] De Weerdt, J. (2004): "Risk Sharing and Endogenous Network Formation", in *Insurance Against Poverty*, Stephan Dercon (ed.), Oxford University Press.

[26] De Weerdt, J. and S. Dercon (2006): "Risk Sharing Networks and Insurance against Illness", *Journal of Development Economics*, 81(2), 337-356.

[27] De Weerdt, J. and M. Fafchamps (2011): "Social Identity and the Formation of Health Insurance Networks", *Journal of Development Studies*, 47(8), 1152-1177.

[28] Dijkstra, E. (1959): "A note on two problems in connexion with graphs", *Numerische Mathematik,* 1, 269-271.

[29] Dutta, B. and S. Mutuswami (1997): "Stable Networks", *Journal of Economic Theory*, 76(2), 322-344.

[30] Eagle N., A. Pentland and D. Lazer (2009): "Inferring friendship network structure by using mobile phone data", *Proceedings of the National Academy of Sciences of the United States of America*, 106(36), 15274-15278.

[31] Fafchamps, M. (1992): "Solidarity Networks in Preindustrial Societies: Rational Peasants with a Moral Economy", *Economic Development and Cultural Change*, 41(1), 147-174.

[32] Fafchamps, M. and F. Gubert (2007): "The Formation of Risk Sharing Networks", *Journal of Development Economics*, 83(2), 326-350.

[33] Fafchamps, M. and S. Lund (2003): "Risk Sharing Networks in Rural Philippines", *Journal of Development Economics*, 71(2), 261-287.

[34] Farber, H. (1983): "The Determination of Union Status of Workers", *Econometrica*, 51(5), 1417-1438.

[35] Freedman, D. (1984): "On Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models", *The Annals of Statistics*, 12(3), 827-842.

[36] Galeotti, A., S. Goyal and J. Kamphorst (2006): "Network Formation with Heterogeneous Players", *Games and Economic Behavior*, 54(2), 353-372.

[37] Genicot, G. and D. Ray (2003): "Group Formation in Risk Sharing Arrangements", *Review of Economic Studies,* 70, 87-113.

[38] Goeree, Y, M. McConnell, T. Mitchell, T. Tromp and L. Yariv (2010): "The $1/d$ Law of Giving", *American Economic Journal: Microeconomics*, 2(1), 183-203.

[39] Goldstein, M., A. De Janvry and E. Sadoulet (2005): "Is a Friend in Need a Friend Indeed?" in *Insurance against Poverty*, Stephan Dercon (ed.), Oxford University Press.

[40] Good, P. (2000): "*Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*", Springer ed.

[41] Gourieroux, C., J. Laffont and A. Monfort (1980): "Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes", *Econometrica*, 48(3), 675-695.

[42] Goyal, S. and S. Joshi (2006a): "Unequal Connections", *International Journal of Game Theory,* 34, 319-349.

[43] Goyal, S. and S. Joshi (2006b): "Bilateralism and Free-trade", *International Economic Review,* 47(3), 749-778.

[44] Graham, B. (2015): "Methods of identification in social networks", *Annual Review of Economics* 7: 465 - 485.

[45] Grimard, F. (1997): "Households Consumption Smoothing through Ethnic Ties: Evidence form Core D'Ivoire", *Journal of Development Economics*, 53, 391-422.

[46] Heckman, J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica* , 46(4), 931-959.

[47] Hitsch, G., A. Hortaçsu, and D. Ariely (2010): "Matching and Sorting in Online Dating", *American Economic Review*, 100(1), 130-163.

[48] Hsieh, C. and L. Lee (2015): "A Social Interactions Model with Endogenous Friendship Formation and Selectivity", *Journal of Applied Econometrics*, forthcoming.

[49] Hubert, L. and J. Schultz (1976): "Quadratic assignment as a general data analysis strategy", *British Journal of Mathematical and Statistical Psychology*, 29, 190-241.

[50] Jackson, M. (2004): "A Survey of Models of Network Formation: Stability and Efficiency", in *Group Formation in Economics: Networks, Clubs and Coalitions*, Cambridge University Press.

[51] Jackson, M. (2008): "*Social and Economic Networks*", Princeton University Press.

[52] Jackson, M., T. Rodriguez-Barraquer and X. Tan (2012) "Social Capital and Social Quilts: Network Patterns of Favor Exchange", *American Economic Review*, 102(5): 1857-97.

[53] Jackson, M. and B. Rogers (2007): "Meeting Strangers and Friends of Friends: How Random are Social Networks?", *American Economic Review*, 97(3), 890-915.

[54] Jackson, M. and A. Van Den Nouweland (2005): "Strongly Stable Networks", *Games and Economic Behavior*, 51(2), 420-444.

[55] Jackson, M. and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks", *Journal of Economic Theory*, 71, 44-74.

[56] Karlan, D. (2007): "Social Connections and Group Banking", *Economic Journal*, 117(517), F52-F84.

[57] Kelejian, H, and G. Piras (2014): "Estimation of Spatial Models with Endogenous Weighting Matrices, and an Application to a Demand Model for Cigarettes", *Regional Science and Urban Economics*, 46: 140-149.

[58] Konig, M. (2015): “The Formation of Networks with Local Spillovers and Limited Observability”, *Theoretical Economics*, forthcoming.

[59] Konig, M. X. Liu, and Y. Zenou (2014): “R&D networks: Theory, empirics, and policy implications”, Technical report, CEPR Discussion Paper 9872.

[60] Krackhardt, D. (1987): “QAP partialling as a test of spuriousness”, *Social Networks*, 9, 171-186.

[61] Krishnan, P. and E. Sciubba (2009): “Links and Architecture in Village Networks”, *Economic Journal*, 119(537), 917-949.

[62] Leung, M. (2015): “Two-Step Estimation of Network-Formation Models with Incomplete Information”, *Journal of Econometrics*, 188(1), 182-195.

[63] Ligon, E., J. Thomas and T. Worrall (2000): “Mutual Insurance, Individual Savings and Limited Commitment”, *Review of Economic Dynamics*, 3, 216-246.

[64] Liu, X., E. Patacchini, Y. Zenou and L. Lee (2013): “Criminal Networks: Who is the Key Player?”, mimeograph.

[65] Maddala, G. (1983): “*Limited Dependent and Qualitative Variables in Econometrics*”, Cambridge University Press.

[66] Mele, A. (2015): “A Structural Model of Segregation in Social Networks”, mimeograph.

[67] Mitti R. and T. Rweyemamu (2001): “A Socio-Economic Profile of Nyakatoke”, *Social Security, Poverty and Economic Reforms in Kagera Region*, Report Nr. 1.

[68] Miyauchi, Y. (2014): “Structural Estimation of a Pairwise Stable Network with Nonnegative Externality”, SSRN WP 2165884.

[69] Morrill, T. (2011): “Network formation under negative degree-based externalities”, *International Journal of Game Theory*, 40(2), 367-385.

[70] Nyblom, J., S. Borgatti, J. Roslakka, and M. Salo (2003): “Statistical Analysis of Network Data: An Application to Diffusion of Innovation", *Social Networks*, 25, 175-95.

[71] Platteau, J. (1995): “An Indian Model of Aristocratic Patronage”, *Oxford Economic Papers*, 47(4), 636-662.

[72] Platteau, J. and P. Sekeris (2010): “On the feasibility of power and status ranking in traditional setups”, *Journal of Comparative Economics*, 38(3), 267-282.

[73] Platteau, J. and E. Seki (2007): "Heterogeneity, Social Esteem, and Feasibility of Collective Action", *Journal of Development Economics*, 83(2), 302-325.

[74] Poirer, D. (1980): "Partial Observability in Bivariate Probit Models", *Journal of Econometrics*, 12, 209-217.

[75] Shalizi, C. and A. Rinaldo (2012): "Consistency under Sampling of Exponential Random Graph Models," ArXiv 1111.3054v3.

[76] Sheng, S. (2014): "A Structural Econometric Analysis of Network Formation Games", mimeograph.

[77] Singh, J. (2005): "Collaborative networks as determinants of knowledge diffusion patterns", *Management Science*, 51(5), 756–770.

[78] Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria", *Review of Economic Studies*, 70, 147-167.

[79] Townsend, R. (1994): "Risk and Insurance in Village India," *Econometrica*, 62, 539-559.

[80] Udry, C. (1994): "Risk and Insurance in a Rural Credit Market: an Empirical Investigation in Northern Nigeria", *Review of Economic Studies*, 61, 495-526.

[81] Vandenbossche, J. and T. Demuynck (2012): "Network formation with heterogeneous agents and absolute friction," *Computational Economics*, 1-23.