

IZA DP No. 8896

Some Challenges in the Empirics of the Effects of Networks

Vincent Boucher
Bernard Fortin

February 2015

Some Challenges in the Empirics of the Effects of Networks

Vincent Boucher

Université Laval, CIRPÉE

Bernard Fortin

*Université Laval, CIRPÉE, CIRANO
and IZA*

Discussion Paper No. 8896
February 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Some Challenges in the Empirics of the Effects of Networks^{*}

We study some recent developments and challenges in the empirics of the effects of social networks. We focus in particular on researchers' ability to make policy recommendations based on a standard linear econometric model. We examine the potential compatibility between this type of econometric model and a microeconomic theoretical approach based on fundamentals, such as preferences, technology and decision processes. We discuss sources of identification for the social multiplier as well as for the identity of the key player. We study the possibility of testing endogeneity in network formation. We analyse the use of proxy variables and their impact for the causal interpretation of the peer effect coefficients. Our analysis suggests that greater care should be taken in grounding econometric network models to sound and credible theoretical underpinnings.

JEL Classification: A14, C33, C36, D85, Z13

Keywords: social networks, social multiplier, network formation, identification, proxy variables, policy analysis

Corresponding author:

Bernard Fortin
Economics Department
Pavillon J.-A. De Sève
Laval University
Quebec (QC)
Canada
E-mail: Bernard.Fortin@ecn.ulaval.ca

^{*} Prepared for the *Oxford Handbook on the Economics of Networks*, Y. Bramoullé, B. Rogers and A. Galeotti (Eds.), Oxford University Press, forthcoming in 2015. We are grateful to Yann Bramoullé, Rokhaya Dieye, Andrea Galeotti, Brian Rogers and Myra Yazbeck for helpful comments. Support for this work has been provided by the Canada Chair of Research in Economics of Social Policies and Human Resources, and le Fonds de Recherche sur la Société et la Culture.

1. Introduction

Aristotle, the Greek philosopher, writes "Man is by nature a social animal... He who lives without society is either a beast or God" (*Politics*, Book I, Part II). In economics, the importance of social interactions outside the market is now well recognized. Individuals share information, learn from others and influence each other in many contexts. This point is particularly important for policy evaluation. To estimate the overall effect of a social intervention, it is crucial to account not only for its direct impact on the treated but also its indirect impact on their peers (Manski, 2013). Recently, a growing literature has attempted to investigate social interactions at the theoretical, empirical and econometric levels using the *social network* approach.

At the theoretical level, social network theory has developed as a rigorous language to analyse social interactions (Jackson, 2011). This theoretical framework explores in particular 1) how social networks influence outcomes and 2) how this in turn affects network formation (Jackson, 2010).

At the empirical level, while the literature on the formation of social networks is limited but burgeoning (Fafchamps and Gubert, 2007; Comola, 2008; Mayer and Puller, 2008; Christakis et al., 2010; Mele, 2013; Boucher, 2014b),¹ applications on the effects of social networks on outcomes is vast and expanding at a rapid pace. To name a few, we can mention Bertrand et al. (2000) on welfare participation, Cassar (2007) on coordination and cooperation, Patacchini and Zenou (2008) on criminal activity, Trogdon et al. (2008) on obesity, Karlan et al. (2009) on risk-sharing, and Calvó-Armengol et al. (2009) on education. Moreover, some recent papers attempt to simultaneously evaluate network formation and network effects (Conti et al., 2012; Hsieh and Lee, 2011; Goldsmith-Pinkham and Imbens, 2013; Badev, 2013; Boucher, 2014a).

At the econometric level, a rich literature has emerged focusing on issues raised by iden-

¹ See also the survey of Chandrasekhar in this Handbook.

tification of peer effects in econometric models of social networks (Manski, 1993; Bramoullé et al., 2009; Lee et al., 2010; Blume et al., 2013). Also, issues of estimation have been addressed either using experimental data and most often based on a reduced form approach (Sacerdote, 2001; Barrera-Osorio et al., 2011) or using non-experimental data (Lin, 2010; Card and Giuliano, 2013). In the latter case, a number of studies have been inspired by the spatial econometric literature (Bramoullé et al., 2009; Lee et al., 2010; Lin, 2010). Distribution-free methods such as generalized two-stage-least squares and GMM, or approaches involving more structure such as (quasi-)maximum likelihood with or without autoregressive disturbances have been proposed.

As pointed out by Blume et al. (2013), the theoretical, empirical and econometric literature on social networks lack integration. Regarding the integration between theoretical and empirical social network approaches, the so-called *identifiability problem* (Chiappori and Ekeland, 2009) is rarely addressed. For instance, assume a non-cooperative (Nash) game over networks in which each individual maximizes his utility, given his characteristics and the behaviours and characteristics of his reference group. In this case, the identifiability problem is whether the individual best response (or reaction) functions can be uniquely rationalized by a microeconomic theoretical model based on fundamentals, such as the utility parameters characterizing private and social preferences. As we will show, a positive answer to this question may help the researcher to simulate overall (direct plus indirect) impact of hypothetical alternative policy reforms, such as a higher taxes on tobacco or an increase in penalties on crime, as long as he has full information on the best response functions. This issue is related to the measurement of the so-called *social multiplier*.

Regarding the integration of empirical and econometric approaches, a crucial issue deals with the possibility of recovering the (stochastic) best response functions from some *a priori* knowledge of the data-generating process, that is, the *identification problem*. Assume a standard linear-in-means model where the outcome of each individual depends linearly on his own characteristics, on the mean outcome of his reference group (*endogenous peer effect*)

and on its mean characteristics (*contextual peer effects*). Assume also that the outcome is continuous.² Identifying the best response functions involves specific econometric problems, such as 1) the *reflection problem* (as one’s own behaviour depends on peer behaviour, which itself simultaneously depends on one’s own behaviour), 2) the presence of *correlated effects* (e.g., due to endogeneity of network formation or network-invariant unobservables), and 3) *incomplete information* (e.g., due to partial knowledge of networks’ architecture and individual outcomes). When individuals interact in groups³ and the group sizes are the same, Manski (1993) shows that the reflection problem makes the best response model unidentified.⁴ However, Bramoullé et al. (2009) shows that there is no reflection problem when individuals do not interact through groups, which is most likely to be the case. Therefore, two basic issues remain as related to the identification problem: the endogeneity of network formation⁵ and problems of information. Problems of information happen either because the game played by the individuals is a game of incomplete information (Blume et al., 2013; Yang and Lee, 2014) or because the econometrician has only partial knowledge of the network (Chandrasekhar and Jackson, 2013) or of the characteristics and behaviours of the individuals (Liu et al., 2013a; Boucher et al., 2014). One important issue related with the latter point is whether the use of proxy variables (e.g., Cooley Fruehwirth, 2013), such as hours of work or body mass index as proxies for work effort or healthy life habits, may allow us to identify peer effects in the structural econometric model.

In this chapter, we focus on some issues associated with the identifiability and the identification problems in the social network approach. In particular, we put the accent on *four* aspects that we think are crucial for *policy analysis* and on which the chapter makes

² A discrete outcome generally helps to identify the model but may be the source of multiple equilibria (Brock and Durlauf, 2007).

³ This means that the population is partitioned into groups, and that individuals are affected by all others in their group and by none outside of it.

⁴ Manski implicitly assumes that the size of groups is infinite, which is akin to assuming that they are the same.

⁵ The presence of network-invariant unobservables can be treated by introducing network fixed effects into the model (Bramoullé et al., 2009; Lee et al., 2010).

a number of contributions.

To give a broad perspective on the relevant literature, we first provide a discussion of recent empirical contributions of research on the effects of social networks on outcomes with a focus on their implications for policymaking. We insist, in particular, on the importance of the *social multiplier* and *key player* concepts to understand the role of the network structure in evaluating the impact of policy interventions on aggregate outcomes.

Second, we show that the presence or absence of a social multiplier crucially depends on the source of social interactions (complementarity *vs.* pure conformity). We prove that the social multiplier is generally not identified even when the peer effects can be recovered (identifiability problem). However, we also show that information on isolated individuals within the network generally allows us to identify this effect.

Third, we extend our analysis to the case where the structure of a network, as summarized by the *social interaction matrix*,⁶ is not fixed but stochastic. In this context, two scenarios can be obtained depending on whether this matrix is exogenous, that is, not correlated with the random term, or endogenous. When the social interaction matrix is exogenous, it may depend on some individual variables. In that case, we show that it is important to account for the effect of a policy shock on the matrix when evaluating the impact of a policy. Also, we discuss the ability to test for the endogeneity of the network structure (identification problem).

Finally, we focus on issues of identification of the best response functions when the econometrician does not observe the true interaction variable but only a proxy (identification problem). We insist on the importance of developing a microeconomic model based on sound and credible theoretical underpinnings to determine the parameters of the best response functions that are identified.

This chapter mostly focuses on the standard linear-in-means model. Most papers on social interactions have considered the econometric framework since it is naturally related

⁶ Its mathematical definition is given in section 3.1.

to the standard simultaneous linear model. Nonetheless, the essence of most of our results and discussions extends to other linear models such as in Blume et al. (2013), Boucher (2014a) and Liu et al. (2014), as well as to non-linear models of social interactions.

The remainder of the chapter is organized as follows. Section 2 provides a survey of the recent contributions on the effects of networks on outcomes with an emphasis on their policy implications. Section 3 discusses the concept of social multipliers and presents our basic results as related to its identifiability. Section 4 discusses various types of network structures. We introduce the notion of stochastic network structure and discuss the cases where the network is exogenous or endogenous. Section 5 analyses identification issues when the right interaction variable is unobservable and a proxy variable is used. Section 6 concludes.

2. What Has Been Learned About the Effects of Social Networks on Policy Outcomes ?

One important interest of using the social network framework is for policy analysis purposes. This section considers a number of concepts with applications illustrating various types of peer effects and other externalities that can be fruitfully investigated in a network setting.

One of the central concepts in social networks related to policy analysis involves *social multipliers* (Glaeser et al., 2003). It arises when a change in a *common* (policy) shock exerts both a *direct effect* on individual action, and an *indirect effect* through social influence. The social multiplier can be defined as the ratio of the total (direct plus indirect) effect to the direct effect, and therefore exceeds 1 (around a stable equilibrium) under the assumption of positive spillovers or complementarity within the network.⁷ The relevance of social multipliers in policy analysis stems from the fact that the impact of a (small) intervention

⁷ This result is formally proven in section 3.1.

may be strongly amplified at the aggregate level when social interactions across individuals are large.

Using social multipliers, policymakers can obtain information on the impacts of shocks on various networks and involving different types of externalities, such as academic performance at school (Graham, 2008), labour participation of mothers (Maurin and Moschion, 2009), tax evasion among small businesses and professionals (Galbiati and Zanella, 2012), and fast food consumption among teenagers in schools (Fortin and Yazbeck, 2014). Importantly, in many papers, the social multiplier in a linear-in-means model is computed as $1/(1 - \delta)$,⁸ where δ is the endogenous peer effect, that is, the impact of the individual reference group's mean outcome on his/her own outcome. However, we show in section 3.1 that this equality does not always hold and that the social multiplier depends on the mechanisms of social interactions (*e.g.*, complementarity vs. conformity). We also show that those mechanisms may be hard to identify in the linear-in-means model even if peers effects are identified. However, identification may be possible when there are isolated individuals in the network (see section 3.2).

While social multipliers apply to a common policy shock, it may be the case that the policy intervention directly affects only a subset of the social network. This point was discussed, among others, in Moffitt (2001) and empirically analysed in Babcock and Hartman (2010) and Dieye et al. (2014), using a treatment evaluation approach. For instance, let us suppose that a fraction of students are given free textbooks to learn the material presented in class. This intervention will not only influence their academic performance but may also indirectly affect the performance of those who did not receive a textbook, through social interactions. This suggests that when one's treatment status influences the actions of others, experimental data are not generally sufficient to identify a causal treatment impact. A social network setting can be useful to provide an adequate analysis of this type of

⁸ There is a natural analogy between this formula and the standard Keynesian multiplier.

intervention (Dieye et al., 2014).⁹

Related to the previous point, one may ask which individuals in the class should be given a textbook to maximize the overall impact of the intervention on academic performance (for a given budget). This normative issue has been called the key-player problem in the social network literature (Borgatti, 2006). At the econometric level, it has been addressed in a number of papers (e.g., Ballester et al., 2006; Liu et al., 2014) that have been surveyed by Zenou in this Handbook. In section 3.3 we provide a simple approach that shows how to determine the key player in a linear-in-means model.

Up to now, it has been assumed that the shock does not influence the structure of the network. In contrast, Comola and Prina (2013) propose an approach in which the social network can be influenced by the exogenous shock. They apply it to experimental data collected to evaluate the impact of a new savings technology in Nepal. Their analysis accounts for the change in the network of financial exchanges between all villagers before and after the shock. According to their results, a non-negligible share of the peer effect in savings comes from the change in the structure of the network. In section 4.1, we provide an analysis of peer effects when the network architecture is exogenous, but stochastic, i.e., it may be affected by changes in exogenous variables.

One of the most difficult problems arising when using network data for policy analyses involves the structure of the network. It is likely to be stochastic endogenous, at least when no experimental data are available.¹⁰ For instance, individual preferences within a friendship network are likely to be similar. Thus, one should expect to observe the presence of homophily on *unobserved* individual characteristics, that is, the tendency of

⁹ Moffitt (2001) points out that this intervention also helps with the identification in a linear-in-means model when individuals interact in groups and the size of groups is very large. The reason is that the policy introduces an exclusion restriction in the model since the students who do not receive a textbook are not directly affected by the intervention.

¹⁰ Using experimental data also raises its own problems such as external validity and the use of an incomplete network. See Blume et al. (2011) for a critical analysis of identification problems when using experimental data.

friends to link with similar other individuals. This source of spurious correlation between friends' outcomes may generate a serious identification problem and produce potentially large biases in the estimators of the structural econometric model.

Recently, some researchers (Conti et al., 2012; Hsieh and Lee, 2011; Goldsmith-Pinkham and Imbens, 2013; Badev, 2013; Boucher, 2014a) have made attempts to develop econometric models allowing for the joint estimation of network formation and network interactions. However, empirical results using Add Health data and focusing on outcomes such as smoking, sleep behaviour and academic performance among students at the secondary level in the U.S. do not seem to detect much difference in peer effects when networks are assumed exogenous and when they are allowed to be endogenous. Among the many reasons which could explain this is the fact that the Add Health database includes a very large number of observable characteristics that can be introduced in the econometric model (Liu et al., 2013a). Another explanation is that, although statistically significant, the explanatory power of the individual characteristics on the probability that two individuals are friends is extremely small (Boucher, 2014a). However, as noted by Badev (2013), even with this absence of bias on the estimated parameters, the endogeneous nature of the network structure may still strongly affect the impact of policy shocks.¹¹ In section 4.1, we show that this argument holds even when the network is exogeneous. In section 4.3, we provide a discussion on how to test for network endogeneity in a particular class of models (latent space models) where the probability of a relation between individuals depends on their positions in a partially observed “social space” (Hoff et al., 2002).

As stressed by Manski (1993), one key issue faced by a researcher who studies social networks is: Who interacts with whom? Except for cases when data from highly controlled field or lab experiments are available (*e.g.*, Mas and Moretti (2009) and Beugnot et al. (2014), in experiments on work effort), problems such as incomplete information on

¹¹ Badev (2013) finds that neglecting the endogeneity of the friendship network leads to a downward bias of 10% to 15% on the predicted impact of public policies.

the structure of networks or partial observability of outcomes are ubiquitous. Davezies et al. (2009) has provided a simple solution to the latter issue when individuals interact through a group. Boucher et al. (2014) has implemented their approach to estimate peer effects in academic achievements in Canadian secondary schools. The former issue is much more severe when individuals interact through more complex social networks and networks architecture is only partially known. Chandrasekhar and Lewis (2011) have developed a generalized instrumental approach that yields consistent estimators of the best response functions when the researcher observes only a sample of the network. Liu et al. (2013a) study the special case where the network is fully observed, but the outcome variable is only observed for a subset of the individuals.

A related important observability issue arises when some crucial variables of the model of social interactions are not available. For instance, in the literature on the social epidemic of obesity (Christakis and Fowler, 2007; Trogon et al., 2008; Cohen-Cole and Fletcher, 2008), the variable of an individual's effort to maintain a normal weight is usually not available. Therefore, researchers often regress the individual's BMI on his peers' average BMI and other variables to estimate peer effects. Does this approach allow to identify true peer effects? In section 5, we show that the answer to this question crucially depends on the nature of the microeconomic model incorporating both the technical relationship between the individual's weight and his effort, and the individual's utility function.

Up to now, the discussion has focused on models with linear social interactions. However, as stressed by a number of researchers (Sacerdote, 2011), the standard linear model may hide large heterogeneity in the effects experienced by different types of individuals. For instance, in the education sector, Sacerdote (2011) reports that when one uses nonlinear models, one prevalent finding is larger peer effects in which high ability students benefit from the presence of other high ability students. Also, when the outcome of interest is discrete (*e.g.*, smoking, marijuana use, criminality among teenagers), the network model becomes intrinsically nonlinear. This raises the problem of multiple equilibria (Brock and

Durlauf, 2001; Card and Giuliano, 2013; Lee et al., 2013; Badev, 2013).

3. Social Multiplier in Networks

As discussed in the previous section, one major policy implication of the peer effect literature is the potential presence of a social multiplier. In this section, we discuss the researcher's ability to identify the presence or not of a social multiplier from the structural econometric model.

3.1. A negative identification result

This sub-section provides a negative result regarding the identification of the primitive preferences parameters (the fundamentals of the microeconomic model). We assume that the individual strategy within a network obeys a linear structure as it is commonly assumed in the literature and that no individual is isolated.¹² We show that it is not possible to recover the social multiplier even when peer effects are identified.¹³

To see why, let us focus on a simple static model inspired by Bramoullé et al. (2009) and Blume et al. (2013). Assume workers with size n are interacting through a network within a firm and are producing an output. The $n \times n$ network adjacency matrix \mathbf{A} is such that $a_{ij} = 1$ if the worker i is influenced by the worker j and $= 0$, otherwise. We assume the links do not differ in strength. Worker i 's reference group with size n_i is the set of other workers by which he is influenced. For the remainder of this subsection, we assume that the social interaction matrix \mathbf{G} represents a row-normalization of the adjacency matrix \mathbf{A} such that rows of \mathbf{G} sum to 1. This implies that there are *no isolated workers*. Therefore $g_{ij} = 1/n_i$ if i is influenced by j , and 0 otherwise. Suppose that \mathbf{G} is a *non-stochastic* (or fixed) and known social interaction matrix.

¹² This assumption is relaxed in the next section.

¹³ Blume et al. (2013) provides a similar negative result on the identification of preferences but does not discuss its link with the social multiplier.

Consider the following quadratic utility functions:¹⁴

$$u_i(\mathbf{y}, c_0) = (c_0 + \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i)y_i - \frac{y_i^2}{2} + \alpha y_i \mathbf{g}_i \mathbf{y} \quad (1)$$

$$v_i(\mathbf{y}, c_0) = (c_0 + \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i)y_i - \frac{y_i^2}{2} - \frac{\lambda}{2} (y_i - \mathbf{g}_i \mathbf{y})^2 \quad (2)$$

where \mathbf{x}_i is the vector of individual i 's observable characteristics and ε_i is a random term representing his unobservable characteristics. One assumes that \mathbf{x}_i are strictly exogenous, and that $\alpha \in [0, 1)$ and $\lambda \geq 0$.¹⁵ Both utility functions are separable into two components: a private and a social sub-utility. The first two expressions on the right-hand side of these equations are the same and correspond to the private sub-utility. The third expression, which describes the social sub-utility, features social interactions and differs across equations.

Preferences given by these utility functions come from two alternative sources of social interactions: *complementary* and *pure conformity*. A worker whose preferences are given by u_i could be selecting his production, y_i , taking into account the benefit received per unit produced, $c_0 + \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$, the cost of producing, $y_i^2/2$, and the synergy created by working with his co-workers, $\alpha y_i \mathbf{g}_i \mathbf{y}$.¹⁶ The latter expression, which reflects the social sub-utility of u_i , means that an increase in the average co-workers' production, $\mathbf{g}_i \mathbf{y}$, positively influences his marginal productivity (*complementary*).¹⁷ Alternatively, a worker whose preferences are given by v_i could be also choosing his production taking into account the benefit received from unit produced, $c_0 + \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$, and the cost of producing, $y_i^2/2$. However, in this case

¹⁴ To simplify the discussion, we ignore contextual peer effects. Including the latter would not change the basic argument.

¹⁵ Those conditions on α and λ imply that the set of best response functions is a contraction, which implies the existence of a unique Nash equilibrium of the game.

¹⁶ Calvó-Armengol et al. (2009) took this approach in modeling peer effects in academic achievement.

¹⁷ Note that *substitutability* in social interactions (with $\alpha < 0$), where the average peers' production reduces a worker's utility, is also possible as long as $|\alpha| < 1$.

there is no complementarity. Rather, his social sub-utility is positively affected by the degree to which he conforms with his peers' level of production, $-\frac{\lambda}{2}(y_i - \mathbf{g}_i \mathbf{y})^2$. This refers to *pure conformity* due, for instance, to social norms or emulation between workers.¹⁸

Note that even if the utilities u_i and v_i represent different sources of social interactions, they lead to *observationally equivalent* sets of reduced form models:¹⁹

$$\mathbf{y}^u(c_0) = (\mathbf{I} - \alpha \mathbf{G})^{-1} (c_0 \mathbf{1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}) \quad (3)$$

$$\mathbf{y}^v(c_0) = (\mathbf{I} - \tilde{\alpha} \mathbf{G})^{-1} (\tilde{c}_0 \mathbf{1} + \mathbf{X} \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}) \quad (4)$$

where $\tilde{\alpha} = \lambda/(1 + \lambda)$, $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}/(1 + \lambda)$, $\tilde{c}_0 = c_0/(1 + \lambda)$ and $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}/(1 + \lambda)$ and where $\mathbf{1}$ is the vector of 1s of dimension n . This is important as only the model with complementarity features a social multiplier!

In other words, the presence (or not) of a social multiplier cannot be identified from the agents' decisions. Formally:

Proposition 1. *Let $c_1 = c_0 + c$, then $\Delta \mathbf{y}^u = \frac{c}{1-\alpha} \mathbf{1}$ and $\Delta \mathbf{y}^v = c \mathbf{1}$,*

with $\Delta \mathbf{y}^u = \mathbf{y}^u(c_1) - \mathbf{y}^u(c_0)$ and $\Delta \mathbf{y}^v = \mathbf{y}^v(c_1) - \mathbf{y}^v(c_0)$.

Suppose that the change (Δ) from c_0 to c_1 is a policy shock represented by the introduction of a wage subsidy. Consider first a worker whose preferences are represented by u_i . Without social interactions ($\alpha = 0$), this leads to a change in the production of any worker by c . With social interactions, however, a worker also indirectly benefits from the other workers' change in pay so the overall impact of the change is larger: $|c|/(1 - \alpha) > |c|$. In this case, the social multiplier is $1/(1 - \alpha)$, which is strictly greater than 1.

¹⁸ As is well-known, the literature on game theory associates the expression *strategic complementarity* or *supermodularity* to the case where an increase in a player's outcome increases the marginal utility (or payoff) of the other player's outcome. Here it is easy to show that both complementarity (as we define it) and pure conformity generate strategic complementarity.

¹⁹ These *restricted* reduced forms of equations correspond to the (unique) Nash equilibrium implied by the set of best-response functions for the two models.

Now consider a worker whose preferences are represented by v_i . In that case, there is no social multiplier (it is equal to 1), that is, the impact of the wage subsidy on his production is not amplified by his peers' mean production. The basic intuition is that the subsidy induces the same direct effect (*i.e.*, without social interactions) on a worker's production and on that of his peers. Therefore, the *gap* between these variables, or $(y_i - \mathbf{g}_i \mathbf{y})^2$, is not influenced by the policy shock. As a result, in this pure conformity model, social interactions between individuals will generate no indirect effect.

Note that it is possible that a worker's social utility has both complementary and pure conformity features. For instance, a worker may be affected both by the synergy in productivity from his co-workers and by his incentives to imitate their behaviour.²⁰ In that case, the corresponding social multiplier will be equal to $\frac{1}{1-\alpha}$ and is independent of the parameter λ reflecting pure conformity.

In this situation, the best-response functions, obtained from maximizing each worker's utility with respect to his own production, *given* the others' production, are given by:

$$\mathbf{y} = \tilde{c}_0 \mathbf{1} + \mathbf{X} \tilde{\boldsymbol{\beta}} + \delta \mathbf{G} \mathbf{y} + \tilde{\boldsymbol{\varepsilon}} \quad (5)$$

where $\delta = \frac{\alpha + \lambda}{1 + \lambda}$. In (5), the parameters $\tilde{\boldsymbol{\beta}}$ capture the individual effects and δ captures the endogenous peer effect. When individuals do not interact through groups, the parameters of the linear-in-means model (5) are identified.²¹ However, the parameters α and λ defining δ are not identified. Again, this shows that the social multiplier as well as the sources of social interactions are not identified.

Our analysis also shows the importance of providing the micro foundations (*e.g.*, in terms of preferences, technology and decision process) of a social interactions model to fully address the identifiability problem, as was carefully done for instance by Blume et al.

²⁰ In that case, the worker's utility function is given by: $(c_0 + \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i) y_i - \frac{y_i^2}{2} + \alpha y_i \mathbf{g}_i \mathbf{y} - \frac{\lambda}{2} (y_i - \mathbf{g}_i \mathbf{y})^2$.

²¹ The parameters are also identified when individuals interact through groups and there are at least three groups with different sizes (Davezies et al., 2009).

(2013). Thus, focusing the analysis only on the linear-in-means structural model does not allow to derive any identifiability result on the social multiplier.²²

3.2. Room for identification when some individuals are isolated

The argument developed up to now depends on one important assumption: no individual is isolated. This implicitly follows from the assumption that the social matrix interaction \mathbf{G} is row-normalized. This assumption is usually made to obtain identification results in the presence of correlated effects due to fixed network unobservables (Bramoullé et al., 2009; Lee et al., 2010; Kwok, 2013). It is usually assumed that removing those individuals from the database can be done without any great loss of generality because, by definition, they do not affect the correlation structure of the network game. However, we will see that those isolated individuals provide identifiability conditions for the presence of a social multiplier as well as the source of social interactions.

Suppose now that the firm is also composed of isolated workers (referred to by $\tilde{\cdot}$). Since their preferences do not include any social sub-utility component, their corresponding utilities are given by:

$$\check{u}_i(y_i, c_0) = (c_0 + \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i)y_i - \frac{y_i^2}{2} \quad (6)$$

$$\check{v}_i(y_i, c_0) = (c_0 + \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i)y_i - \frac{y_i^2}{2} \quad (7)$$

For those workers, both complementary and pure conformity models lead to the *same* best-response functions:

$$\tilde{\mathbf{y}} = c_0\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (8)$$

²² This point was also made in Liu et al. (2013b) where they present a model featuring both complementarity and pure conformism. They provide a test for distinguishing between what they call the local aggregate (complementarity) and the local average (pure conformism). Their test relies on the fact that their local aggregate effect is defined on the non-normalised adjacency matrix. Our point is stronger: restricting the analysis to row-normalized matrices, the econometric model does not allow to identify the source of the interaction (pure conformity or complementarity).

where β and c_0 are identified when \mathbf{X} is full rank and can be estimated by a simple OLS.

We argue that isolated workers allow to test for the presence of a social multiplier. To see why, consider (5), where both sources of social interactions are potentially present, as well as (8) for isolated individuals. Since β is identified from (8) and $\tilde{\beta}$ is identified from (5), λ is identified.²³ Also, since δ is identified from (5), the identification of λ is sufficient for the identification of α .²⁴ Then, all the parameters are identified, which is sufficient to test for the presence of a social multiplier.²⁵

This argument, however, relies on the researcher's ability to distinguish between isolated and non-isolated individuals, which is strongly dependent on the observation of the exact network structure. Specifically, if some links are unobserved, one may incorrectly infer that an individual is isolated. In that case, the estimated parameters are likely to be biased and researchers should proceed to robustness checks, for instance using simulations.

Here an important remark is in order. In some cases, as for example when the network structure is endogenous (see section 4.2), one may want to estimate the model with the full network matrix. Let \mathbf{G} be such a matrix, where rows of isolated individuals sum to 0, while the other rows sum to 1.²⁶ Consider the following model:

$$\mathbf{y} = c_0 \mathbf{1} + \mathbf{X}\beta + \alpha \mathbf{G}\mathbf{y} + \varepsilon$$

With this specification, and since \mathbf{G} includes isolated and non-isolated individuals, the model *imposes* that the individual effects are the same for every individual. This implies that the economy *has* to feature a social multiplier. The model is then incompatible with preferences such as v_i and the social interaction parameter should not (unless explicit preferences are specified) be interpreted as an effect of conformism.²⁷

²³ Recall that $\tilde{\beta} = \beta/(1 + \lambda)$.

²⁴ Recall that $\delta = (\alpha + \lambda)/(1 + \lambda)$.

²⁵ Recall that the social multiplier is equal to $\frac{1}{1-\alpha}$.

²⁶ This definition of the interaction structure is used, among others, by Bramoullé et al. (2009).

²⁷ When the interaction matrix \mathbf{G} is not row-normalized, however, the full model may not be incompatible

3.3. Social multiplier or key player?

In the previous sections, we have defined the social multiplier as the impact of a *common* shock (*i.e.*, the wage subsidy). However, even when social interactions do not lead to the presence of a social multiplier, the policymaker may want to know which individuals to target (e.g., Ballester et al., 2006; Graham, 2014). Suppose that the policymaker wants to affect only a subset $I \subset N$ of individuals. Which ones should he target in order to maximize the overall impact? Which criminals, once put in jail, generate the highest possible reduction in aggregate crime levels (Liu et al., 2014)? In this section, we briefly describe the identity of those key-players.²⁸

Consider the following model where \mathbf{G} is row-normalized:

$$\mathbf{y} = \tau (\mathbf{I} - \delta \mathbf{G})^{-1} (c_0 \mathbf{1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

where $\delta \in (0, 1)$ and $\tau = \frac{1}{1+\lambda}$, with $\lambda \geq 0$. Remark that this specification includes the models of complementarity ($\lambda = 0$ and $\delta = \alpha$) and pure conformity ($\lambda > 0$ and $\delta = \frac{\lambda}{1+\lambda}$) as special cases. The overall impact of a shock c on all individuals $i \in I$ is given by:

$$\tau c \mathbf{1}' (\mathbf{I} - \delta \mathbf{G})^{-1} \mathbf{e}_I$$

where \mathbf{e}_I is a vector taking 1 at the i 'th position for all $i \in I$ and 0 elsewhere. Letting $\mathbf{b} = (\mathbf{I} - \delta \mathbf{G}')^{-1} \mathbf{1}$, this is equivalent to $\tau c \mathbf{b}' \mathbf{e}_I = \tau c \sum_{i \in I} b_i$. The vector \mathbf{b} is a well-known measure of centrality called the PageRank (Brin and Page, 1998) and is equal to the Katz-Bonacich centrality for the transpose of \mathbf{G} (Calvó-Armengol et al., 2009). Then, the policymaker should target the most central individuals. Note that the identity of those individuals strongly depending on δ (\mathbf{b} is a function of δ), but *not* on τ and c , which only

with conformism. See, for example, Boucher (2014a) where the interaction matrix is given by the Laplacian matrix of the network.

²⁸ For a complete survey, see the chapter of this Handbook by Zenou.

affect the magnitude of the shock. To know the identity of the key players, the source of social interaction (complementarity or pure conformity) is then irrelevant, as long as δ is precisely estimated.

4. Types of Network Structures

Networks are convenient to represent the structure of interactions between individuals. Interactions of different natures lead to different types of network structures. Despite the important differences between those types (see below), the literature rarely distinguishes between them. In this section, we distinguish between three types of network structures: non-stochastic network structures, stochastic exogenous network structures and stochastic endogenous network structures. We also discuss the problem of identification of the structural econometric model in these three cases.

For the remaining of this section, we assume that the interaction matrix \mathbf{G} represents a row-normalization of the adjacency matrix \mathbf{A} such that rows for non-isolated individuals sum to 1 while the other rows are 0. To clarify the exposition, we assume that there are no endogenous effects and that the structural econometric model is given by:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{G}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{9}$$

where $\varepsilon_i \sim_{iid} N(0, \sigma)$.

4.1. Exogenous or non-stochastic?

Let's assume that the network structure \mathbf{G} is non-stochastic. Prominent economic examples of such network structures include group interactions (Lee, 2007; Boucher et al., 2014), provided that groups are non-stochastic, as well as panel data models. As discussed in Bramoullé et al. (2009), panel data models represent a special case of network interactions when individual i at time t has only one peer: himself at time $t - 1$. Econometric methods

based on non-stochastic network structures include for instance Lee et al. (2010).²⁹ When \mathbf{G} is non-stochastic, parameters in (9) are identified as long as matrix \mathbf{X} is of full rank and $\mathbb{E}(\boldsymbol{\varepsilon}|\mathbf{X}) = \mathbf{0}$. Moreover, they can be estimated by OLS.

Now, assume that \mathbf{G} is stochastic, that is, the observed network \mathbf{G} is the realisation of a random variable. Assume also that \mathbf{G} is exogenous in the sense that $\mathbb{E}(\boldsymbol{\varepsilon}|\mathbf{X}, \mathbf{G}) = \mathbf{0}$. For example, \mathbf{G} could represent the (row-)normalization of the adjacency matrix \mathbf{A} , defined as follows:

$$a_{ij} = \mathbf{I}(r(\mathbf{x}_i, \mathbf{x}_j) + \nu_{ij} > 0)$$

for some function r and where $\mathbf{I}(B)$ is an indicator function where $\mathbf{I}(B) = 1$ if B is true and $\mathbf{I}(B) = 0$ if B is false, and where $\nu_{ij} \sim_{iid} N(0, 1)$. In that example, we immediately see that $\mathbb{E}(\boldsymbol{\varepsilon}|\mathbf{X}, \mathbf{G}) = \mathbf{0}$, so the parameters in (9) can again be estimated by OLS.

Assuming that the network is non-stochastic or (stochastic and) exogenous therefore has the same implication for the identification and the estimation of the model. However, the impact for the interpretation, and related policy recommendations, of the model are *not* the same. For example, suppose that (9) represents workers' productivity, \mathbf{y} , as a function of the workers' network \mathbf{G} within a firm, and their individual characteristics \mathbf{X} , which may include, for instance, wages. Now, suppose that the government introduces a wage subsidy (*i.e.*, a shock on \mathbf{X}). That policy may also influence the structure of the workers' network within the firm. This must be taken into account in the evaluation of the impact of the policy on the workers' productivity. In general, for exogenous networks, one cannot predict the impact of a policy shock without imposing assumptions on the process generating the network.

This might seem like a simple remark. However, very few models of peer effects include a description of network formation process. When they do, the emphasis is placed on whether or not the network is endogenous (*i.e.*, whether or not $\mathbb{E}(\boldsymbol{\varepsilon}|\mathbf{X}, \mathbf{G}) = \mathbf{0}$ holds). Our

²⁹ Note that despite the fact that Lee et al. (2010) considers non-stochastic network structures, the method has been applied in practice to cases where the network is explicitly stochastic.

point is different: for policy’s sake, any model of social interactions based on a stochastic network structure should include a description of the network formation process. This does not mean that researchers should always provide an explicit network formation process, but they should at least discuss the expected impact of their proposed policy shocks on the structure of the interaction network.

The distinction between stochastic and non-stochastic network structures is crucial in order to describe the impact of a policy shock. In the next section, we discuss the third type of network structure, which is when the network is stochastic and endogenous (*i.e.*, when $\mathbb{E}(\boldsymbol{\varepsilon}|\mathbf{X}, \mathbf{G}) \neq \mathbf{0}$). Specifically, we discuss our ability to test for the endogeneity of the network structure.

4.2. Testing for network endogeneity in latent space models

Assume that the model in (9) can be decomposed as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{G}\mathbf{X}\boldsymbol{\gamma} + \underbrace{\boldsymbol{\rho}\boldsymbol{\nu} + \boldsymbol{\xi}}_{\boldsymbol{\varepsilon}} \quad (10)$$

where $\boldsymbol{\xi}$ is an exogenous error term. Assume also that the network structure is stochastic and that \mathbf{G} is the (row-)normalization of the adjacency matrix \mathbf{A} defined as follows:

$$a_{ij} = \mathbf{1}\{\kappa - \phi|x_i - x_j| - \mu|\nu_i - \nu_j| + \eta_{ij} > 0\} \quad (11)$$

where η_{ij} is an exogenous error term. The network formation processes characterized by (11), where the likelihood of a link between i and j can be represented as a function of the agents’ individual characteristics, are called *latent space models* in the statistical literature (Hoff et al., 2002) and have been recently used in economics by Goldsmith-Pinkham and Imbens (2013) and Graham (2014). When ϕ and μ are greater than zero, this indicates the presence of homophily, that is, the closer two individuals are in terms of their observable and unobservable characteristics, the higher is the probability that they are linked.

Here, the network structure is endogenous due to the fact that the econometrician does not observe the relevant variable ν , which enters both the outcome equation (10) and the network formation equation (11). Note that this is different from the endogeneity that would result from self-selection on the outcome variable y , as in Badev (2013) and Boucher (2014a). Unless $\rho = 0$ or $\mu = 0$, the outcome structural equation is unidentified due to the endogeneity created by ν . This makes it impossible to obtain consistent estimates of its parameters. Goldsmith-Pinkham and Imbens (2013) discusses the possibility of testing for this particular problem. The idea is the following.

Suppose that $\rho \neq 0$ and $\mu > 0$. Let \hat{a}_{ij} be the predicted value of a_{ij} assuming that $\mu = 0$ (that is, for a consistent estimator of κ and ϕ , conditional on $\mu = 0$). Similarly, let $\hat{\varepsilon}_{ij} = |\hat{\varepsilon}_i - \hat{\varepsilon}_j|$ where $\hat{\varepsilon}_i$ is the residual from $y_i - \hat{y}_i$, again assuming $\rho = 0$.

Now, consider the set of pairs such that $a_{ij} = 0$ (no link). Suppose that, for some pair of individuals (i, j) , \hat{a}_{ij} is “high” compared with the predicted value for the other linkages. This suggests that the link between i and j was more likely to be created than the other links (but is not linked in the data). Then, if $\mu > 0$, this implies that the value of $\hat{\varepsilon}_{ij}$ should be relatively large. Put differently, if the predicted value of the link is large, but no link is created, it implies that the unobserved shock on the link is strongly negative. If the model features homophily, this implies that the distance in the unobservables is large (*e.g.*, strong difference in preferences).

To our knowledge, no formal test of endogeneity exists which fully exploits the information in \hat{a}_{ij} and $\hat{\varepsilon}_{ij}$. In practice, researchers have used the fact that if the network is endogenous, “low-values” of \hat{a}_{ij} should be correlated with $\hat{\varepsilon}_{ij}$ (Liu et al., 2013a). However, as discussed above, this is far from exploiting all the information provided by \hat{a}_{ij} and $\hat{\varepsilon}_{ij}$. Also, if $\rho = 0$ or $\mu = 0$, the errors should not only be uncorrelated, but *independent*. Nonetheless, the same intuition holds. If $\rho = 0$ or $\mu = 0$, the residuals $\hat{\varepsilon}_{ij}$ from (10) should not provide any information on the probability that a link is created (conditional on the observables).

To explore further the empirical content of the model, we present numerical simulations. We show results for simple correlation tests, as well as for the Hilbert-Schmidt Independence Criterion tests.³⁰ Note also that this “test” is only valid when the network formation process is given by (11), *i.e.*, when the utility of the network is separable across links (Bramoullé and Fortin, 2010). However, we suspect that the same principle could be applied to more general models of network formation such as in Hsieh and Lee (2011), Mele (2013) or Chandrasekhar and Jackson (2013).

4.2.1. Simulations

We provide an example using Monte Carlo simulations for the model in (10), assuming that:

$$a_{ij} = \mathbf{1}\{\kappa - \phi|x_i - x_j| - \mu|\tilde{\nu}_i - \tilde{\nu}_j| + \eta_{ij} > 0\} \quad (12)$$

We use two specifications of the model. An “exogenous” version where ν and $\tilde{\nu}$ are two *independent and identically distributed* variables, and an “endogenous” version where $\nu = \tilde{\nu}$ (*i.e.* the same variable as in (10)).³¹

We let $x, \nu, \tilde{\nu}, \varepsilon, \eta \sim N(0, 1)$. Parameters are fixed such that $\phi = 0.5$, $\beta = 2$, $\gamma = 10$, $\kappa = 4$, $\mu = 6$ and $\rho = 20$. Parameters are chosen so as to facilitate the visual representation. We simulate 100 groups of sizes drawn from an *uniform* (10, 50). The total population is comprised of 2,944 individuals and 49,304 within-group pairs.

We first regress the predicted value of the link (*i.e.*, \hat{a}_{ij}) on the distance in the residuals of the outcome equation (*i.e.*, $\hat{\varepsilon}_{ij}$), and on the distance between the residuals when individuals are linked (*i.e.*, $\hat{\varepsilon}_{ij} * a_{ij}$) and on the status (linked or not) of the pair (*i.e.*, a_{ij}). If the

³⁰ The Hilbert-Schmidt Independence Criterion is a kernel-based independence test. It tests whether the (Hilbert-Schmidt) norm of the cross-covariance operator is equal to zero. For generic kernels, the Hilbert-Schmidt Independence Criterion between two random variables is equal to zero if and only if the variables are independent. See Gretton et al. (2005) for details. We computed the Hilbert-Schmidt Independence Criterion using Arthur Gretton’s (2007) *hsicTestGamma.m* function. See <http://people.kyb.tuebingen.mpg.de/arthur/indep.htm>.

³¹ This strategy is used so that the distribution of the unobservables in (10) and (12) will be the same, which facilitates the visual representation (see Figures 1-4).

network structure is exogenous, the only predictor of \hat{a}_{ij} should be a_{ij} . Looking at Table 1 (endogenous version), we see that we reject the null hypothesis that the network is exogenous since we measure a significant correlation between $\hat{\varepsilon}_{ij}$ and \hat{a}_{ij} . This result is confirmed by looking at Table 2 (exogenous version), where no correlation is measured.

The endogeneity created by the omitted variable ν can also be seen by looking at the conditional joint densities in Figures 1-4. If the network structure is exogenous, the realized network, (*i.e.*, the set of pairs of individuals that are actually linked or not) should not affect the joint distribution of $\hat{\varepsilon}_{ij}$ and \hat{a}_{ij} . When the network is endogenous, we see a large difference between the joint density for linked pairs (Figure 1), and the joint density of unlinked pairs (Figure 2). When the network is exogenous, we do not (and should not; see Figures 3 and 4). This visual analysis is confirmed by the Hilbert-Schmidt independence test in Table 3.

Since there exists no formal test exploiting all the empirical content of the model, we suggest that, in practice, researchers should proceed to a visual inspection of the conditional joint densities (as in Figures 1 and 2) and that they perform non-parametric independence tests such as the Hilbert-Schmidt Independence Criterion. Note that the visual inspection of the joint densities may provide more information as it gives an idea of the importance of the endogeneity. If joint densities are relatively similar, we can expect the bias due to the endogeneity to be small, even if evidence of endogeneity is found.

5. Using Proxy Variables

As in many economic applications, the researcher does not always observe the “true” variable and must sometimes use a proxy. In the peer effects literature, prominent examples are the use of student achievement as a proxy for effort, or time studying (Sacerdote, 2011), and body mass index (BMI) as a proxy for effort, or healthy life habits (Yakusheva et al., 2014). In this section, we use the example of BMI and present two model specifications. Since BMI is not directly transmitted from an individual to another nor can it be directly

chosen by the individuals, peer effects will take source in the unobserved individual effort. As we will see, this distinction is important as it deeply affects the causal interpretation of peer effects.

We assume that individual's i 's BMI (denoted by y_i) is linked to his observable characteristics, \mathbf{x}_i , his effort, e_i , and a random term, ν_i , in a linear way. The set of equations are given by:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\psi} + \rho\mathbf{e} + \boldsymbol{\nu} \quad (13)$$

where $\rho < 0$. We also assume that preferences are as follows:

$$u_i(\mathbf{e}, \mathbf{y}) = -(\mathbf{x}'_i\boldsymbol{\beta} + \varepsilon_i)y_i - \frac{e_i^2}{2} + \alpha e_i \mathbf{g}_i \mathbf{e} \quad (14)$$

The individual suffers a (linear) cost of being overweight (provided that $\mathbf{x}'_i\boldsymbol{\beta} + \varepsilon_i$ is positive) as well as a (quadratic) cost of making an effort. However, assuming complementarity, the cost of effort is lower if the individual's peers are also exerting effort ($\alpha > 0$).

Replacing (13) in (14) and solving for the first order conditions leads to: $\mathbf{e} = -\rho\mathbf{X}\boldsymbol{\beta} + \alpha\mathbf{G}\mathbf{e} - \rho\boldsymbol{\varepsilon}$. Note that this model only features endogenous interactions. Again, using (13), we can rewrite the model as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \alpha\mathbf{G}\mathbf{y} - \alpha\mathbf{G}\mathbf{X}\boldsymbol{\psi} - \boldsymbol{\eta} \quad (15)$$

where $\boldsymbol{\theta} = [\boldsymbol{\psi} - \rho^2\boldsymbol{\beta}]$ and $\boldsymbol{\eta} = \boldsymbol{\nu} - \alpha\mathbf{G}\boldsymbol{\nu} - \rho^2\boldsymbol{\varepsilon}$.

This model is then a special case of Lee et al. (2010) so all the parameters are identified, with the exception of ρ and $\boldsymbol{\beta}$. This reflects the loss of information due to the fact that the econometrician does not observe the effort.

Remark that (15) results from a model with an *endogenous* peer effect on the effort only. There is *no* direct contribution of the peers' characteristics on the optimal level of effort (that is, no contextual effects). This is fundamental for the economic interpretation of the composite parameter $\alpha\boldsymbol{\psi}$. It is usually interpreted as a peer effect coefficient capturing the

direct impact of the individuals' characteristics on the outcome. Here, however, it is simply a residual of the "true" social interaction parameter α , and not a second source of social interaction. The same reasoning holds for models of conformism.

5.1. Exogenous effect

Consider the following alternative model:

$$u_i(\mathbf{e}, \mathbf{y}) = -(\mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i) y_i - \frac{e_i^2}{2} - \frac{\lambda}{2} (y_i - \mathbf{g}_i \mathbf{y})^2 \quad (16)$$

Again, the individual suffers a cost of being overweight as well as a cost of making an effort. The individual also suffers a conformity cost on physical appearance. By substituting (13) in (16) and solving for the optimal choice of effort, we have:

$$\mathbf{e} = \mathbf{X}\boldsymbol{\zeta} + \tilde{\lambda}\mathbf{G}\mathbf{y} + \boldsymbol{\eta}$$

where $\boldsymbol{\zeta} = \frac{-e}{1+\lambda e^2} [\boldsymbol{\beta} + \lambda \boldsymbol{\psi}]$, $\tilde{\lambda} = \frac{\lambda e}{1+\lambda e^2}$, and $\boldsymbol{\eta} = \frac{-e}{1+\lambda e^2} [\boldsymbol{\varepsilon} + \lambda \boldsymbol{\nu}]$. We see that the optimal effort exerted by the individuals is a function of their peers' appearance. Rewriting the model in terms of the individual's BMI, we have:

$$\mathbf{y} = \mathbf{X}\check{\boldsymbol{\zeta}} + \check{\lambda}\mathbf{G}\mathbf{y} + \check{\boldsymbol{\eta}}$$

where $\check{\boldsymbol{\zeta}} = \frac{-1}{1+\lambda e^2} [\rho^2 \boldsymbol{\beta} - \boldsymbol{\psi}]$, $\check{\lambda} = \frac{\lambda e^2}{1+\lambda e^2}$, and $\check{\boldsymbol{\eta}} = \frac{-1}{1+\lambda e^2} [\rho^2 \boldsymbol{\varepsilon} - \boldsymbol{\nu}]$. Here, remark that the main parameter of interest, λ , is not identified. Identification results for linear models imply that $\frac{\rho^2 \boldsymbol{\beta} + \boldsymbol{\psi}}{1+\lambda e^2}$ and $\frac{\lambda e^2}{1+\lambda e^2}$ are identified, but it is not possible to recover λ from those two expressions. This is important for the interpretation of the model in terms of causal impact of social interactions. The correlation between y_i and $\mathbf{g}_i \mathbf{y}$ (given by $\frac{\lambda e^2}{1+\lambda e^2}$) *does not* have an interpretation in terms of peer effects as it also captures the impact of the individual's

effort on his BMI.³² This point is related to Angrist’s criticism (Angrist 2014), where the author is highly sceptical about the causal interpretation of the “peer effect” parameter $\check{\lambda}$. Angrist’s argument (see in particular section 6) partly relies on the fact that one can always build a microeconomic model, free of any peer interaction, that will lead to a positive $\check{\lambda}$, casting doubts on the causal interpretation of this parameter in terms of peer effects. Here, we argue that models of social interactions accounting for unobserved effort choices (as was carefully done by Cooley Fruehwirth (2013) for instance), brings credibility to the causal interpretation of the “social interaction” parameters.

6. Conclusion

In this chapter, we first present a survey of recent developments and challenges in the empirics of social networks as they are related to policy analysis. We then discuss the concept of social multiplier (Glaeser et al., 2003). We show that in the standard linear-in-means model the social multiplier is generally not identified except if we impose structure on the mechanisms of the social interactions (*e.g.*, complementarity *vs* pure conformity). We show that the presence of isolated individuals in the network helps to identify the social multiplier. However, we prove that estimating the linear-in-means model while including isolated individuals in the network will lead to biased estimators of peer effects when at least part the source of social interactions comes from conformity. We also provide a brief analysis of the key player (Borgatti, 2006; Ballester et al., 2006) and how one can identify him/her within a network.

We then discuss three types of social networks: non-stochastic (or fixed), stochastic exogenous and stochastic endogenous networks. We show that assuming that the network is non-stochastic or stochastic and exogenous has the same implication for the identification and the estimation of the model. However, the impact for the interpretation and policy

³² Note that testing $\check{\lambda} = 0$ allows to test for the *presence* of peer effects as $\check{\lambda}$ is proportional to the peer effect λ (assuming $\varrho < 0$). However, $\check{\lambda}$ does not provide an estimate of the size of the peer effect.

recommendations flowing from the model are not the same. The reason is that when the network is stochastic and exogenous, an exogenous policy intervention may affect the structure of the network. This must be taken into account in the evaluation of the effect of the policy. As far as endogenous networks are concerned, we suggest performing a Hilbert-Schmidt Independence Test to check the exogeneity in the network formation.

Finally, we analyse situations where some choice variables of the model are unobservable to the researcher (*e.g.*, healthy life habits), although proxy variables such as body mass index are observable. We show that, in this case, the issue of identifiability of the parameters crucially depends on the specification of the fundamentals of the model (preferences, technology, decision processes).

One basic message of our chapter is that in networks econometric studies, more care should be taken in developing a microeconomic model based on sound theoretical underpinnings to determine the fundamentals that are identified. This theoretical *identifiability problem* (Chiappori and Ekeland, 2009) is as important as the standard statistical *identification problem* (Manski, 1993; Bramoullé et al., 2009; Blume et al., 2013) in order to verify whether estimated peer effects can be interpreted as true *causal effects*.

Further research is needed to determine the conditions under which one can recover the fundamentals of the social network microeconomic model from its corresponding structural econometric model. Indeed, knowledge of the fundamentals is required in order to allow the researcher to simulate the impact of hypothetical reforms on individual outcomes within the network. Moreover, our chapter shows the crucial importance of advancing research on testing for network endogeneity.

References

- Babcock, P., Hartman, J., December 2010. Networks and workouts: Treatment size and status specific peer effects in a randomized field experiment. National Bureau of Economic Research Working Paper Series (16581).
- Badev, A., 2013. Discrete games in endogenous networks: Theory and policy, mimeo.
- Ballester, C., Calvó-Armengol, A., Zenou, Y., 2006. Who's who in networks. wanted: the key player. *Econometrica* 74 (5), 1403–1417.
- Barrera-Osorio, F., Bertrand, M., Linden, L. L., Perez-Calle, F., April 2011. Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia. *American Economic Journal: Applied Economics* 3 (2), 167–95.
- Bertrand, M., Luttmer, E. F. P., Mullainathan, S., August 2000. Network effects and welfare cultures. *The Quarterly Journal of Economics* 115 (3), 1019–1055.
- Beugnot, J., Fortin, B., Lacroix, G., Villeval, M.-C., 2014. Social networks and peer effects at work, mimeo.
- Blume, L. E., Brock, W. A., Durlauf, S. N., Ioannides, Y. M., 2011. Identification of Social Interactions. Vol. 1B of *Handbook of Social Economics*. North Holland, pp. 853–964.
- Blume, L. E., Brock, W. A., Durlauf, S. N., Jayaraman, R., 2013. Linear social interactions models, mimeo.
- Borgatti, S. P., 2006. Identifying sets of key players in a social network. *Comput Math Organiz Theor* 12.
- Boucher, V., 2014a. Conformism and self-selection in social networks, mimeo.
- Boucher, V., 2014b. Structural homophily. *International Economic Review*.
- Boucher, V., Bramoullé, Y., Djebbari, H., Fortin, B., 2014. Do peers affect student achievement? evidence from canada using group size variation. *Journal of Applied Econometrics* 29 (1), 91–109.
- Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. *Journal of Econometrics* 150 (1), 41–55.
- Bramoullé, Y., Fortin, B., 2010. social networks: econometrics. *The New Palgrave Dictionary of Economics*, ed. S. N. Durlauf, S. N. and L. E. Blume. Palgrave Macmillan.
- Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks* 30.

- Brock, W. A., Durlauf, S. N., 2001. Discrete choice with social interactions. *The Review of Economic Studies* 68 (2), pp. 235–260.
- Brock, W. A., Durlauf, S. N., 2007. Identification of binary choice models with social interactions. *Journal of Econometrics* 140.
- Calvó-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. *The Review of Economic Studies* 76 (4), 1239–1267.
- Card, D., Giuliano, L., 2013. Peer effects and multiple equilibria in the risky behavior of friends. *Review of Economics and Statistics* 4.
- Cassar, A., 2007. Coordination and cooperation in local, random and small world networks: Experimental evidence. *Games and Economic Behavior* 58 (2), 209–230.
- Chandrasekhar, A., Jackson, M., 2013. Tractable and consistent random graph models, mimeo.
- Chandrasekhar, A. G., Lewis, R., 2011. Econometrics of sampled networks, mimeo.
- Chiappori, P.-A., Ekeland, I., 2009. The microeconomics of efficient group behavior: Identification. *Econometrica* 77 (3), 763–799.
- Christakis, N. A., Fowler, J. H., July 2007. The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine* 357 (4), 370–379, access to full text is subject to the publisher’s access restrictions.
- Christakis, N. A., Fowler, J. H., Imbens, G. W., Kalyanaraman, K., 2010. An empirical model for strategic network formation, mimeo.
- Cohen-Cole, E., Fletcher, J. M., 2008. Is obesity contagious? social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics* 27 (5), 1382 – 1387.
- Comola, M., 2008. The network structure of informal arrangements: Evidence from rural tanzania, mimeo.
- Comola, M., Prina, S., 2013. Do Interventions Change the Network? A Dynamic Peer Effect Model Accounting for Network Changes, Mimeo, Paris School of Economics.
- Conti, G., Galeotti, A., Mueller, G., Pudney, S., October 2012. Popularity. Working Paper 18475, National Bureau of Economic Research.
- Coolley Fruehwirth, J. C., 2013. Identifying peer achievement spillovers: Implications for desegregation and the achievement gap. *Quantitative Economics* 4 (1), 85–124.

- Davezies, L., d'Haultfoeuille, X., Fougère, D., 2009. Identification of peer effects using group size. *Econometrics Journal* 12, 397–413.
- Dieye, R., Djebbari, H., Barrera-Osorio, F., 2014. Accounting for peer effects in treatment response, mimeo.
- Fafchamps, M., Gubert, F., 2007. Risk sharing and network formation. *American Economic Review* 97 (2), 75–79.
- Fortin, B., Yazbeck, M., 2014. Peer effects, fast food consumption and adolescent weight gain, mimeo.
- Galbiati, R., Zanella, G., 2012. The tax evasion social multiplier: Evidence from italy. *Journal of Public Economics* 96 (56), 485 – 494.
- Glaeser, E. L., Scheinkman, J. A., Sacerdote, B. I., 2003. The social multiplier. *Journal of the European Economic Association* 1 (2/3), pp. 345–353.
- Goldsmith-Pinkham, P., Imbens, G. W., 2013. Social networks and the identification of peer effects. *Journal of Business and Economic Statistics* 31 (3), 253 – 264.
- Graham, B. S., 2008. Identifying social interactions through conditional variance restrictions. *Econometrica* 76, 643–660.
- Graham, B. S., 2014. Methods of identification in social networks, mimeo.
- Gretton, A., Bousquet, O., Smola, A., Schölkopf, B., 2005. Measuring statistical dependence with hilbert-schmidt norms. In: *Algorithmic learning theory*. Springer, pp. 63–77.
- Hoff, P. D., Raftery, A. E., Handcock, M. S., 2002. Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97.
- Hsieh, C.-S., Lee, L.-F., 2011. A social interactions model with endogenous friendship formation and selectivity, mimeo.
- Jackson, M. O., 2010. *Social and economic networks*. Princeton University Press.
- Jackson, M. O., 2011. An overview of social networks and economic applications. *The Handbook of Social Economics.*, ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson. Elsevier Press.
- Karlan, D., Möbius, M., Rosenblat, T., Szeidl, A., 2009. Trust and social collateral. *The Quarterly Journal of Economics* 124 (3), 1307–1361.
- Kwok, H. H., 2013. Identification problems in linear social interaction models: A general analysis based on matrix spectral decompositions, mimeo.

- Lee, L., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140 (2), 333–374.
- Lee, L.-f., Li, J., Lin, X., 2013. Binary choice models with social network under heterogeneous rational expectations. *Review of Economics and Statistics* (00).
- Lee, L.-f., Liu, X., Lin, X., 2010. Specification and estimation of social interaction models with network structures. *The Econometrics Journal* 13 (2), 145–176.
- Lin, X., 2010. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics* 28 (4), 825–860.
- Liu, X., Patacchini, E., Rainone, E., Nov. 2013a. The Allocation of Time in Sleep: A Social Network Model with Sampled Data.
- Liu, X., Patacchini, E., Zenou, Y., 2013b. Peer effects: Social multiplier or social norms?, mimeo.
- Liu, X., Patacchini, E., Zenou, Y., Lee, L.-F., 2014. Criminal networks: Who is the key player?, mimeo.
- Manski, C., 2013. Identification of treatment response with social interactions. *Econometrics Journal* 16.
- Manski, C. F., 1993. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60 (3), 531–542.
- Mas, A., Moretti, E., 2009. Peers at work. *The American Economic Review* 99 (1), 112–45.
- Maurin, E., Moschion, J., 2009. The social multiplier and labor market participation of mothers. *American Economic Journal: Applied Economics* 1 (1), pp. 251–272.
- Mayer, A., Puller, S. L., 2008. The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics* 92 (12), 329 – 347.
- Mele, A., 2013. A structural model of segregation in social networks, mimeo.
- Moffitt, R. A., 2001. Policy Interventions, Low-Level Equilibria, and Social Interactions. MIT Press, pp. 45–82.
- Patacchini, E., Zenou, Y., 2008. The strength of weak ties in crime. *European Economic Review* 52 (2), 209–236.
- Sacerdote, B., 2001. Peer effects with random assignment: Results for dartmouth roommates. *Quarterly Journal of Economics* 116 (2), 681–704.

- Sacerdote, B., June 2011. Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? Vol. 3 of Handbook of the Economics of Education. Elsevier, Ch. 4, pp. 249–277.
- Trogdon, J. G., Nonnemaker, J., Pais, J., 2008. Peer effects in adolescent overweight. *Journal of Health Economics* 27 (5), 1388 – 1399.
- Yakusheva, O., Kapinos, K. A., Eisenberg, D., 2014. Estimating heterogeneous and hierarchical peer effects on body weight using roommate assignments as a natural experiment. *Journal of Human Resources* 49 (1), 234–261.
- Yang, C., Lee, L., 2014. Social interactions under incomplete information with heterogeneous expectations, mimeo.

Figure 1: Kernel density estimates, endogenous model, linked

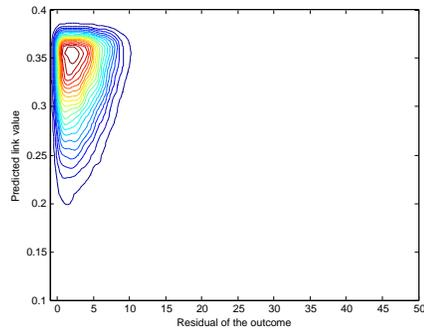


Figure 2: Kernel density estimates, endogenous model, no link

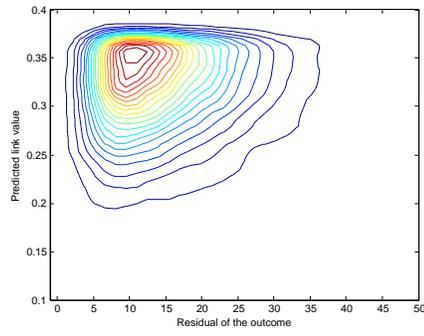


Figure 3: Kernel density estimates, exogenous model, linked

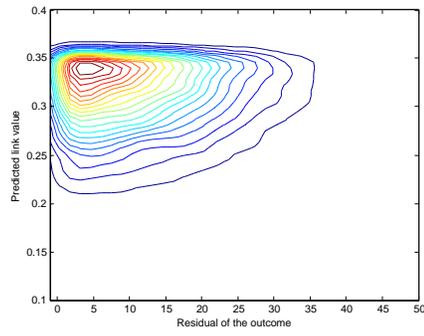


Figure 4: Kernel density estimates, exogenous model, no link

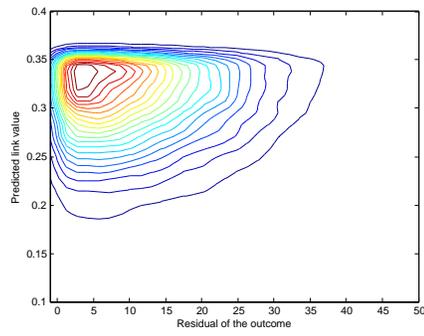


Table 1: Regression on predicted value of link (\hat{a}_{ij}): endogenous version

Variable	Coefficient	(Std. err.)
$\hat{\epsilon}_{ij}$	0.301**	(0.028)
$\hat{\epsilon}_{ij} * a_{ij}$	1.898**	(0.145)
a_{ij}	5.228**	(0.736)
Intercept	310.639**	(0.471)
Joint test for $[\hat{\epsilon}_{ij}] = [\hat{\epsilon}_{ij} * a_{ij}] = 0$: $F(2, 52385) = 177.45$		

Table 2: Regression on predicted value of link (\hat{a}_{ij}): exogenous version

Variable	Coefficient	(Std. Err.)
$\hat{\epsilon}_{ij}$	0.035	(0.024)
$\hat{\epsilon}_{ij} * a_{ij}$	-0.028	(0.042)
a_{ij}	7.187**	(0.617)
Intercept	301.958**	(0.342)
Joint test for $[\hat{\epsilon}_{ij}] = [\hat{\epsilon}_{ij} * a_{ij}] = 0$: $F(2, 49300) = 1.15$		

Table 3: Hilbert-Schmidt Independence Criterion

Test	Stat.	Thresh(0.025)	$\phi_c = 0.05$
Exogenous ($h_{ij} = 0$)	0.3237	0.7434	0.6732
Exogenous ($h_{ij} = 1$)	0.3918	0.7367	0.6636
Endogenous ($h_{ij} = 0$)	1.2885	0.7925	0.7391
Endogenous ($h_{ij} = 1$)	1.3692	0.6557	0.5823

Note: For computational reasons, tests are performed on a 10% sub-sample.

Appendix

PROOF (OF PROPOSITION 1). Using the fact that $(\mathbf{1} + \alpha\mathbf{G})^{-1} = \sum_{k=0}^{\infty} \alpha^k \mathbf{G}^k$, we have:

$$\mathbf{y}^u(c_1) = \mathbf{y}^u(c_0) + c \sum_{k=0}^{\infty} \alpha^k \mathbf{G}^k \mathbf{1} \text{ and } \mathbf{y}^v(c_1) = \mathbf{y}^v(c_0) + \frac{c}{1+\lambda} \sum_{k=0}^{\infty} \tilde{\alpha}^k \mathbf{G}^k \mathbf{1}.$$

Since $\mathbf{1}$ is an eigenvector of \mathbf{G} associated with the eigenvalue 1, this leads to $\mathbf{y}^u(c_1) = \mathbf{y}^u(c_0) + c \sum_{k=0}^{\infty} \alpha^k \mathbf{1} = \mathbf{y}^u(c_0) + \frac{c}{1-\alpha} \mathbf{1}$ and $\mathbf{y}^v(c_1) = \mathbf{y}^v(c_0) + \frac{c}{1+\lambda} \sum_{k=0}^{\infty} \tilde{\alpha}^k \mathbf{1} = \mathbf{y}^v(c_0) + c\mathbf{1}$, or $\Delta\mathbf{y}^u = \frac{c}{1-\alpha} \mathbf{1}$ and $\Delta\mathbf{y}^v = c\mathbf{1}$.