Causal Effects in Social Networks

Marcel Fafchamps*

This paper reviews the current literature on the estimation of causal peer effects. After a discussion of causality in general, I introduce the standard peer effect model in networks and illustrate the reflection problem. I then present approaches to causal inference with observational data before introducing experimental approaches. I review estimation issues arising from measurement and sampling errors, and discuss how they affect causal inference in network and peer effect experiments. The last section of the paper broadens the discussion to encompass dynamic peer effects and link formation and illustrates the different meanings that causality can take in the estimation of peer effects.

L'ESTIMATION CAUSALE DES EFFETS DE PAIRS DANS LES RÉSEAUX SOCIAUX

Cet article passe en revue la littérature récente sur l'estimation causale des effets de pairs. Après une discussion de la causalité en général, j'introduis le modèle standard d'effets de pairs dans les réseaux et je donne une illustration du problème de réflexion. Je présente ensuite différentes approches de l'inférence causale avec des données d'observation, avant d'introduire les approches expérimentales. Je passe en revue les problèmes d'estimation qui sont soulevés par les erreurs de mesure et d'échantillonnage, et je discute la façon dont ils affectent l'inférence causale dans les expériences d'effets de réseau et d'effets de pairs. La dernière section de l'article élargit la discussion pour couvrir les effets de pairs dynamiques et la formation des liens, et illustre les différentes significations que peut revêtir le concept de causalité dans l'estimation des effets de pairs.

Classification *JEL* : C31, C21, C90

* Stanford University. *Correspondence:* Freeman Spogli Institute for International Studies, Stanford University, 616 Serra Street, Stanford CA 94305 (USA). *Email:* fafchamp@stanford.edu.

This paper was written in preparation for a keynote address to the XXXI[e] Journées de microéconomie appliquée, 5-6 June 2014, CERDI-CNRS, Université d'Auvergne, Clermont-Ferrand, France. I thank Yann Bramoullé, conference participants, and the journal editor for their valuable comments.

INTRODUCTION

Economists often talk of peer effects and network effects. But, as I will argue in this paper, there is considerable ambiguity as to what is meant by these expressions. The use of the word "effect" implies some sort of causality. What is less clear is what causality means in the context of social networks.

In the presentation, I start by reminding us of what economists today mean by "causal." I use a simple example to illustrate the inherent difficulties associated with a study of causality. I then draw upon the lessons learned from the literature to illustrate different notions of causal peer effects, and to outline the counterfactual experiments required to test them.

Since the purpose of this article is to provide a documented version of my keynote presentation, the reader will hopefully excuse the fact that I borrow extensively from my own research when providing empirical examples of the various issues discussed here.

CAUSALITY IN ECONOMICS

Randomized controlled trials and endogeneity

The literature on randomized controlled trials has had a profound impact on how causality is studied in economics. Randomized controlled trials (RCTs) are the official procedure by which new medical treatments are authorized for use. To summarize, in an RCT the administration of a treatment w is randomized across $i \in \mathrm{N}$ patients. Some proportion of patients, say half, are randomly selected to receive the treatment while the rest do not. The former are called the treated, the latter are called the controls. The effectiveness of the treatment is verified by comparing outcomes y between treated and control subjects. If outcomes are statistically different between treated and controls, e.g.,

$$\mathrm{E}[y \mid w = 1] \neq \mathrm{E}[y \mid w = 0]$$

the treatment is said to have a causal effect on outcomes—hopefully, a beneficial one.

While this approach has proved useful and influential in economic practice, it is a somewhat reductionist view of causality. Before RCTs made their way into economics, causality was modeled through flow charts and their mathematical representation, i.e., economic models. The statistical representation of these causal models is the system of simultaneous equations in econometrics. In these models, variables are thought to influence each other in a system. A simple example of such a system is:

$$\begin{aligned} y_1 &= \alpha_0 + \alpha_1 y_2 + \alpha_2 z_1 + u_1 \\ y_2 &= \beta_0 + \beta_1 y_1 + \beta_2 z_2 + u_2 \end{aligned} \tag{1}$$

where variables y_1 and y_2 are said to be endogenous—they influence each other—while variables z_1 and z_2 are said to be exogenous, since they are not

influenced by either y_1 or y_2. The availability of exogenous variables z_1 and z_2 ensures that system (1) is identified, in the sense that its coefficients can be estimated by using z_1 as an instrument for y_1 in equation 2, and z_2 as instrument for y_2 in equation 1. RCTs are seen as a way of generating "perfect" instruments since randomly assigned treatment is, by design, exogenous.

It is also possible to estimate a reduced-form version of system (1) by regressing endogenous on exogenous variables:

$$y_1 = a_0 + a_1 z_1 + a_2 z_2 + e_1$$

$$y_2 = b_0 + b_1 z_1 + b_2 z_2 + e_2$$

and there is a direct mathematical relationship between the α's and β's—the structural coefficients—and the a's and b's—the reduced-form coefficients. As long as the system is identified, all structural coefficients can be recovered from the reduced-form coefficients.

Causes and theories

RCTs are best at producing reliable estimates of reduced-form coefficients. By themselves they do not provide much information on what the structural model might be. For the approval of medical treatments, a reduced-form model may be all that we need: as long as the biology of the human body is sufficiently self-contained and uniform across individuals, statistical evidence about the efficacy of a treatment in a small sample of individuals is able to predict the efficacy of this treatment in the population at large. Better treatments, however, may be found if we have a better understanding of the channels by which treatment affects outcomes. Medicine would not be what it is today if RCTs were its only tool and it had no understanding of how the human body functions as a system of complex mutual causation.

In economics, it is often difficult to predict the effect of a treatment when it is applied outside the context of a specific RCT. This is because, in social and economic phenomena, the process of causation involves many sub-systems that vary across time and space. The efficacy of a treatment in a particular context may not predict its efficacy in another—a problem known as the "external validity" of experimental results. As in medicine, our capacity to predict causal effects outside the experiment ultimately rests on our "structural models," that is, our understanding of the channels of causation. For instance, even if we find that poor Kenyan children sleeping under insecticide-treated bednets are healthier, we do not recommend that poor Canadian children sleep under bednets. This is because we believe bednets improve health because they protect people against malaria-bearing mosquitoes, which are absent from Canada.

Channels are, by definition, endogenous. But this does not mean that they are not causal. To illustrate, imagine that a primitive tribe observes people entering their room in a recently built tourist hotel. They note that, most of the time, light appears in a room shortly after someone enters. To ascertain whether the effect is causal, they run an RCT and randomly assign people to rooms. Results confirms that people entering a room cause light to appear. They try this in their home village, but with no success. They conclude that their research lacks external validity because they do not know the channel of causation.

They go back to the hotel and note that two things typically happen before light appears: 1) the person says something—such as "it is so dark in here!"; and 2) they activate a switch in the wall. They do not know whether it is the words (the incantation) or the action (or both) that cause light to appear—or if the incantation is required before they can touch the switch. They run more RCTs—sometimes preventing people to speak when entering the room, sometimes covering the switch, sometimes both. From their research, they conclude that speaking is not necessary. But touching the switch is nearly always essential.[1] Based on there findings, they install switches in their huts—but again have no success in creating light.

This example illustrates several things. First, channels are causal even though they are endogenous: people get light in their room because they switch the light on. Being endogenous does not make this channel less essential in causing the outcome. Second, an easy way of verifying whether a channel is essential or necessary is to remove it and see what happens. In our example there are numerous channels of causation that are necessary: a switch connected to a lamp with a working bulb, itself connected to an electricity source that is operational, etc. All these elements are necessary: should one of them be missing, light will not materialize. If the researcher wishes to identify a sufficient cause, then the RCT must introduces a new channel. If the introduction of this channel causes the outcome, then this channel is sufficient—within the context of the experiment—even though it may require an endogenous behavioral response. If there exist alternative chains of causation, e.g., torchlight or candle, inference about causal channels is even more complicated.

Thirdly, it is possible to devise RCTs that examine each of the elements in the chain to isolate the necessary ones. But this search can only be fruitful if the researcher starts with a set of hypotheses that include the correct ones. Because channels of causation can be multiple and complex, relying on RCTs alone is an ineffective way to study causality. RCTs need to be complemented by theory, that is, by an intellectual understanding of the possible processes of causation. Furthermore, some hypotheses can be erroneous—e.g., light is caused by magical power—even though falsifying them would be difficult. It may be easier to reject these hypotheses on a priori grounds. To summarize, causation in economics cannot typically be understood equipped solely with RCTs and statistical inference.

PEER EFFECTS IN DIFFUSION NETWORKS

With these few words of caution, we are now in a better position to examine inference about peer effects in networks. I focus first on the estimation of peer effects in social networks connecting many individual agents. Throughout I conform to convention and sometimes call each agent a node, and a connection between two agents a link. A network is a collection of agents and the links between them. What the links represent depends on the topic of study.

1. Some tourists have a torchlight.

The estimation of peer effects raises a host of issues that have been extensively discussed in the literature—see Moffitt [2001] for an excellent review, and Bramoullé, Djebbari and Fortin [2009] for a clear exposition of the issues surrounding identification and estimation of peer effect models. Following this literature, we assume that the researcher is interested in estimating a network autoregressive model of the form:

$$\mathrm{Y}_i = a + \mathrm{X}_i\beta + \rho g_i \mathrm{Y} + \delta g_i \mathrm{X} + u_i \tag{2}$$

where Y_i is the decision or outcome of node i, X_i is a set of node-specific variables thought to influence Y_i, g_i is the vector of links between other nodes and i, and Y is a vector of Y_i over all observations. Parameter ρ captures so-called endogenous peer effects, that is, the influence that the outcomes of other nodes have on node i. Parameter δ captures so-called exogenous peer effects, that is, the influence that characteristics of peers have on node i. Model (2) focuses on diffusion effects—e.g., epidemic, information cascade, adoption of a new product. If Y_i is an outcome variable, ρ and δ measure network externalities. If Y_i is a choice variable, ρ captures strategic complementarity or strategic substitution effects.

Model (2) is closely related to the linear-in-means model which assumes that network effects depend not on the *absolute* number but on the *proportion* of adopting neighbors:

$$\mathrm{Y}_i = a + \mathrm{X}_i\beta + \rho \hat{g}_i \mathrm{Y} + \delta \hat{g}_i \mathrm{X} + u_i \tag{3}$$

with

$$\hat{g}_i \equiv \left[\frac{g_{ij}}{\sum_j g_{ij}} \right] = \left[\frac{g_{ij}}{d_i} \right]$$

where d_i is the degree (i.e., number of links) of noted i.

The choice between (2) or (3) depends on which type of diffusion process is most appropriate a priori: if diffusion occurs by contact (e.g., epidemic), model (2) is the obvious choice; if diffusion occurs through conformism or imitation, model (3) is more appropriate. Whether peer effects are endogenous or exogenous depends on the type of diffusion mechanism at work. To illustrate, suppose that Y_i represents adoption and X_i represents information. Exogenous peer effects corresponds to the case where adoption by i depends on whether i's neighbors are informed, irrespective of whether they adopt themselves. Endogenous peer effects arise if adoption by i depends on whether i's neighbors themselves adopted.

We now discuss how to obtain consistent estimates of models (2) and (3).

Identification and the reflection problem

In a widely cited paper, Manski [1993] discusses the estimation problems raised by the linear-in-means model (3). He considers a specific network structure in which people are divided into mutually exclusive, fully connected groups. He shows that, in this special case, parameters ρ and δ cannot be

separately estimated—they are not identified. He attributes this to the fact that what *i* does affects what *j* does and vice versa, and calls this the "reflection problem." Similar problems arise for model (2).

Moffitt [2001] illustrates the reflection problem with a simple example, which I reproduce here.[2] In this example each group has size 2. We have, in each group *g*:

$$y_{1g} = \theta_0 + \theta_1 x_{1g} + \theta_2 y_{2g} + \theta_3 x_{2g} + u_{1g}$$

$$y_{2g} = \theta_0 + \theta_1 x_{2g} + \theta_2 y_{1g} + \theta_3 x_{1g} + u_{2g}$$

Solving this system of simultaneous linear equations for the reduced forms yields:

$$y_{1g} = \alpha + \beta x_{1g} + \gamma x_{2g} + v_{1g}$$

$$y_{2g} = \alpha + \beta x_{2g} + \gamma x_{1g} + v_{1g}$$

where:

$$\alpha = \frac{\theta_0(1+\theta_2)}{1-\theta_2^2}$$

$$\beta = \frac{\theta_2\theta_3 + \theta_1}{1-\theta_2^2}$$

$$\gamma = \frac{\theta_2\theta_1 + \theta_3}{1-\theta_2^2}$$

We see that we cannot recover the four structural coefficients θ_0, θ_1, θ_2 and θ_3 from the three reduced form estimated parameters α, β, and γ. It is easy to verify that the same reasoning applies for larger group sizes. Moffitt, however, notes that if $\theta_2 = \theta_3 = 0$, then $\gamma = 0$. This means that parameter γ does identify the *combined* endogenous and exogenous peer effect. This is informative in itself.[3]

Several kinds of approaches have been offered to solve the identification problem in (2) or (3). One approach, which is ad hoc but has often been used in practice, is to omit one x_k variables from the list of exogenous peer effects. This yields an instrument $\hat{g}_i x_k$ for $\hat{g}_i \mathrm{Y}$ and permits identification. Unfortunately, it is often difficult to justify the exclusion of x_k from exogenous peer effects—unless this instrument has been engineered in an experimental or quasi-experimental way.

2. There is a small difference between Manski and Moffitt in that Manski includes *i* in the mean of *i*'s peers (reasoning that the mean is actually the expected behavior of the peers) while Moffitt does not (reasoning that *i* does not influence itself). As it turns out, this difference does not really matter in the group model—see Bramoullé, Djebarri and Fortin [2009].

3. Brock and Durlauf [2001] note that identification can be obtained if *y* is a dichotomous variable. In this case, model (3) is no longer a linear regression and identification is possible through the curvature of the likelihood function for *y*. Because identification is achieved through the curvature of the likelihood function, it is unclear how robust the method is in practice.

Manski and Moffit only consider mutually exclusive groups of equal size.[4] Identification is easier in more general networks. The approach, which has a GLS flavor, identifies endogenous $\rho\hat{g}_i\mathrm{Y}$ and exogenous $\delta\hat{g}_i\mathrm{X}$ peer effects by relying on the correlation of residuals. The intuition behind the approach is that if residuals are i.i.d., once we control for exogenous effects any remaining correlation in outcomes across individuals can only arise through endogenous peer effects. To illustrate, I show how the endogenous peer effect coefficient ρ can be estimated from the correlation of residuals. I express Y_i in deviation relative to the mean, which lets us ignore the constant term. For now I omit exogenous regressors. The simplified regression model (2) can then be written in matrix form as:

$$\mathrm{Y} = \rho\mathrm{GY} + u \tag{4}$$

where $\mathrm{G} \equiv [g_{ij}]$ is the network matrix. Equation (4) can be inverted to yield:

$$\mathrm{Y} = (\mathrm{I} - \rho\mathrm{G})^{-1}u$$

Forming the outer product of the above and taking expectations we get:[5]

$$\mathrm{E}[\mathrm{YY}'] = (\mathrm{I} - \rho g)^{-1}\Sigma(\mathrm{I} - \rho g)'^{-1} \tag{5}$$

where $\Sigma \equiv \mathrm{E}[u \cdot u']$ is the covariance matrix of the errors. If we are willing to impose sufficient structure on Σ it becomes possible to estimate ρ. For instance, if we assume that errors are i.i.d.. with variance σ^2, it follows that $\Sigma = \sigma^2\mathrm{I}$ and we get:

$$\mathrm{E}[\mathrm{YY}'] = \sigma^2(\mathrm{I} - \rho\mathrm{G})^{-1}(\mathrm{I} - \rho\mathrm{G})'^{-1} \tag{6}$$

This creates a relationship between the data—the outer product of the Y_i's in $\mathrm{E}[\mathrm{YY}']$—and the parameter of interest ρ. Given the network matrix G it is possible to recover ρ from the data using GMM or MLE.[6] In Appendix 1, we illustrate this observation with a simple example.

The approach can be generalized to include dependence on X_i and $g_i\mathrm{X}$. In matrix form, we get:

$$\begin{aligned} \mathrm{Y} &= \rho\mathrm{GY} + (\beta + \delta\mathrm{G})\mathrm{X} + u \\ \mathrm{Y} &= (\mathrm{I} - \rho\mathrm{G})^{-1}(\beta + \delta\mathrm{G})\mathrm{X} + (\mathrm{I} - \rho\mathrm{G})^{-1}u \end{aligned} \tag{7}$$

Here X and GX play a role similar to impulses in time series analysis.[7] Identification is achieved in the same manner as above. Assuming strict exogeneity of X, estimation can be achieved by applying MLE or GMM to (7).

4. Identification is in principle possible from data on fully connected groups of different size. This is because larger groups have stronger multiplier effects than small groups. This generates variation across groups that can be used to estimate ρ—see Bramoullé, Djebbari and Fortin [2009] and the references cited therein.

5. I have assumed that the network matrix g is non-stochastic. I return to this issue later.

6. This is not very different qualitatively from time series analysis where the chronological nature of the data is used to obtain an estimate of the autocorrelation coefficient ρ. Here, unlike in time series data, there is two-way influence, hence the square matrix.

7. This is because ρ "transmits" the effect of u across peers, while δ does not. It follows that any correlation in errors u_i between peers would constitute evidence of ρ-type of effects, i.e., endogenous peer effects.

The problem with the above approach is that we have to assume i.i.d.. errors to obtain a consistent estimator of the causal peer effect ρ. This ignores the possibility of contextual effects (as Manski calls them), that is, external or environmental influences common to a group or subset of agents. If contextual effects are uncorrelated with X, we can still use the above approach by expanding Σ to include group-specific random effects. If they are correlated with X, we can add a vector of group dummies to X. Both of these approaches, however, break down if the *u* are correlated across observations in arbitrary ways, or in the presence of heteroskedasticity of unspecified form. For these reasons, the literature has developed estimators of peer effect models that do not require assumptions about the specific form of Σ.

Causal inference with observational data

The problem with (7) is that stringent assumptions about the error structure are needed to estimate ρ. To overcome this problem, Bramoullé, Djebbari and Fortin [2009] propose an IV method that addresses the identification problem without making such assumptions. The canonical network autoregressive model is:

$$\mathrm{Y} = \rho\mathrm{GY} + \beta\mathrm{X} + \delta\mathrm{GX} + u \tag{8}$$

The idea is to use $\mathrm{G}^2\mathrm{X}$—the characteristics of the neighbors of the neighbors of *i*—as instrument for GY. This is similar to the concept of spatial lags used in spatial econometrics. Bramoullé, Djebbari and Fortin [2009] also extend the IV method to allow for sub-population fixed effects, and derive conditions under which this strategy is identified.

Spatial models

The approach can be generalized by borrowing from the literature on spatial autoregressive models ([Anselin [1988], Drukker, Prucha and Raciborski [2001]). This literature seeks to estimate a canonical model of the form:

$$\mathrm{Y} = \rho\mathrm{GY} + \beta\mathrm{X} + u \tag{9}$$

$$u = \lambda\mathrm{M}u + e \tag{10}$$

where *e* is i.i.d.. and everything is in matrix form. Matrix G represents the spatial or network dependence across outcome variables Y and M represents the spatial or network dependence across the errors. It is possible to assume that $\mathrm{G} = \mathrm{M}$, but it is not necessary. Note that, as written here, (9) does not include exogenous peer effects of the form GX. It immediately follows that endogenous peer effects GY can be instrumented using GX, so that the model is identified. More about this later.

In spatial data, matrices M and G can be formed of inverse distances, e.g., $\mathrm{M} = [m_{ij}]$ and $m_{ij} = \dfrac{1}{dist_{ij}}$ where $dist_{ij}$ is the distance between *i* and *j*. Sometimes these distances are truncated upwards, e.g., setting $m_{ij} = 0$ if $dist_{ij} > 500\ km$, for instance. M and G can also be contiguity matrices, e.g., $m_{ij} = 1$ if *i* and *j* share a border. Network data fit perfectly in this setup since a

network matrix basically constitutes a (non-Euclidian) contiguity matrix. Econometric methods developed to deal with spatial models with contiguity matrices are thus applicable to network autoregressive models.

To estimate the system (9) and (10), we must be able to solve the following reduced-form equations:

$$\mathrm{Y} = (\mathrm{I} - \rho\mathrm{G})^{-1}\beta\mathrm{X} + (\mathrm{I} - \rho\mathrm{G})^{-1}u$$

$$u = (\mathrm{I} - \lambda\mathrm{M})^{-1}e$$

which clearly requires 1) that $\mathrm{I} - \lambda\mathrm{M}$ and $\mathrm{I} - \rho\mathrm{G}$ are non-singular, so that they can be inverted; and 2) that the inverses are finite. The first requirement means that it is not possible for G and M to have l's everywhere, as would be the case if the network is complete. The second requirement puts a restriction on the size of the largest eigen value of $\mathrm{I} - \lambda\mathrm{M}$ and $\mathrm{I} - \rho\mathrm{G}$ so that the network multiplier effect does not "explode." This requirement ensures that the equilibrium value of Y is interior. It de facto rules out network interaction models with multiple equilibria, which cannot be estimated using this method.

We also note that:

$$\mathrm{E}[(\mathrm{G}y)u'] = \mathrm{G}(\mathrm{I} - \rho\mathrm{G})^{-1}\Omega_u \neq 0$$

where Ω_u is the covariance matrix of u. This is another way of saying that, in equation (9), the regressor GY is endogenous since it is correlated with u.

Estimation

One avenue for estimating model (9) and (10) is to assume that errors e are homoskedastic and distributed normally, and apply maximum likelihood. We have:

$$\Omega_u = \mathrm{E}[uu'] = \sigma^2(\mathrm{I} - \lambda\mathrm{M})^{-1}(\mathrm{I} - \lambda\mathrm{M}')^{-1}$$

Solving (9) and (10) for y in terms of e yields:

$$\mathrm{Y} = (\mathrm{I} - \rho\mathrm{G})^{-1}\beta\mathrm{X} + (\mathrm{I} - \rho\mathrm{G})^{-1}(\mathrm{I} - \lambda\mathrm{M})^{-1}e \tag{11}$$

This expression, which is similar to (7), can be used to construct a log-likelihood function.

Various 2SLS and GMM estimators have also been proposed to estimate regression model (9) using instrumental variables to deal with the endogeneity of GY. Kelejian and Prucha [1998], for instance, suggest using the following instruments: X, GX, $\mathrm{G}^2\mathrm{X}$ as well as MX, MGX, $\mathrm{MG}^2\mathrm{X}$. The authors suggest beginning with X, GX, $\mathrm{G}^2\mathrm{X}$ as instruments and applying standard 2SLS. A more efficient estimator can be obtained using the same instruments in a GMM setting.

As discussed earlier, Bramoullé, Djebbari and Fortin [2009] generalize this approach to allow for exogenous peer effects. Their model is of the form:

$$\mathrm{Y} = \rho\mathrm{GY} + \beta\mathrm{X} + \delta\mathrm{GX} + u$$

This means that now GX enters as a regressor directly into the model and cannot, therefore, be used as instrument for GY. Instrument $\mathrm{G}^2\mathrm{X}$ nevertheless remains, and this is the approach suggested by the authors. Building on the work of

Lee and Liu [2010], Liu *et al.* [2012] combine the approaches of Kelejian and Prucha [1998] with that of Bramoullé, Djebbari and Fortin [2009] into a GMM approach. They also propose a correction of the GMM estimator for the presence of small-sample bias.

Causal inference using experimental approaches

More convincing causal inference on network peer effects can be achieved using data from controlled or natural experiments. A growing number of studies have sought to resolve the endogeneity of $g_i\mathrm{Y}$ by using experimental methods. These studies can be roughly divided into two unequal groups: those, the most common, that "seed" the network exogenously with an experimental treatment to instrument Y but regard existing links g_i as given; and those, still rare, that seek to instrument g_i using experimental or quasi-experimental variation.

Seeding the network

Examples of estimation of peer effects using experimental data but existing links are found in Fafchamps and Vicente [2013], Giné and Mansuri [2011], and Fafchamps, Vaz and Vicente [2014]. Many of the estimation issues raised by such experiments are discussed in Baird *et al.* [2014]. In these experiments, a treatment T_i is introduced that has a direct effect on Y_i. This treatment can also have two kinds of indirect effects: a diffusion effect to individuals not directly treated but whose neighbors were treated; and a reinforcement effect on treated individuals whose neighbors were also treated. Variation in treatment across neighbors or in *i*'s number of neighbors can be used to obtain identification.

The reduced-form or intent-to-treat version of the model is of the form:

$$\mathrm{Y}_i = a + \mathrm{X}_i\beta + \theta\mathrm{T}_i + \rho' g_i\mathrm{T} + \delta g_i\mathrm{X} + u_i \tag{12}$$

The IV or structural version uses $g_i\mathrm{T}$ as instrument for $g_i\mathrm{Y}$ in (13):

$$\mathrm{Y}_i = a + \mathrm{X}_i\beta + \theta\mathrm{T}_i + \rho' g_i\mathrm{Y} + \delta g_i\mathrm{X} + u_i \tag{13}$$

Since $g_i\mathrm{T}$ is nothing but the number[8] of *i*'s neighbors who were treated, the results from the reduced form (12) and structural form (13) are typically very similar since variation[9] in the instrumented $\widehat{g_i\mathrm{Y}}$ is nothing but a multiple of variation in $g_i\mathrm{T}$. When both T and Y are dichotomous variables, applying IV to (13) using $g_i\mathrm{T}$ is the standard LATE estimator.

In practice, it is common for researchers to subsume contextual effects $g_i\mathrm{X}$ either in network/group fixed effects, or in individual fixed effects. With individual fixed effects the researcher needs data on Y_i from before and after treatment. With such data, we can estimate:

$$\Delta\mathrm{Y}_{it} = \beta\Delta\mathrm{X}_{it} + \theta\Delta\mathrm{T}_{it} + \rho\Delta(g_{it}\mathrm{Y}_t) + \delta\Delta(g_{it}\mathrm{X}_t) + \Delta u_{it} \tag{14}$$

8. Proportion, if we use the linear-in-means version of the model.
9. I mean the part of the variation that is uncorrelated with or not explained by X_i and $g_i\mathrm{X}$.

where Δ denotes the first-difference estimator, i.e., $\Delta z_{it} = z_{it} - z_{i,t-1}$ and $\Delta(g_{it}\mathrm{T}_{it})$ is used as instrument for $\Delta(g_{it}\mathrm{Y}_t)$. If the vector of contextual effects X_t does not change over time (e.g., because it only contains time-invariant characteristics) the first regressor drops out of (14). In addition, if g_{it} does not change over time either (e.g., because the researcher only collected the information at baseline), (14) simplifies to:

$$\Delta \mathrm{Y}_{it} = \theta \Delta \mathrm{T}_{it} + \rho g_i \Delta \mathrm{Y}_t + \Delta u_{it} \tag{15}$$

using $g_i \Delta \mathrm{T}_t$ as instrument. This requires that T_t changes over time, which is typically ensured by collecting data prior to treatment (i.e., baseline data) and after treatment (i.e., follow-up data), and setting treatment such that $\mathrm{T}_0 = 0$ for all and $\mathrm{T}_1 > 0$ for some randomly selected individuals.

If X_t does not change but g_{it} changes over time—and the researcher is willing to regard this change as exogenous—then (14) can be rewritten as:

$$\begin{aligned}
\Delta \mathrm{Y}_{it} &= \theta \Delta \mathrm{T}_{it} + \rho \Delta(g_{it}\mathrm{Y}_t) + \delta \Delta g_{it}\mathrm{X} + \Delta u_{it} \\
&= \theta \Delta \mathrm{T}_{it} + \rho[g_{it}\mathrm{Y}_t - g_{it-1}\mathrm{Y}_{t-1}] + \delta \Delta g_{it}\mathrm{X} + \Delta u_{it} \\
&= \theta \Delta \mathrm{T}_{it} + \rho[g_{it}\mathrm{Y}_t - g_{it-1}\mathrm{Y}_t + g_{it-1}\mathrm{Y}_t - g_{it-1}\mathrm{Y}_{t-1}] + \delta \Delta g_{it}\mathrm{X} + \Delta u_{it} \\
&= \theta \Delta \mathrm{T}_{it} + \rho[\Delta g_{it}\mathrm{Y}_t + g_{it-1}\Delta \mathrm{Y}_t] + \delta \Delta g_{it}\mathrm{X} + \Delta u_{it}
\end{aligned}$$

In practice, researchers often have only one data point for g_i. In this case, $\Delta g_{it} = 0$ and the above simplifies to (15). Contextual effects within locations can be corrected for by clustering standard errors by location.

Network sampling and measurement error

Chandrasekhar and Lewis [2012] discuss sampling issues surrounding the estimation of non-AR(1) reduced form model (12) that include indirect effects terms. Dropping exogenous peer effects for simplicity of exposition, the estimated model they investigate is of the form:

$$\mathrm{Y}_i = a + \rho g_i \mathrm{T} + \gamma h(\mathrm{GT}) + u_i \tag{16}$$

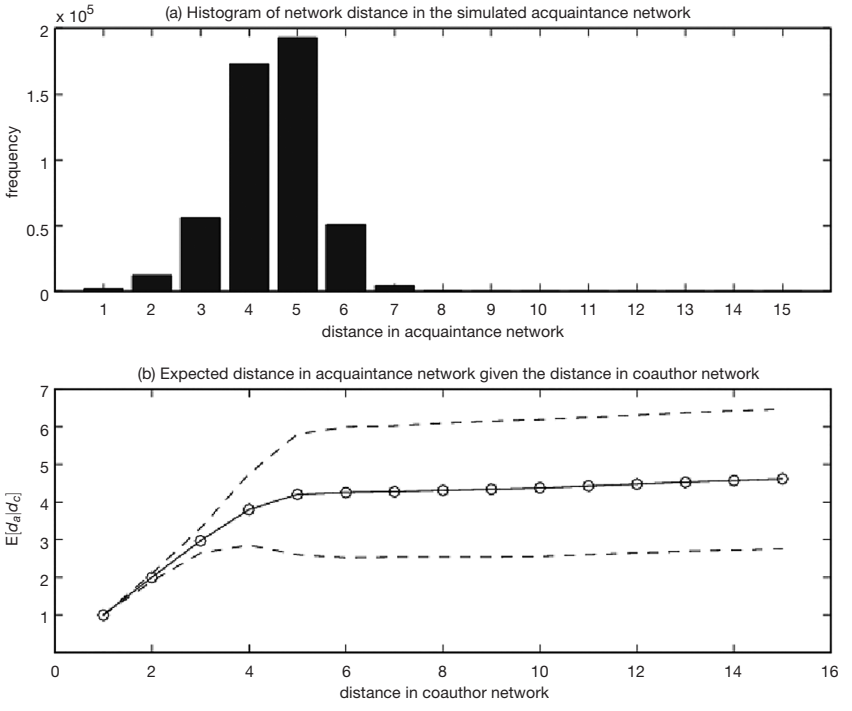
where $h(\mathrm{GT})$ is a vector of aggregate statistics on treatment network GT. In Banerjee *et al.* [2012]'s application of this methodology to the diffusion of microfinance, GT includes the network distance from i to the "seeded" or treated individual k, together with various measures of the centrality of k.

The issue that Chandrasekhar and Lewis focus on is the measurement error bias in γ that results from constructing GT statistics from a sample of the total network G. To illustrate the problem, imagine that we sample only 50% of the population and focus on the distance between i and k. Let the true distance be d_{ik}. In practice, the estimate of d_{ik} that we can construct from a 50% sample overestimates d_{ik} since we do not observe all the links in G. This measurement error in turn biases γ. Chandrasekhar and Lewis illustrate the size of the bias for different sampling ratios, and show that the bias gets large if the sampling ratio falls below 50%. They propose a bias correction methods that uses the information available from G and the sampling ratio to bracket what the true d_{ik} should be, assuming that G is a random sample of the total population.

This problem affects AR(1) network autoregressive models as well *except* if the experiment only treats individual in the sample. If other, unsampled individuals are treated too, there will be a bias in ρ. To illustrate, imagine that some locations are treated and others not. In treated locations, for each sampled individual who was treated, there is one unsampled individual who was also treated. Further imagine that each sampled individual *i* has exactly two friends, one sampled and one unsampled. We estimate the first difference model (15). The regressor of interest g_iT is the number of treated friends. By assumption the true value of g_iT in treated locations is 2 but the observed value is 1. In other words, g_iT is mismeasured to be exactly half of its true value. It follows that when we estimate (15) using available information, the coefficient estimate $\hat{\rho}$ will capture the effect of having two friends and thus will be twice its true value, i.e., we have $\hat{\rho} = 2\rho$. The fact that $\hat{\rho}$ is biased does not, however, imply that inference about the existence of peer effects is biased: if $\rho = 0$ then $\hat{\rho} = 0$ as well. Hence if we reject that $\hat{\rho} = 0$ it implies that ρ is also not equal to zero. This example can be extended to more complex cases: as long as $E[\hat{\rho}] > 0$ if $\rho > 0$ and vice versa, we can draw inference on whether $\rho > 0$ from $\hat{\rho}$ (see for instance Fafchamps and Vicente [2013]).

In a related vein, Fafchamps, Goyal and Van der Leij [2010] discuss how to conduct inference when the researcher only observes a subset of the links between individuals. In this case, the researcher observes the whole population but not all the links. This is a typical situation for secondary data on certain links, like

Figure 1. *Simulated acquaintance networks*



collaborations, citations, or phone calls: there are social interactions that are not observed by the researcher. In this case, the network distance in the observed network d_c is weakly larger than the distance in the denser true network d_a. If the observed network is embedded in the observed network, however, d_c provides some useful information regarding d_a. Since the observed network is a subset of the unobserved network, we must have: $d_a \leqslant d_c$. It follows that $\mathrm{E}[d_a \mid d_c]$ increases with d_c. In other words, d_c provides information about unknown d_a since the average value of unobserved d_a increases monotonically with observed d_c.

This is illustrated by Fafchamps, Goyal and Van der Leij [2010] with a simple computer experiment. We first generate a random "acquaintance network" of 1 000 nodes and 2 500 links between randomly chosen pairs of nodes. Figure 1(a) shows a histogram of this simulated acquaintance network. Next, let us randomly select 1 000 links from the "acquaintance network" to obtain an "observed network." As the observed network is a subgraph of the acquaintance network, the distance in the acquaintance network between two nodes is bounded from above by the distance in the observed network. We then analyze the relation between d_a and d_c in these simulated networks. Figure 1(b) shows the results. As expected, $\mathrm{E}[d_a \mid d_c]$ increases monotonically with d_c. Given that there is a monotonic relation between d_c and d_a, d_c is a valid proxy variable for d_a. The requirement is that d_c is not so much above the distribution of d_a that $\partial \mathrm{E}[d_a \mid d_c] / \partial d_c \to 0$. Above this point, d_c is no longer informative about $\mathrm{E}[d_a]$.

Identification and the speed and strength of diffusion

Paradoxically, causal peer effects may be difficult to detect if they are too fast or too strong. This is easily illustrated with the following example. Imagine that a fire breaks out in an office building. One person observes the outbreak and passes the information to her immediate office neighbors, who quickly pass it on to their own office neighbors, and so on. Within minutes everyone is informed and leaves the building. In this example diffusion is so rapid that an external researcher is unable to estimate regression (16). The reason is that Y_i—awareness of the fire—is rapidly common to all workers, irrespective of their local network g_i. Hence there is no correlation between Y_i and i's position in the network. Yet, it is obvious in the example that peers are essential to the diffusion of the information. This means that diffusion must be slow enough, relative to the type of data available, for it to be observed by the researcher. It is also true that, if diffusion is too slow relative to the researcher's time frame, network effects will also fail to be observed.

Even if diffusion is rapid, peer effects may nevertheless be detectable provided diffusion is incomplete. This arises if the network is split into distinct components or parts. To illustrate with the example above, imagine that workers on a different floor of the building are not made aware of the fire because there is no communication between them and the rest of the building. As a result they do not leave the building on time—an outcome that is observable. This shows that network effects in fast diffusion processes can be observed if the network is split into distinct components—as long as "seeding" is sufficiently uncommon so that some components are not seeded.

This is easily illustrated with a simple diffusion model. Let Y_{it} be individual i's awareness of the fire at time t. The more aware of the fire someone is, the

more likely they are to take evasive action, e.g., to leave the building. Peer effects are modeled as a reinforcing process of awareness that takes place over time: the more aware my network neighbors are of the fire, the more I become aware of it myself. Let $Y_t = [Y_{1t}, ...Y_{Nt}]'$ be the vector of awareness of all individuals in the network at time t. We have:

$$Y_t = \rho G Y_{t-1} + u_t \tag{17}$$

where u_t denotes a vector of external impulses, e.g., the "seeding" of information from an exogenous source. We first note that if $u_{it} = 0$ for all i and all t, $Y_{it} = 0$ is the equilibrium of the system: no awareness of a fire is an equilibrium if no one in the network has received any information about the fire. Now consider what happens when one individual receives a message u_i about an imminent fire, and this message remains constant over time (i.e., it is not a false alarm). How does awareness of the fire among the network evolve over time?

We first check whether process (17) has a finite resting point or equilibrium, that is, a point at which:

$$Y = \rho G Y + u \tag{18}$$

where $u = [0, ..., u_i, ...0]$. A finite resting point would mean that awareness about the fire does not "explode" in a fleeing frenzy. If such a point exists, it is given by:

$$Y = (I - \rho G)^{-1} u$$

A sufficient condition for a finite solution to exist is that ρ be smaller than the largest eigenvalue of G—which is for instance guaranteed if ρ is smaller than 1 over the maximum degree of any agent.[10] When ρ is large enough, an interior equilibrium does not exist: awareness spreads rapidly and in an exponential manner, and all individuals connected through the network become aware of the fire, which we define as reaching a value of Y about some threshold $\overline{Y}$.

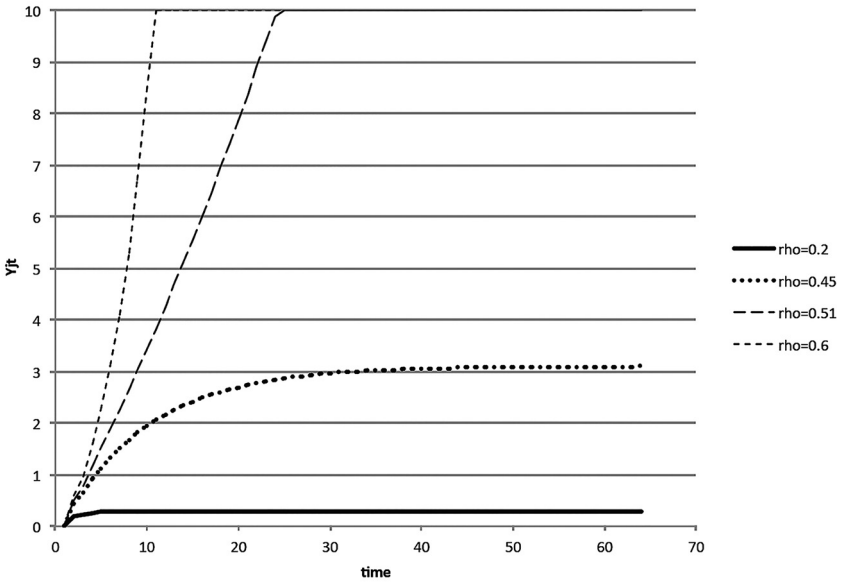
A high ρ also implies a rapid diffusion of the information, meaning that the awareness level of an arbitrary network member j rises rapidly over time and quickly reaches $\overline{Y}$. This is illustrated in figure 3 for different values of ρ for a fully connected network of size 3.[11] Since in this example the average degree is 2, a sufficient condition for Y to increase without bound is $\rho > 1/2$, which is indeed what we find. For $\rho < 1/2$, Y_{jt} converges to a constant value; for $\rho > 1/2$, it increases without bounds. We also note that increasing ρ above $1/2$ shortens the time it takes Y_{jt} to reach $\overline{Y}$—which we take to be the awareness level required to observe fire awareness, e.g., fleeing behavior.

If the researcher observes fleeing behavior at intervals longer than the time required for information to spread to the entire network—e.g., every $T > 20$ periods in figure 3 when $\rho = 0.51$—no information can be revealed about the precise value of ρ from detailed information about the initial impulse u_i and the structure of the network G: before the network is seeded with u_i, $Y_i = 0$ for all i;

10. This is related to the concept of Bonacich centrality in network theory. See Jackson [2009] for details.

11. We set $u_i = 1$ for one node i from $t = 1$ onwards, and we trace the evolution of Y_{jt} for one of the other two nodes. We set $\overline{Y} = 10$ and only report values of Y_{jt} up to $\overline{Y}$.

Figure 2. *Evolution of Y_{jt}*



and after awareness has spread, $Y_{jt} \geqslant \overline{Y}$ for all those connected to i in the network. In both cases, there is no variation in fleeing behavior across individuals that is associated with proximity to individual i, and hence no econometric identification of ρ. The only causal identification that the effect of seeding in disjoint components, that is, networks that are unconnected: unseeded networks remain at $Y = 0$ and they do not flee, while seeded networks quickly converge to $Y \geqslant \overline{Y}$ and they flee.

Creating exogenous links

As we have seen, estimating (13) by experimentally seeding the network with information (or another treatment) requires that g_i be exogenous to treatment. This requires either that information about g_i be collected prior to treatment, or that the network information collected after treatment be orthogonal to treatment—e.g., family ties.[12]

Even so, we cannot completely rule out the possibility that inference is affected by network self-selection. To see why, imagine that we "seed" farmers with information about a new fertilizer. We observe that farmers A and B, who are network neighbors, both adopt although only farmer A was informed about the new technology. It is tempting to conclude from this that A passed the information to B and thus that we observed network diffusion. We cannot, however, rule out the possibility that A told everyone about the new fertilizer but only his friend B adopted because both A and B are "modern" farmers interested in new technology, and this is why they are friends in the first place. Here susceptibility

12. Even here, though, recall bias may generate measurement error that is correlated with treatment if people remember better those family ties that received treatment.

to treatment is correlated with network proximity: people for whom the effect of treatment is large (small) are linked to others for whom the effect of treatment is also large (small). Hence we observed correlated adoption along network neighbors as a result of treatment—even though there is no diffusion along social networks.

Only exogenously assigned links can solve the problem of network self-selection in a fully convincing way. A number of papers have used random assignment of students to teams to identify peer effects. In a seminal study, Sacerdote [2001] uses random assignment of roommates and dormmates at Dartmouth College to demonstrate social network effects on grade point average (GPA) and joining social groups. Lyle [2007], [2009] and Shue [2011] use similar strategies. Fafchamps and Quinn [2013] organize a randomized field experiment that combines a link formation treatment with a information seeding treatment.

To illustrate, consider the following diffusion model:

$$\mathrm{Y}_{it} = a + \mathrm{X}_i\beta + \rho g_i \mathrm{Y}_{t-1} + \delta g_i \mathrm{X} + u_{it}$$

where peer effects are assumed to take time to materialize, hence the lag on Y_{t-1}. A treatment τ_{ij} aims to create new links between i and j such that τ_{ij} is a valid instrument for g_{ij}. Hence we can instrument g_{ij} using τ_{ij}, and use $g_i^p \mathrm{Y}_{t-1}$ as instrument for $g_i \mathrm{Y}_{t-1}$, where g_{ij}^p is the value of g_{ij} predicted by τ_{ij} and where $g_i^p \equiv [g_{i1}^p, ..., g_{in}^p]$. Since X_i and $g_i\mathrm{X}$ are by design orthogonal to the instrumented value of $g_i \mathrm{Y}_{t-1}$, they can be ignored from the regression and subsumed in the error term. We can therefore estimate a model of the form:

$$\mathrm{Y}_{it} = a + \rho g_i \mathrm{Y}_{t-1} + u_{it}^{'}$$

using $g_i^p \mathrm{Y}_{t-1}$ as instrument. This regression model basically estimates the extent to which a treatment that induces people to link also induces them to change their behavior or outcome Y_{it} in a way that is consistent with diffusion along newly created network links. The corresponding intent-to-treat or reduced form version is:

$$\mathrm{Y}_{it} = a + \rho' \tau_i \mathrm{Y}_{t-1} + v_{it}$$

Exclusion bias

Guryan, Kroft and Notowidigdo [2009] argue that exogenously assigning individuals to groups generate a specific *negative* bias in the estimation of peer effects. They estimate a linear-in-means model of the form:

$$\mathrm{Y}_i = a + \rho \hat{g}_i \mathrm{Y} + u_i$$

where $\hat{g}_i \mathrm{Y}$ represents the average of variable Y for those in the same group as i, where group membership $\hat{g}_i$ is randomly assigned within a given experimental population. The authors show that this framework mechanically results in a negative bias in ρ. In their paper Y_i is performance in a golf tournament and u_i is unobserved ability. The intuition is that high ability golfers are matched with individuals who are, on average, of lower ability while low ability golfers are matched with individuals who are, on average, of higher ability. This generates a mechanical *negative* correlation between u_i and $\hat{g}_i \mathrm{Y}$. Caeyers and Fafchamps [2015] call this an exclusion bias, and examine the nature of the bias in detail and

use simulations to quantify the magnitude and sign of the bias with and without different types of fixed effects.

One remedy to exclusion bias is to include a proxy for the average ability of the peers—e.g., based on their performance in earlier tournaments. Because the bias is mechanical, it is also possible to conduct consistent inference about the existence of peer effects by using bootstrapping to correct *p*-values instead. This approach, discussed in detail by Caeyers and Fafchamps [2015], works as follows. Let the model to be estimated be written:

$$Y_i = \beta \sum_{j \in C_i} Y_j + \mu_s + \varepsilon_{is} \tag{19}$$

where C_i is the (randomly assigned) neighborhood of i and μ_s is an experimental session fixed effect for individual i. We wish to test the null hypothesis that $\beta = 0$. The distribution of $\hat{\beta}$ under the null can be simulated as follows.

- Take the pool of subjects assigned to treatment and, within each pool (e.g., an experimental session), randomly reassign subjects to new "lacebo groups" P_i.
- For each subject, generate a new $\sum_{j \in P_i} Y_j$. By construction the only source of correlation between $\sum_{j \in P_i} Y_j$ and Y_i is the exclusion bias.
- Estimate (19) using $\sum_{j \in P_i} Y_j$ in lieu of $\sum_{j \in C_i} Y_j$ and store the estimate of $\hat{\beta}_p$.
- Repeat this a large number of times. The resulting frequency distribution of $\hat{\beta}_p$ simulates the distribution of $\hat{\beta}$ under the null hypothesis that $\beta = 0$. This distribution is not centered around 0 since $\sum_{j \in P_i} Y_j$ is negatively correlated with Y_i as a result of random selection without replacement.
- Use the simulated distribution of $\hat{\beta}_p$ to compute a corrected *p*-value for test that $\beta = 0$ in regression (19).

Deleting links

Experiments aimed at creating new social links ultimately seek to identify sufficient causal processes. Other researchers have documented what happens when social links are removed. This identifies necessary causal processes.

Mevlude and Yuksel [2011], for instance, examine the long-term consequences of Jewish expulsions in Nazi Germany on the educational attainment and political outcomes of German children. They find that they had significant detrimental effects on the human capital and political development of Germans who were at school-age during the Nazi regime. This suggests that the availability of qualified teachers and professors is necessary in order to educate the population, in the sense that an exogenous reduction in the supply of teachers and professors results in lower education levels.

In a similar vein, Patnam [2011] identifies the effect of corporate networks on firms' financial investment and executive pay decisions. The idea behind the paper is that board interlocks provide a channel for non-market interactions amongst firms. Using panel data for all publicly traded companies in India,

Patnam estimates peer effects in firm policies. Identification is achieved by exploiting exogenously changes in board interlocks caused by the death or retirement of shared directors. Significant network peer effects are found that are positively associated with firms' investment strategy and executive compensation.

This paper is particularly relevant to our purpose because board interlocks are, by nature, endogenous: they result from a choice made by firms to share board directors. Even though board interlocks are created endogenously, they may end for exogenous reasons. Patnam's paper thus illustrates how it is possible to identify the causal effect of endogenous social links by studying what happens when they are abruptly severed.

CAUSAL NETWORK EFFECTS REVISITED

The literature on peer effects that we have summarized so far is extensive and varied. But it is all about the diffusion model. The logic of this model is borrowed from epidemiology: a social network is given exogenously, and a diffusion process takes place over it. Although it is possible to construct microeconomic foundations for it (e.g., Liu *et al.* [2012]), the model does not allow any link formation or deletion. The network has to be exogenous. Any suspicion of endogeneity is seen as threatening claims about causal peer effects.

As a result the diffusion framework is ill-suited to the study of many interesting phenomena in which networks play an important role in the causal chain. To illustrate, imagine someone is told of a new job opportunity in their favorite ice cream shop. They call different people and find out how to apply. In the end, they get the job. Would we say that because applying for the job was endogenous, it did not have any causal impact on getting the job? Clearly not: applying is an essential element of the causal chain—like the light switch in our earlier example. The same is true of the different social contacts that were created to locate the information about how to apply: these social links were initiated by the job seeker and are thus endogenous; but they were essential in obtaining the necessary information. These endogenously created links are part of the causal chain, they are a channel by which the job was obtained. The fact that social contacts are endogenous does not imply they can be ignored in policy design. Especially, removing the possibility of social contact would, in this example, result in the job going to someone else.

Dynamic peer effects

Comola and Prina [2014] offer an elegant way of thinking about the causal role of endogenous network effects in an experimental setting. They consider a randomized controlled trial in which treated individuals are located in a baseline social network $\mathrm{G}_0 \equiv [g_{io}]$. The researchers observe an outcome at baseline and endline, denoted Y_0 and Y_1, respectively. They also observe the social network $\mathrm{G}_1 \equiv [g_{i1}]$ after treatment. They write the diffusion model in the following form:

$$\mathrm{Y}_{i0} = a_0 + \rho g_{i0}\mathrm{Y}_0 + \mu_i + e_{i0}$$
$$\mathrm{Y}_{i1} = a_0 + a_1 + \rho g_{i1}\mathrm{Y}_1 + \beta \mathrm{T}_{i1} + \delta g_{i1}\mathrm{T}_1 + \mu_i + e_{i1}$$

where $T_{i1} = \{0,1\}$ denotes treatment (set to 0 at baseline) and μ_i denotes an individual fixed effect. As before ρ captures endogenous peer effects while δ measures exogenous peer effects.

If the social network matrix is time-invariant so that $G_1 = G_0$ the diffusion model boils down to the standard:

$$\Delta Y = a_1 + \rho G \Delta Y + \beta T + \delta G T + \Delta e$$

which can be estimated using the approach discussed earlier.

In what Comola and Prina call a dynamic peer effect model, part of the causal effect of treatment is achieved through a recomposition of the social network. We have:

$$\Delta Y = a_1 + \rho G_0 \Delta Y + \rho \Delta G Y_0 + \rho \Delta G \Delta Y + \beta T + \delta G_0 T + \delta \Delta G T + \Delta e$$

There are three types of endogenous peer effects in this model: $G_0 \Delta Y$ (our earlier endogenous peer effect term), $\Delta G Y_0$ (the network recomposition effect), and the combined effect $\Delta G \Delta Y$. Comola and Prina note that the three types of endogenous peer effects are typically correlated. As a result, ignoring the change in network structure ΔG and dropping the $\Delta G Y_0$ and $\Delta G \Delta Y$ terms results in a biased coefficient ρ. The authors build on Bramoullé, Djebbari and Fortin [2009] and Liu *et al.* [2012] to construct suitable instruments and apply the model to empirical data, allowing ρ and δ to vary across terms. They find that a large share of the peer effects ρ occurs through a recomposition of the network—mainly through the $\Delta G \Delta Y$ term. This suggest that a significant share of the peer effect results from a recomposition of the network to adjust to the change in outcome. Put differently, people form new links to individuals whose outcome has been changed as a consequence of treatment—and they drop links to people whose outcome has not changed—and part of the effect of treatment is mediated through this network recomposition.

To illustrate this idea with an example, imagine a treatment that induces people to adopt a new practice. This treatment effect is magnified by network effects, i.e., people are more likely to adopt if their network neighbors adopt as well. Dynamic peer effects arises when people drop links with non-adopters, create links with adopters, and this leads them to adopt even more. This situation would naturally arise for network goods, that is, goods with strategic complementarities between neighbors. Typical examples include software products whose usefulness increases with the number of using neighbors. In this case, it is rational for people to rewire their network to increase proximity to other adopters. This is an endogenous effect in the sense that it is correlated with own adoption. But it is still a channel through which peer effects magnify the effect of treatment: if people were unable to rewire their network, the total effect of treatment on adoption would be smaller.

Cause or means to an end

Improper inference about peer effects can arise from a fundamental imprecision regarding what a link is. To illustrate, imagine I survey graduate students and ask them, "If you have a question regarding econometrics, who do you turn

to?" I collect their answers and subsequently obtain information about their mark in econometrics. I regress their mark on the type of link they listed. I find that those who listed the econometrics professor obtain higher marks. Is this evidence of a causal effect of the link on exam performance?

There are many potential problems with this example, e.g., links could be correlated with interest—students gifted in econometrics share a common interest with the econometrics professor and are therefore more likely to talk to each other. Another source of difficulty that is often overlooked in empirical work is that people report certain links more than others, or that they report links that do not yet exist but could easily be created. Suppose the latter. In this case, it is conceivable that students who are more confident in their capacity to articulate an econometrics question to the professor are also more likely to list the professor as possible source of information. Here we have a case of response bias—the link does not actually exist in the way the researcher thinks about it—and this bias is correlated with unobserved student ability. In fact, it is probably the case that all students can in principle see the professor, but only some students decide to avail themselves of this opportunity. In this case, the correlation between exam mark and listing the professor says nothing about the importance of network effects—everyone has access to the professor, there cannot be network effects in the sense that those with a link have an advantage—but the correlation says much about student heterogeneity.

Comola and Fafchamps [2014b] propose a methodology for testing whether discordant responses given to the above question are better interpreted as links or as willingness to link. Suppose we use their methodology to test whether students listed an actual link to their econometrics professor, or a willingness to link with the professor should a question arise. Further, suppose that we conclude it is willingness to link. It follows that we have no evidence of a causal effect of the reported "link" on exam performance: all students could approach the econometrics professor if they wanted to, but some are reluctant to do so for reasons that are partly correlated with ability.

Does this imply that linking to the professor is unimportant for exam performance and that we should, therefore, cancel office hours? No, because linking to the professor is the means by which some students achieve a good outcome. Put differently, social links often are the channel by which peer effects are achieved. The ultimate cause of the peer effect is the choice to link, but the peer effect could not be achieved if linking was made impossible. In other words, social links can be the essential channels by which diffusion of information and practices is achieved even though they do not cause diffusion per se.

There are many examples of diffusion that fit this description. Think of Granovetter's [1995] job search example: people activate their network to find job information and do so until a job is found. People who have neglected to build links with others may find it harder to find a new job, but their lack of network may be a consequence of their own earlier choices, which could be driven by their preferences—which in turn could be partly determined by their genes, which in turn could be the consequence of human evolution through mate selection, etc. The search for a cause that is not "endogenous" is largely futile.

Encouraging link formation

Fafchamps and Quinn [2013], [2014] push this idea one step further. They use data from two field experiments in which entrepreneurs are given an opportunity to form new acquaintances with a few other entrepreneurs. They then examine whether entrepreneurs who are given this opportunity are more likely to emulate each other's business practices. They also investigate whether this effect is stronger for new acquaintances predicted by homophily compared to acquaintances that are not. The logic is that people naturally react to an opportunity to make new friends in a small crowd by congregating with people that resemble them—e.g., same age, gender, ethnicity, religion, etc. There is a large literature that documents the fact that nearly all social networks of friendship and acquaintance display a strong element of homophily (Jackson [2009]). One can therefore interpret homophilous links as those that are more likely to happen by default, without specific purpose or goal—other than the pleasure of socializing with people sharing similar interests and background. Non-homophilous links, therefore, are more likely to result from purposive link formation—such as links one would form to access new information or opportunities. Fafchamps and Quinn [2013] test whether diffusion is associated more with homophilous or non-homophilous links. They find some evidence that small entrepreneurs who link with large entrepreneurs are more likely to emulate them than entrepreneurs of similar size.

While the evidence the authors are able to provide remains impressionistic only, the approach illustrates the idea that peer effects may be stronger for links that are formed precisely to access them. This is logical: if agents form certain links to access information and benefit from network externalities, it is natural to expect these links to channel more information and to generate stronger externalities—and hence to have stronger peer effects.

In the limit, it is even conceivable that peer effects arise only through purposively formed links. To illustrate, imagine that a new online game is introduced. Consider an online gamer who is currently not linked to anyone playing the new game. To play the game, the gamer has to form new links to others interested in playing the new game. The fact that these links can be created is essential in the adoption decision: if such links could not be made, the gamer could only play alone and game would have little or no value. In this example, rewiring the network is essential to capture network externalities. Peer effects are a characteristic of the online game that are manifested by the formation of new links among adopters. In this example, purposive link formation is evidence of peer effects, and is an essential ingredient to diffusion.

Purposive link formation

This leads us to wonder whether most social links observed in practice result from a purposive decision process—or arise serendipitously. If economic agents form links to achieve a self-serving purpose—e.g., to access relevant information about jobs, technology, and other market-relevant opportunities—then we expect social links to be created in a way that favors efficiency. Full efficiency may not be achieved if agents do not internalize the externalities they

create for others, but social networks should maximize the individual benefits from link formation. In contrast, if people form links for reasons other than the maximization of self-interest, social networks need not circulate information and other benefits in an efficient manner. It is therefore of considerable interest to study the process by which social links are created in practice.

In the last couple of years, I have, in collaboration with various coauthors, initiated a number of experiments aimed at understanding the motives people have when forming social links that can generate economic gains for themselves and others.

There exists a considerable literature documenting other-regarding preferences in experimental games (Charness and Rabin [2002]). Experimental subjects often make choices that are consistent with altruism, envy, or both (e.g., Fehr and Schmidt [1999]). What is less clear is whether other-regarding preferences also affect link formation. Belot and Fafchamps [2014] conduct a dictator-type experiment in which subjects determine the payoffs of four players—their own, and that of three other players. In one treatment the choice is couched in the form of an allocation process: the subject chooses between two payoff vectors for four players. In another treatment, the choice is framed as a partner selection process whereby the two unselected partners must team up with each other. Payoffs are determined by partner selection. Payoff vectors are identical across the two treatments. We show that subjects with a higher endowment display significantly more altruism in the first than in the second treatment, while subjects with a lower endowment appear to act more spitefully in the partner selection treatment. Partner selection thus appears not to bring out the best in people.

Davies and Fafchamps [2014] implement a simplified version of the gift exchange game of Brown, Falk and Fehr [2004] in an anonymized setting. The game is couched as an employer-employee contract and played by Ghanaian university students. Various treatments are investigated including relational contracting, reputation sanctions, praise and criticism, and worker non-enforceable promises. While Brown, Falk and Fehr [2004], [2012] have found evidence that relational contracting and reputational sanctions discipline workers and increase efficiency, we find considerable under-performance or shirking by subjects playing the role of employee, and persistent loss-making wage offers by subjects acting as employees. These findings are largely insensitive to treatment, and outcomes are inefficient (effort is too low) and inequitable (employers receive near-zero payoffs while employees enjoy large positive gains). Reciprocity has often been thought to be the cornerstone of social networks of favor exchange (e.g., Bloch, Genicot and Ray [2008], Jackson, Rodriguez-Barraquer and Tan [2012]). These findings suggest that reciprocity may not be natural in an anonymous setting. It follows that the formation of social links to serve as favor exchange conduit need not be as easy as often anticipated. A social context may be necessary to trigger social norms that limit cheating.

If economic agents are weary of cheating and abuse by others, this may deter link formation. To investigate this possibility, Fafchamps and Hill [2014] construct an anonymous, zero-feedback experiment in which subjects increase their payoff by joining a group, but also expose themselves either to the destruction or to the appropriation of their payoff by others. In a third treatment, people who join a group can give part of their payoff to others. The experiment was conducted in the UK, Kenya and Uganda. We find that African subjects steal

less and destroy no more than UK subjects. But the fear of payoff destruction acts as a deterrent to group formation, especially among African subjects. We also find that UK subjects give less than African subjects, but are also less likely to join a payoff-increasing group in the giving treatment. These results suggest that the fear of stealing and destruction by others can discourage link formation, more so in some populations than in others. Attitudes towards anonymous giving also affect willingness to form links, with differences across populations as well. These findings suggest that efficiency-enhancing links need not be formed.

To investigate this issue further, Comola and Fafchamps [2014a] use observational data to test whether self-reported favor exchange links are best understood as resulting from unilateral or bilateral link formation. Bilateral link formation corresponds to the concept of mutually agreed contract or arrangement—as for instance assumed in the seminal papers of Coate and Ravallion [1993] and Kocherlakota [1996]. Unilateral link formation is more akin to the idea of social norms of reciprocity making it difficult to refuse demands to help—and offers to help. Using data on favor exchange from Tanzania and India, we find evidence suggesting unilateral link formation: many links exist—and transfers take place—that only benefit one side. Using detailed transfer data from Tanzania, De Weerdt and Fafchamps [2011] examine whether transfer patterns between villagers are consistent with the constraints imposed by rational reciprocity. They find that they do not. In the Philippines, Fafchamps and Gubert [2007] similarly find that most transfers between households take place among relatives and neighbors.

When compared to our experimental findings, the results suggest that the formation of efficiency-enhancing links—such as favor or gift exchange—may need a social context that triggers either feelings of altruism or norms of reciprocity. To investigate this possibility, Caria and Fafchamps [2014a] run a link formation experiment with Indian farmers. The experiment is designed as a treasure hunt: one subject receives information about a treasure that generates a positive payoff. All other subjects who have access to this information also receives the treasure. In one treatment, forming a link to another subject gives access to all the information this subject has. In another treatment, forming a link with someone gives that person all the information the subject has. In the first treatment, it is optimal to link to the person with the most information. In the second treatment, it is efficient—but not privately beneficial—to link to someone who can pass the information to the most subjects. We find that participants understand the game but systematically depart from making efficient choices, either for themselves or for others. When we introduce an artificial identity treatment, we find that it distorts link formation—subjects prefer to link to their own group even if it lowers their individual payoff. Aggregate efficiency, however, is unaffected—probably because it is already low even without the identity treatment. What this suggests is that human subjects are not particularly good at forming links that maximize either their individual payoff or the payoff of the entire group.

To investigate this issue further, Comola and Fafchamps [2014b] conduct an experiment in which participants can form links with two other subjects. Links generate different payoffs for each player. Side transfers are not allowed. The experiment allows for sequential offers and counter-offers, in a way that mimics

a matching algorithm. Payoff matrices are selected such that, if subjects follow myopic best response, rapid convergence to a unique strongly pairwise stable (SPS) equilibrium is guaranteed. The experiment is conducted with French university students. Even though the game is complex and involves multiple offer rounds, we find little departure from the strongly pairwise stable equilibrium and high levels of aggregate efficiency. Departure from the SPS appears to be primarily due to satisficing behavior: some subjects stop making offers too soon. We also find that more information about others' payoffs speeds up convergence. These findings suggest that the human mind is well attuned to competition over partners, so much so that the theoretically predicted equilibrium is reached in nearly all games and sessions—an outcome rarely observed in laboratory games.

Taken together, this evidence suggests that link formation is only partially driven by material gain. Other considerations interfere with the decentralized formation of efficient networks, such as the (justified) fear of being cheated or left out, adherence to (even arbitrary) social identity, and satisficing behavior. Given these constraints it is probably unsurprising that empirical social networks center around shared genes and shared life experiences. In fact, we typically think negatively of individuals who befriend others only when they expect to derive a personal gain. Social opportunism is frowned upon, and people are often asked to demonstrate their absence of social opportunism by incurring costs and undertaking tasks that a purely selfish individual would not do.

This means that humans, as social animals, are not always encouraged to form social links in a purposive, self-interested manner, as a response to changes in external incentives. We are expected to form durable bonds with carefully selected individuals whose help can be mobilized in response to changes in circumstances. If we unilaterally form new bonds in response to external incentives, we are supposed to form them at least in part for altruistic reasons, not purely out of self-interest. In contrast, if we form bonds by mutual agreement, competition with others ensures that formed links are mutually beneficial. In such environment, self-regarding behavior appears to be less problematic, and more efficiency is achieved in experimental conditions. One possible interpretation of this finding is that the human mind is naturally attuned to market exchange, a conclusion that reinforces other market experiments such as those pioneered by Vernon Smith.

Strategic interactions and multiple equilibria

Finally, there are situations in which trying to identify *causal* peer effects is misguided. To illustrate, imagine a short race in which competitors are randomly assigned to teams of two players and must drive a car as quickly as possible between two points. Clearly, they cannot both drive the car at the same time. Since it is a short race, switching drivers is probably not a good idea. Hence it is optimal for one of the competitors to drive and for the other to serve as navigator. Should we infer from this evidence that one player's driving *causes* the other player's to navigate?

Causal inference is the wrong question here: we have multiple equilibria but, because of strong strategic substitution effects, they all have the common

feature that players specialize, i.e., their actions are negatively correlated. This is a consequence of the game structure which creates interdependency between the actions of the agents; it is not driven by causal peer effects.

The example can be expanded to include strategic complements as well. Imagine that each competing team can choose to drive in different cars but they win if *both* team members cross the finish line sooner than all other pairs of competitors. This rule implies that it is optimal for teams to ride in the same car. Hence if they have the choice between different cars, we will observe them getting into the same one: because the rules of the game create strong strategic complementarities, their choice of car will be the same even though, if all cars are equivalent, which car they pick does not matter. Should we infer from this evidence that one player's choice of, say, car 27 *causes* his team player to choose the same car? No: players choose to ride in the same car—in whichever way they decide—because of the incentive structure of the game.[13]

Causal inference is not a particularly interesting question here: there are multiple equilibria with correlated actions. How players choose the car is largely irrelevant: they could talk about it beforehand, they could agree beforehand that one of them will choose for both of them, they could argue over it at the start of the race, it does not matter. The outcome is identical, i.e., they ride in the same car.

What this example illustrates is that evidence about the existence of strategic complements and substitutes can be extracted from the pattern of correlation between choices made by players who are close to each other—without trying to second-guess which of the two influenced the other. If behaviors are strategic substitutes, behaviors will be more negative correlated between individuals who are close to each other, i.e., capable of influencing each other's strategic incentives. Similarly, if behaviors are strategic complements, they will be more positively correlated between nearby individuals. How convincing such evidence is depends on the extent to which correlation in contextual effects could account for the observed patterns.

Fafchamps and Söderbom [2014] use this approach to study the adoption of business practices in Ethiopia and Sudan. They conclude that some business practices are probably characterized by strategic substitution effects—e.g., free riding on vocational training offered by a competitor. In a different context, Fafchamps and Quisumbing [2003] use a similar approach to study specialization in household chores. They find a strong gender divide in the allocation of tasks, but also evidence of strategic substitutes in assignment to task within gender group. They argue that this suggests the presence of returns to specialization—a bit like driving a car: some things are better done when only person does it, even if it does not matter who does it. In the same spirit, Van den Boogaart, Fafchamps and Söderbom [2014] examine the diffusion of mobile money in an African country, looking for tell-tale signs that diffusion is driven by spreading information about the existence and reliability of the new service. They conclude that the diffusion pattern is more consistent with the presence of network externalities.

13. An astute observer may note that it should be possible to organize an RCT whereby one player is assigned to a car, and observe whether this assignment "causes" the team member to select the same car. Finding this to be the case, the observer may correctly conclude that assigning a driver to, say, car *c* causes the team member to select *c* as well. This experiment is not by itself revealing of the true cause of coordination, which is the incentive structure of the game.

CONCLUSIONS

In this paper I have sought to clarify some of the many issues surrounding the estimation of causal effects in social networks. Most of the current econometric literature on this issue focuses on the estimation of an AR(1) network autoregressive model thought to arise from network externalities or strategic complementarities. In this model, diffusion takes place over an existing network, with no formation of new links.

The literature has made considerable progress on how to obtain consistent estimates of this model from observational and experimental data. I discussed many of the issues that arise, such as identification, the role of instrumental variables, and the usefulness of experimental approaches. Next I discussed seeding experiments whereby the researcher introduces information into an existing network and observes its diffusion. While such experiments offer much promise, they are not a magic bullet. The literature has identified a number of sources of bias. I note that the effect of sampling bias varies with the type of model being estimated. Observation bias can arise as well, and I discussed how this type of bias can be circumvented in some cases.

Link self-selection bias has received more attention in the literature, and has been identified as a potential source of bias in many empirical papers seeking to estimate causal network effects. The proposed solution to this problem is to conduct link formation experiments. Such experiments are susceptible to a specific source of bias, called exclusion bias. This bias can fortunately be corrected, but has been largely ignored in empirical analysis of network effects relying on exogenous link formation. I also discussed whether link formation experiments seek to identify sufficient or necessary causes: sufficient causes can be identified with experiments seeking to create new links; necessary causes can be identified by experiments that destroy existing links. The distinction has not been fully recognized in the literature, yet it has important implications for policy design.

Once we allow for the possibility that social networks recompose over time, the notion of a causal network effect becomes elusive: what if a network effect is achieved by creating new links, or by using existing links in new ways—i.e., what if the network effect is endogenous? Does this make it less causal? Using simple examples, I argued that a network effect can be both causal and endogenous—and in fact that many factors that we regard as causes are endogenous in practice. For instance, we all understand that turning on a light switch is what causes light to appear in a room. This is true even though the act of turning on the switch is endogenous, in the sense that it is the result of purposive action on our part.

This led me to discuss the existing evidence regarding the nature of the link formation process. Using results from several recent experiments, I argued that people do not always form social links in the purposeful way that is often assumed by economists. Much link formation appears to be serendipitous—based on shared genes, identity, or life experiences, and less on self-interest or aggregate efficiency concerns. When link formation is unilateral, experimental subjects seem to experience difficulty in forming efficient networks, and their actions are partly driven by other-regarding preferences. Subjects do better when link formation is the result of a competitive process of offers and counter-offers, as typically happens in market exchange. I ended by observing that, in the presence

of strategic interaction and multiple equilibria, (negatively or positively) correlated actions should arise out of the structure of incentives, not because the action of one person "causes" others to behave in a certain way. In these cases, correlated actions are caused by incentives, not by peer effects. Many network effects studied in the empirical literature have this character, and can in principle be studied by looking at the correlation pattern directly.

For lack of space I have left out a number of important issues, such as the process by which economic agents influence others, the shape of the influence function (e.g., Jackson [2009]), and the possibility of saturation effects (e.g., Baird *et al.* [2014]). Even with these omissions, this rapid overview has illustrated that the study of causal effects on social networks is a challenging topic. Much progress has been realized in recent years, but much remains to be done. Hopefully this papers helps to show the way forward.

REFERENCES

Altmann, S., Falk, A. and Huffman, D. B. [2010]. "Implicit Contracts, Unemployment, and Labor Market Segmentation," *IZA Discussion Papers*, 5001.

Anselin, L. [1988]. *Spatial Econometrics: Methods and Models*. Berlin: Springer.

Arcand, J.-L. and Fafchamps, M. [2012]. "Matching in Community Based Organizations," *Journal of Development Economics*, 98 (2): 203–19, July.

Baird, S., Bohren, J. A., McIntosh, C. and Ozler, B. [2014]. "Designing Experiments to Measure Spillover Effects," *Penn Institute for Economic Research Working Paper*, 14-006, February.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O. [2012]. "The Diffusion of Microfinance," *NBER Working Paper*, 17743.

Barr, A., Dekker, M. and Fafchamps, M. [2012]. "Who Shares Risk with Whom under Different Enforcement Mechanisms?," *Economic Development and Cultural Change*, 60 (4): 677–706, July.

Belot, M. and Fafchamps, M. [2014]. "Experimental Evidence on Other-Regarding Preferences in Partnership Formation," *Mimeo*.

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

Bramoullé, Y., Djebbari, H. and Fortin, B. [2009]. "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 150 (1): 41–55.

Brock, W. A. and Durlauf, S. N. [2001]. "Interactions Based Models." In *Handbook of Econometrics*, Amsterdam: North Holland, 5, 3297–3380,

Brown, M., Falk, A. and Fehr, E. [2004]. "Relational Contracts and the Nature of Market Interactions," *Econometrica*, 72 (3): 747–80.

Brown, M., Falk, A. and Fehr, E. [2012]. "Competition and Relational Contracts: The Role of Unemployment as a Disciplinary Device," *Journal of the European Economic Association*, 10 (4): 887–907.

Caria, S. and Fafchamps, M. [2014a]. "Can Farmers Create Efficient Networks? Experimental Evidence from Rural India," *Mimeo*.

Caria, S. and Fafchamps, M. [2014b]. "Cooperation and Expectations in Networks: Evidence from a Network Public Good Experiment in Rural India," *Mimeo*.

Caeyers, B. and Fafchamps, M. [2015]. "Exclusion Bias in Social Interaction Models," *Mimeo*.

Chandrasekhar, A. and Lewis, R. [2012]. "Econometrics of Sampled Networks," *Mimeo*.

CHARNESS, G. and RABIN, M. [2002]. "Understanding Social Preferences with Simple Tests," *The Quarterly Journal of Economics*, 117 (3): 817–69, August.

COATE, S. and RAVALLION, M. [1993]. "Reciprocity without Commitment: Characterization and Performance of informal Insurance Arrangements," *Journal of Development Economics*, 40 (1): 1–24.

COMOLA, M. and FAFCHAMPS, M. [2014a]. "Bilateral Link Formation with Competition: An Experimental Study," *Mimeo*.

COMOLA, M. and FAFCHAMPS, M. [2014b]. "Testing Unilateral and Bilateral Link Formation," *The Economic Journal*, 124 (579): 954–976, September.

COMOLA, M. and PRINA, S. [2014]. "Do Interventions Change the Network? A Dynamic Peer Effect Model Accounting for Network Changes," *SSRN Working Paper*, 2250748, April.

DAVIES, E. and FAFCHAMPS, M. [2014]. "Competition, Enforcement and Reputation in Labor Markets: An Experimental Study in Ghana," *Mimeo*.

DE WEERDT, J. and FAFCHAMPS, M. [2011]. "Social Identity and The Formation of Health Insurance Networks," *Journal of Development Studies*, 47 (8): 1152–77.

DRUKKER, D. M., PRUCHA, I. R. and RACIBORSKI, R. [2001]. "A Command for Estimating Spatial-Autoregressive Models with Spatial-Autoregressive Disturbances and Additional Endogenous Variables," *The Stata Journal*, 1 (1): 1–13.

FAFCHAMPS, M., GOYAL, S. and VAN DER LEIJ, M. [2010]. "Matching and Network Effects," *Journal of the European Economic Association*, 8 (1): 203–31, January.

FAFCHAMPS, M. and GUBERT, F. [2007]. "The Formation of Risk-Sharing Networks," *Journal of Development Economics*, 83 (2): 326–50, July.

FAFCHAMPS, M. and HILL, R. [2014]. "Redistribution and Group Participation: Comparative Experimental Evidence from Africa and the UK," *Mimeo*.

FAFCHAMPS, M. and LA FERRARA, E. [2012]. "Self-Help Groups and Mutual Assistance: Evidence from Kenyan Slums," *Economic Development and Cultural Change*, 60 (4): 707–34, July.

FAFCHAMPS, M. and LABONNE, J. [2013]. "Do Politicians' Relatives Get Better Jobs? Evidence from Municipal Elections in the Philippines," *Mimeo*.

FAFCHAMPS, M. and QUINN, S. [2013]. "Networks and Manufacturing Firms in Africa: Results from a Randomised Field Experiment," *Mimeo*.

FAFCHAMPS, M. and QUINN, S. [2014]. "Social Networks and Business Practices: Evidence from a Randomized Experiment with Microentrepreneurs," *Mimeo*.

FAFCHAMPS, M. and QUISUMBING, A. [2003]. "Social Roles, Human Capital, and the Intrahousehold Division of Labor: Evidence from Pakistan," *Oxford Economic Papers*, 55 (1): 36–80.

FAFCHAMPS, M. and SÖDERBOM, M. [2014]. "Network Proximity and Business Practices in African Manufacturing," *World Bank Economic Review*, 28 (1): 99–129.

FAFCHAMPS, M., VAZ, A. and VICENTE, P. [2013]. "Voting and Peer Effects: Evidence from a Randomized Controlled Trial," *Mimeo*.

FAFCHAMPS, M. and VICENTE, P. [2013]. "Political Violence and Social Networks: Experimental Evidence from a Nigerian Election," *Journal of Development Economics*, 101 (1): 27–48, March.

FAFCHAMPS, M. and WOODRUFF, C. [2014]. "Identifying Gazelles: Expert Panels vs. Surveys as a Means to Identify Firms with Rapid Growth Potential," *Mimeo*, in preparation.

FALK, A. and HUFFMAN, D. [2007]. "Studying Labor Market Institutions in the Lab: Minimum Wages, Employment Protection, and Workfare," *Journal of Institutional and Theoretical Economics*, 163 (1): 30–45, March.

FALK, A., HUFFMAN, D.B. and MACLEOD, W.B. [2008]. "Institutions and Contract Enforcement," *ISA Discussion Papers*, 3435.

FEHR, E. and SCHMIDT, K.M. [1999]. "A Theory Of Fairness, Competition, And Cooperation," *The Quarterly Journal of Economics*, 114 (3): 817–68, August.

GINÉ, X. and MANSURI, G. [2011]. "Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan," *The World Bank Policy Research Working Paper*, June.

GRANOVETTER, M.S. [1995]. *Getting a Job: A Study of Contacts and Carreers*. Chicago (Ill.): University of Chicago Press, 2nd edition.

GURYAN, J., KROFT, K. and NOTOWIDIGDO, M. J. [2009]. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal: Applied Economics*, 1 (4): 34–68, October.

JACKSON, M. O. [2009]. *Social and Economic Networks*. Princeton (N. J.): Princeton University Press.

JACKSON, M. O., RODRIGUEZ-BARRAQUER, T. and TAN, X. [2012]. "Social Capital and Social Quilts: Network Patterns of Favor Exchange," *American Economic Review*, 102 (5): 1857–97, August.

KELEJIAN, H. and PRUCHA, I. R. [1998]. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17 (1): 99–121.

KOCHERLAKOTA, N. R. [1996]. "Implications of Efficient Risk Sharing without Commitment," *Review of Economic Studies*, 63 (4): 595–609, October.

LEE, L.-F. and LIU, X. [2010]. "Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances," *Econometric Theory*, 26 (1): 187–230, February.

LIU, X., PATACCHINI, E., ZENOU, Y. and LEE, L.-F. [2012]. "Criminal Networks: Who Is the Key Player?," *FEEM Working Paper*, 39.2012, June.

LYLE, D. [2007]. "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *The Review of Economics and Statistics*, 89 (2): 289–99.

LYLE, D. [2009]. "The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point," *American Economic Journal: Applied Economics*, 1 (4): 69–84.

MANSKI, C. [1993]. "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60 (3): 531–42.

MEVLUDE, A.-Y. and YUKSEL, M. [2011]. "The Long-Term Direct and External Effects of Jewish Expulsions in Nazi Germany," *IZA Discussion Papers*, 5850.

MOFFITT, R. [2001]. "Policy Interventions, Low-Level Equilibria, and Social Interactions." In DURLAUF, S. and YOUNG, P. (eds), *Social Dynamics*. Cambridge (Mass.): The MIT Press, p. 45–82.

PATNAM, M. [2011]. "Corporate Networks and Peer Effects in Firm Policies," *Mimeo*.

SACERDOTE, B. [2001]. "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 116 (2): 681–704.

SHUE, K. [2011]. "Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers," *Mimeo*.

VAN DEN BOOGAART, M., FAFCHAMPS, M. and SÖDERBOM, M. [2014]. "Social Networks and the Diffusion of Airtime Transfers in Rwanda," *Mimeo*.

APPENDIX

To illustrate how the network autoregressive parameter can be identified from data, imagine 5 nodes in a circle and suppose that each node is influenced by its left neighbor. $I - \rho G$ has the form:

$$I - \rho G = \begin{bmatrix} 1 & 0 & 0 & 0 & -\rho \\ -\rho & 1 & 0 & 0 & 0 \\ 0 & -\rho & 1 & 0 & 0 \\ 0 & 0 & -\rho & 1 & 0 \\ 0 & 0 & 0 & -\rho & 1 \end{bmatrix}$$

The corresponding matrix $(I - \rho G)^{-1}$ is:

$$(I - \rho G)^{-1} = \frac{1}{1-\rho^5} \begin{bmatrix} 1 & \rho^4 & \rho^3 & \rho^2 & \rho \\ \rho & 1 & \rho^4 & \rho^3 & \rho^2 \\ \rho^2 & \rho & 1 & \rho^4 & \rho^3 \\ \rho^3 & \rho^2 & \rho & 1 & \rho^4 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

The exponent of ρ is the distance from *i* to *j* in the directed network G. The covariance matrix of Y is:

$$E[YY'] = \frac{\sigma^2}{(1-\rho^5)^2} \begin{bmatrix} a & b & c & c & b \\ b & a & b & c & c \\ c & b & a & b & c \\ c & c & b & a & b \\ b & c & c & b & a \end{bmatrix}$$

$$a \equiv 1 + \rho^2 + \rho^4 + \rho^6 + \rho^8$$

$$b \equiv \rho + \rho^3 + \rho^4 + \rho^5 + \rho^7$$

$$c \equiv \rho^2 + \rho^3 + \rho^4 + \rho^5 + \rho^6$$

which demonstrates how ρ can be estimated from the sample moments of matrix $E[YY']$.