# Who is the Key Player? A Network Analysis of Juvenile Delinquency*†

Xiaodong Liu‡ Eleonora Patacchini§ Yves Zenou¶ Lung-Fei Lee‖

October 7, 2014

**Abstract**

We propose a new methodology to test who the key player is in delinquent networks, i.e. the delinquent who, once removed from the network, generates the highest possible reduction in aggregate delinquency level. Using data on adolescent delinquents in the United States, we then provide new results regarding the identification of peer effects and determine the key player in each delinquent network. We show that, compared to a policy that removes the most active delinquent from the network, a key player policy engenders a much higher delinquency reduction.

**Key words**: Juvenile delinquency, Katz-Bonacich centrality, estimation of peer effects, crime policies.

**JEL Classification:** A14, D85, K42, Z13

*We would like to thank the editor, Fabrizio Zilibotti, as well as six referees for very helpful comments. We also thank the participants of the Einaudi Institute of Economics and Finance lunch seminar, the conference on "The Economics of Coordination and Communication: Theory and Empirics" at Ravello in 2012, the Iowa State University departmental seminar, the IIES seminar, and the Italian Association of Labor Economics conference in 2011 for their comments, in particular, Jason Fletcher, David Levine, Matthew Lindquist, Marcus Mobius, Luigi Pistaferri and Tanya Rosenblat. We are grateful to Tiziano Arduini and Edoardo Rainone for excellent research assistance and Dina Neiman for her editorial services. This paper was previously circulated under the title "Criminal Networks: Who is the Key Player?".

†This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis.

‡University of Colorado Boulder, USA. E-mail: xiaodong.liu@colorado.edu.

§Cornell University, EIEF and CEPR. E-mail: ep454@cornell.edu.

¶Corresponding author. Stockholm University, Research Institute of Industrial Economics (IFN) and GAINS. E-mail: yves.zenou@ne.su.se.

‖The Ohio State University, USA. E-mail: lee.1777@osu.edu.

## 1 Introduction

There are 2.3 million people behind bars at any one time in the United States, and that number continues to grow. It is the highest level of incarceration per capita in the world. Moreover, since the crime explosion of the 1960s, the prison population in the United States has multiplied fivefold, to one prisoner for every hundred adults — a rate unprecedented in American history and unmatched anywhere in the world.[1] However, in spite of a continuously falling crime rate, the prisoner head count continues to rise, and poor people as well as minorities still bear the brunt of both crime and punishment.

One possible way to reduce crime is to detect, apprehend, convict, and punish criminals. This is what has been done in the United States and all of these actions cost money, currently about $200 billion per year nationwide. For example, in California, even if this "brute force" policy has partly worked (the rate of every major crime category is now less than half of what it was 20 years ago), the cost of this policy has been tremendous. Consider, for example, that the cost of the justice system is higher than the cost of education.[2]

In his recent book published in 2009, Mark Kleiman argues that simply locking up more people for lengthier terms is no longer a workable crime-control strategy. But, says Kleiman, there has been a revolution in controlling crime by means other than brute-force incarceration: substituting swiftness and certainty of punishment for randomized severity, concentrating enforcement resources rather than dispersing them, communicating specific threats of punishment to specific offenders, and enforcing probation and parole conditions to make community corrections a genuine alternative to incarceration. As Kleiman shows, "zero tolerance" is nonsense: there are always more offenses than there is punishment capacity.

Is there an alternative to brute force? In this paper, we argue that concentrating efforts by targeting "key players", i.e. players who once removed generate the highest possible reduction in aggregate activity level in a network. The "key players" policy can be more effective in reducing delinquency and crime because of the snow-ball effects or "social multipliers" at work (see, in particular, Kleiman, 2009; Glaeser et al., 1996; Verdier and Zenou, 2004; Calvó-Armengol and Zenou, 2004).[3] Furthermore, the impact of social networks may be particularly important for adolescents because this developmental period overlaps with the initiation and continuation of many risky, unhealthy, and delinquent behaviors and is a period of maximal response to peer pressure (Thornberry et al., 2003; Warr, 2002).

It is indeed well-established that delinquency and crime are, to some extent, a group phenomenon, and that the sources of delinquency and crime are located in the intimate social networks of individuals (see e.g. Sarnecki, 2001; Warr, 2002; Haynie, 2001; Patacchini and Zenou, 2012). Delinquents often have friends who have themselves committed several offences, and social ties among delinquents are seen as a means whereby individuals exert an influence over one another to commit

[1] See Cook and Ludwig (2010) and the references therein.

[2] For example, the "Three Strikes" law passed in California in 1994 mandates extremely long prison terms (between 29 years and life) for anyone previously convicted in two serious or violent felonies (including residential burglary) when convicted of a third felony, even for something as minor as petty theft.

[3] See Goyal (2007), Jackson (2008), Ioannides (2012) and Jackson and Zenou (2014) for overviews on network theory.

crimes. In fact, not only friends, but also the *structure* of social networks, matters in explaining an individual's own delinquent behavior. This suggests that the underlying structural properties of friendship networks must be taken into account to better understand the impact of peer influence on delinquent behavior and to address adequate and novel delinquency-reducing policies.

The aim of this paper is to investigate the importance of the key player in delinquent networks by *proposing a new methodology* that helps detect the key player in the real world. For that, we extend the theoretical model of Ballester et al. (2006) to incorporate local-aggregate (i.e. *strategic complementarity*) as well as local-average (i.e. *taste for conformity*) effects in the utility function and identify the key player. We then bring this model to the data. In order to determine the key player, we need to estimate the intensity of interactions between criminals in each network. Most of the network models in the literature use the identification approach proposed by Bramoullé et al. (2009), which assumes that the adjacency matrix (or sociomatrix) is *row-normalized* (the *local-average model*). With a row-normalized adjacency matrix, a criminal's effort level depends on the *average* effort level of her criminal friends (Patacchini and Zenou, 2012). However, key-player policies also make sense in the context of the *local-aggregate model* developed by Ballester et al. (2006) with a *non-row-normalized* adjacency matrix, where it is the *sum* of friends' efforts that affects own effort. In this paper, we provide a new identification strategy for the local-aggregate model which is based on the variation of the row sum of the adjacency matrix. We show that, in general, the identification condition for the local-aggregate model is weaker than that for the local-average model developed by Bramoullé et al. (2009).

Using the Add Health data of adolescents in the United States, using the general model with both local-average and local-aggregate effects, we then determine who is the key player in each of our 103 networks. We find that the key player is *not* necessarily the most active delinquent in the network. We also find that it is *not* straightforward to determine which delinquent should be removed from a network only based on her position in the network. Compared to other students, the key players are older and in higher grades, are less likely to be white, have lower math scores, have lower self esteem, and are less likely to feel being part of school or being safe in school. They are also less likely to feel that their parents care about them, and are less likely to come from families where parents work as professionals.

We finally discuss the policy implications of our results. First, we show that targeting the most active delinquent is less effective than the key-player policy. Second, our key-player policy can be directly applied to reduce crime in the real world when criminal network data are available (we provide a selective list of such data). In fact, some similar policies aiming at reducing crime have already been implemented in the United States but without the analytical tools of the key-player policy. Finally, we show that we can use our methodology to determine the key player in other types of networks and activities, where network data are easier to obtain. This is particularly true for financial networks, R&D networks, networks in developing countries and political networks.

The rest of the paper unfolds as follows. In the next section, we discuss the related literature and explain our contribution. The theory is exposed in Section 3. Our data are described in Section 4. In Section 5, the identification of the econometric network model is discussed while the estimation and empirical results of the impact of peer effects on crime are provided. Section 6 characterizes the key players. In Section 7, we discuss the implications of the key-player policy. Finally, Section 8 concludes.

## 2 Related literatures

Our paper lies at the intersection of different literatures. We would like to expose them in order to highlight our contribution.

**Empirical studies of peer effects in crime** There is a growing body of empirical literature suggesting that peer effects are very strong in criminal decisions. Ludwig et al. (2001) and Kling et al. (2005) study the relocation of families from high- to low-poverty neighborhoods using data from the Moving to Opportunity (MTO) experiment. They find that this policy reduces juvenile arrests for violent offences by 30 to 50 percent, relative to a control group. This also suggests very strong social interactions in crime behaviors. Patacchini and Zenou (2012) find that peer effects in crime are strong, especially for petty crimes. Bayer et al. (2009) consider the influence that juvenile offenders serving time in the same correctional facility have on each other's subsequent criminal behavior. They also find strong evidence of learning effects in criminal activities since exposure to peers with a history of committing a particular crime increases the probability that an individual who has already committed the same type of crime recidivates that crime.

Our contribution here is to structurally estimate a network model and to propose a way to identify the key player in criminal networks. This paper is the first to present an empirical implementation of the key player policy, which is grounded on a precise behavioral foundation.

**Econometrics of networks** The literature on identification and estimation of social network models has progressed significantly recently (see Blume et al., 2011 and Durlauf and Ioannides, 2010, for recent surveys). In his seminal work, Manski (1993) introduces a linear-in-means social interaction model with endogenous effects, contextual effects, and correlated effects. Manski shows that the linear-in-means specification suffers from the "reflection problem" and the different social interaction effects cannot be separately identified.[4] Bramoullé et al. (2009) generalize Manski's linear-in-means model to a general local-average social network model, whereas the endogenous effect is represented by the average outcome of the peers. They provide some general conditions for the identification of the local-average model (i.e. when the adjacency matrix is row-normalized) using an indirect connection's characteristics as an instrument for the endogenous effect.

Our contribution to this literature is to provide an identification condition for the local-aggregate model (i.e. when the adjacency matrix is not row-normalized). We show that, in general, the identification condition of the local-aggregate model is weaker than that of the local-average model.

**The key-player problem** The problem of identifying key players has a long tradition in the sociological literature, which have proposed different measures of network centralities to define the key players (see, in particular, Wasserman and Faust, 1994). Borgatti (2003, 2006) was among the first researchers to analytically study the issue of key players by explicitly measuring the contribution of a set of actors to the cohesion of a network. Borgatti measures the amount of *reduction* in

[4] Lee (2007a) considers a model with multiple networks where an agent is equally influenced by all the other agents in the same network. Lee's social interaction model is identifiable only if there is variation in network size in the sample. The identification, however, can be weak if all of networks are large.

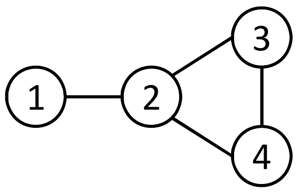
cohesiveness of the network that would occur if some nodes were not present. In the economics literature, Ballester et al. (2006, 2010) were the first to define the key-player problem in terms of *behavior* of agents so that total activity is measured as the sum of efforts of all agents at a Nash equilibrium.

This is the *first paper* that proposes a methodology to test the key-player policy. Our contribution is thus mainly *methodological*. We also provide concrete recommendations on how to implement the key player policy in the real world, primarily from criminal networks but also for financial, R&D, development, political and tax-evasion networks.

## 3 Theoretical framework

### 3.1 Model and Nash equilibrium

Suppose $N = \{1, \ldots, n\}$ is a finite set of agents, each corresponds to a node in a network $g \equiv g_N$. We keep track of social connections in network $g$ through its adjacency matrix $\mathbf{G} = [g_{ij}]$, where $g_{ij} = 1$ if nodes $i$ and $j$ $(i \neq j)$ are connected and $g_{ij} = 0$ otherwise.[5] We set $g_{ii} = 0$. For network $g$ with adjacency matrix $\mathbf{G}$, the $k$-th power of $\mathbf{G}$ given by $\mathbf{G}^k$ keeps track of direct and indirect connections in the network. More precisely, the $(i, j)$-th entry of $\mathbf{G}^k$ gives the number of paths of length $k$ from node $i$ to node $j$ in network $g$. In particular, $\mathbf{G}^0 = \mathbf{I}$. Note that, by definition, a path between $i$ and $j$ needs not follow the shortest possible route between those nodes. For instance, if $g_{ij} = g_{ji} = 1$, then the sequence $i \to j \to i \to j$ constitutes a path of length three from $i$ to $j$. Furthermore, let $\mathbf{G}^* = [g^*_{ij}]$, with $g^*_{ij} = g_{ij} / \sum_{j=1}^n g_{ij}$,[6] denote the row-normalized adjacency matrix. An example is given in Figure 1 for a tree network with four agents.




$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \qquad \mathbf{G}^* = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

Figure 1: a tree network and corresponding adjacency matrices.

Agents in network $g$ decide how much effort to exert in delinquent activities. We denote by $y_i$ the effort level of agent $i$ and by $\mathbf{y} = (y_1, ..., y_n)'$ the population effort profile in network $g$. Each agent $i$ selects an effort $y_i \geq 0$, and obtains a payoff $u_i(\mathbf{y}, g)$ that depends on the effort profile $\mathbf{y}$ and on the underlying network $g$, in the following way:

$$u_i(\mathbf{y}, g) = \underbrace{(\pi_i + \lambda_1 \textstyle\sum_{j=1}^n g_{ij} y_j) y_i}_{\text{benefits}} - \underbrace{[p\, f\, y_i + \frac{1}{2} y_i^2 + \frac{1}{2}\lambda_2 (y_i - \textstyle\sum_{j=1}^n g^*_{ij} y_j)^2]}_{\text{cost}} \tag{1}$$

with $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. This utility has a standard cost-benefit structure (as in Becker, 1968). The

[5] For the ease of the presentation, we focus on undirected unweighted networks so that $\mathbf{G}$ is a $(0, 1)$ symmetric square matrix. All our theoretical results also hold for directed and/or weighted networks since the (a)symmetry of the adjacency matrix $\mathbf{G}$ does not play any role in the proof of our theoretical results.

[6] For simplicity, we assume none of the agents in the network are isolated so that $\sum_{j=1}^n g_{ij} \neq 0$ for all $i$.

payoff increases in own effort $y_i$ with the marginal benefit given by $\pi_i^* + \lambda_1 \sum_{j=1}^n g_{ij} y_j$. The term $\pi_i^*$ denotes the *exogenous heterogeneity* of agent $i$'s "productivity" in delinquent activities, and is given by

$$\pi_i = \mathbf{x}_i' \boldsymbol{\beta}_1^* + \textstyle\sum_{j=1}^n g_{ij}^* \mathbf{x}_j' \boldsymbol{\beta}_2^* + \xi + \epsilon_i^*. \tag{2}$$

where $\mathbf{x}_i = (x_{1,i}, ..., x_{k_x,i})'$ is a $k_x \times 1$ vector that captures the *observable* exogenous characteristics of agent $i$ (e.g. age, sex, race, parental education, etc.), $\sum_{j=1}^n g_{ij}^* \mathbf{x}_j'$ is the *average* exogenous characteristics of agent $i$'s connections, with its coefficient vector $\boldsymbol{\beta}_2^*$ representing *contextual effects* (Manski, 1993), $\xi$ denotes the *unobservable* (to the researcher) exogenous network characteristics, e.g., the prosperity level of the neighborhood of network $g$ (i.e. more prosperous neighborhoods may lead to higher proceeds from delinquent activities) and $\epsilon_i^*$ denotes the *unobservable* (to the researcher) characteristics of agent $i$.

Compared to the standard crime model (Becker, 1968), the new element in the benefit part of (1 is the term $\lambda_1 \sum_{j=1}^n g_{ij} y_j$, which reflects the influence of the total effort of an agent's connections on her own "productivity". Indeed, an agent may benefit directly from the effort of her connections if they are co-offenders in some delinquent activity. An agent may also benefit indirectly through the form of know-how sharing about delinquent behavior with her friends.[7] We assume that the more delinquent connections an agent has and the more these connections are involved in delinquent activities, the higher is the marginal payoff of the agent's own delinquent effort. Thus, we call $\lambda_1$ the *social-multiplier* coefficient.

The cost part of the utility has three components. The cost of being caught is captured by the probability of being caught $0 < p < 1$ times the fine $f\, y_i$, which increases with the effort level $y_i$, as the severity of the punishment increases with one's involvement in delinquent activities. Also, individuals have a *direct* cost of exerting effort given by $\frac{1}{2} y_i^2$. Finally, different from Ballester et al. (2006, 2010), the cost in (1) has an additional term $\frac{1}{2}\lambda_2 (y_i - \sum_{j=1}^n g_{ij}^* y_j)^2$, which represents the moral cost due to deviation from the *social norm* of the reference group (i.e., the average effort of agent $i$'s connections). We call $\lambda_2$ the *social-conformity* coefficient.

At equilibrium, each agent maximizes her utility (1) and the best-response function of each agent $i$ is given by:

$$y_i = \phi_1 \sum_{j=1}^n g_{ij} y_j + \phi_2 \sum_{j=1}^n g_{ij}^* y_j + \alpha_i, \tag{3}$$

where $\phi_1 = \lambda_1/(1+\lambda_2)$, $\phi_2 = \lambda_2/(1+\lambda_2)$, and

$$\alpha_i = (\pi_i - p\, f)/(1+\lambda_2) = \mathbf{x}_i' \boldsymbol{\beta}_1 + \textstyle\sum_{j=1}^n g_{ij}^* \mathbf{x}_j' \boldsymbol{\beta}_2 + \eta + \epsilon_i \tag{4}$$

with $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*/(1+\lambda_2)$, $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^*/(1+\lambda_2)$, $\eta = (\xi - p\, f)/(1+\lambda_2)$ and $\epsilon_i = \epsilon_i^*/(1+\lambda_2)$. As $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, we have $\phi_1 \geq 0$ and $0 \leq \phi_2 < 1$. The coefficient $\phi_1$ captures the *local-aggregate* endogenous peer effect. As $\phi_1 \geq 0$, this coefficient reflects *strategic complementarity* in efforts. The coefficient $\phi_2$ captures the *local-average* endogenous peer effect, which reflects the *taste for conformity.* Note

[7] Sutherland (1947) and Akers (1998) expressly argue that criminal behavior is learned from others in the same way that *all* human behavior is learned. Indeed, young people may be influenced by their peers in all categories of behavior - music, speech, dress, sports, and *delinquency.*

that, $\phi_1/\phi_2 = \lambda_1/\lambda_2$. That is, the relative magnitude of $\phi_1$ and $\phi_2$ is the same as that of the social-multiplier coefficient $\lambda_1$ and the social-conformity coefficient $\lambda_2$.

Let $\bar{g}_i = \sum_{j=1}^{n} g_{ij}$ denote the degree of node $i$ in network $g$. Let $\bar{g}^{\max} = \max_i \bar{g}_i$ denote the highest degree in network $g$. It is easily shown that, if $\phi_1 \geq 0$, $\phi_2 \geq 0$ and $\bar{g}^{\max}\phi_1 + \phi_2 < 1$, the network game with utility (1) has a unique Nash equilibrium in pure strategies given by:

$$\mathbf{y}^* = (\mathbf{I}_n - \phi_1\mathbf{G} - \phi_2\mathbf{G}^*)^{-1}\boldsymbol{\alpha}, \tag{5}$$

where $\mathbf{y}^* = (y_1^*, ..., y_n^*)'$ is the equilibrium effort vector, $\mathbf{I}_n$ is the $(n \times n)$ identity matrix, and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)'$.

Two special cases of this network game are of particular interest. The first case is when $\lambda_2 = 0$ in the utility (1). In this case, $\phi_2 = 0$ and the best-response function becomes

$$y_i = \phi_1 \sum_{j=1}^{n} g_{ij} y_j + \alpha_i,$$

with equilibrium effort vector given by

$$\mathbf{y}^* = (\mathbf{I}_n - \phi_1\mathbf{G})^{-1}\boldsymbol{\alpha}.$$

As the equilibrium effort of an agent depends on the aggregate effort of her connections, we call this case the *local-aggregate* network game. The other case is when $\lambda_1 = 0$ in the utility (1). In this case, $\phi_1 = 0$ and the best-response function becomes

$$y_i = \phi_2 \sum_{j=1}^{n} g_{ij}^* y_j + \alpha_i,$$

with equilibrium effort vector given by

$$\mathbf{y}^* = (\mathbf{I}_n - \phi_2\mathbf{G}^*)^{-1}\boldsymbol{\alpha}.$$

As the equilibrium effort of an agent depends on the average effort of her connections, we call this case the *local-average* network game.

### 3.2 Finding the key player

A *key player* is the agent whose removal from the network leads to the largest reduction in the aggregate effort level in a network. Let $\mathbf{M}(g, \phi_1, \phi_2) = (\mathbf{I} - \phi_1\mathbf{G} - \phi_2\mathbf{G}^*)^{-1}$, with its $(i, j)$-th entry denoted by $m_{ij}(g, \phi_1, \phi_2)$. Let

$$\mathbf{b}(g, \phi_1, \phi_2, \boldsymbol{\alpha}) = \mathbf{M}(g, \phi_1, \phi_2)\boldsymbol{\alpha}$$

with its $i$-th entry denoted by $b_i(g, \phi_1, \phi_2, \boldsymbol{\alpha}) = \sum_{j=1}^{n} m_{ij}(g, \phi_1, \phi_2)\alpha_j$. Let $B(g, \phi_1, \phi_2, \boldsymbol{\alpha}) = \sum_{i=1}^{n} b_i(g, \phi_1, \phi_2, \boldsymbol{\alpha}) = \boldsymbol{\iota}_n'\mathbf{M}(g, \phi_1, \phi_2)\boldsymbol{\alpha}$ denote the aggregate effort level in network $g$, where $\boldsymbol{\iota}_n$ is an $n \times 1$ vector of ones. Let $g^{[-i]}$ denote the network with agent $i$ removed. Let $\mathbf{G}^{[-i]}$ and

$\boldsymbol{\alpha}^{[-i]}$ denote the adjacency matrix and vector of covariates corresponding to the remaining agents in network $g^{[-i]}$. Then, the key player $i^*$ in network $g$ is given by $i^* = \arg\max_i d_i(g, \phi_1, \phi_2, \boldsymbol{\alpha})$, where

$$d_i(g, \phi_1, \phi_2, \boldsymbol{\alpha}) = B(g, \phi_1, \phi_2, \boldsymbol{\alpha}) - B(g^{[-i]}, \phi_1, \phi_2, \boldsymbol{\alpha}^{[-i]}). \tag{6}$$

### 3.2.1 The key player in the local-aggregate network game

Ballester et al. (2006, 2010) and Ballester and Zenou (2014) have studied the key-player policy for the *local-aggregate* network game with $\lambda_2 = 0$ in (1). Observe that, in this case, $\phi_2 = 0$ and $b_i(g, \phi_1, 0, \boldsymbol{\iota})$ is known as the *Katz-Bonacich centrality* of node $i$ (Ballester et al., 2006). Let $\boldsymbol{\alpha}^{[i]}$ denote the vector of covariates calculated based on the network consisting $g^{[-i]}$ and the isolated $i$. It follows from Ballester and Zenou (2014) that the key player can be determined by the *generalized intercentrality.*

**Proposition 1** *For network $g$, let the generalized intercentrality of node $i$ be denoted by*

$$d_i(g, \phi_1, 0, \boldsymbol{\alpha}) = \underbrace{\boldsymbol{\iota}_n' \mathbf{M}(g, \phi_1, 0)(\boldsymbol{\alpha} - \boldsymbol{\alpha}^{[i]})}_{contextual\ variable\ change\ effect} + \underbrace{\frac{b_i(g, \phi_1, 0, \boldsymbol{\alpha}^{[i]}) \sum_{j=1}^n m_{ji}(g, \phi_1, 0)}{m_{ii}(g, \phi_1, 0)}}_{network\ structure\ change\ effect}. \tag{7}$$

*Then, agent $i^*$ is the key player of the local-aggregate network game if and only if $i^*$ has the highest generalized intercentrality in network $g$.*

The generalized intercentrality (7) highlights the fact that when an agent is removed from a network, two effects are at work. The first one is the *contextual variable change effect*, which is due to the change in $\boldsymbol{\alpha}$ after the removal of an agent. The second effect is the *network structure change effect*, which captures the change in $\mathbf{G}$ when an agent is removed. More generally, the generalized intercentrality measure accounts both for one's exposure to the rest of the group and for one's contribution to every other exposure.

### 3.2.2 The key player in network games with the local-average peer effect

To the best of our knowledge, nobody has studied the key-player policy for the local-average model. To have some intuition, consider the local-average network game for the network in Figure 1. Suppose that the agents are homogeneous *ex ante* with $\mathbf{x}_i = 1$ and $\eta = \epsilon_i = 0$ for $i = 1, 2, 3, 4$. We assume $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = 1$ and thus by (4)

$$\boldsymbol{\alpha} = \begin{bmatrix} \mathbf{x}_1 + \mathbf{x}_2 \\ \mathbf{x}_2 + \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_3 + \mathbf{x}_4) \\ \mathbf{x}_3 + \frac{1}{2}(\mathbf{x}_2 + \mathbf{x}_4) \\ \mathbf{x}_4 + \frac{1}{2}(\mathbf{x}_2 + \mathbf{x}_3) \end{bmatrix} = 2\boldsymbol{\iota}_4.$$

For the local-average network game with $\phi_1 = 0$, the equilibrium effort vector is given by $\mathbf{b}(g, 0, \phi_2, \boldsymbol{\alpha}) = \mathbf{M}(g, 0, \phi_2)\boldsymbol{\alpha} = 2(\mathbf{I} - \phi_2 \mathbf{G}^*)^{-1}\boldsymbol{\iota}_4 = 2\boldsymbol{\iota}_4/(1 - \phi_2)$. Observe that, although the 4 agents have different positions (e.g. different degrees) in the network, their equilibrium effort levels are identical. The aggregate effort level is $B(g, \phi_1, \phi_2, \boldsymbol{\alpha}) = 8\boldsymbol{\iota}_4/(1 - \phi_2)$.

Suppose agent 1 is removed from the network. Then,

$$\mathbf{G}^{[-1]*} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\alpha}^{[-1]} = \begin{bmatrix} \mathbf{x}_2 + \frac{1}{2}(\mathbf{x}_3 + \mathbf{x}_4) \\ \mathbf{x}_3 + \frac{1}{2}(\mathbf{x}_2 + \mathbf{x}_4) \\ \mathbf{x}_4 + \frac{1}{2}(\mathbf{x}_2 + \mathbf{x}_3) \end{bmatrix} = 2\boldsymbol{\iota}_3$$

and the equilibrium effort vector is given by $\mathbf{b}(g^{[-1]}, 0, \phi_2, \boldsymbol{\alpha}^{[-1]}) = \mathbf{M}(g^{[-1]}, 0, \phi_2)\boldsymbol{\alpha}^{[-1]} = 2(\mathbf{I} - \phi_2 \mathbf{G}^{[-1]*})^{-1}\boldsymbol{\iota}_3 = 2\boldsymbol{\iota}_3/(1-\phi_2)$. Observe that the equilibrium effort of the remaining agents are not affected by the removal of agent 1. Similarly, we can show that $\mathbf{b}(g^{[-3]}, 0, \phi_2, \boldsymbol{\alpha}^{[-3]}) = \mathbf{b}(g^{[-4]}, 0, \phi_2, \boldsymbol{\alpha}^{[-4]}) = 2\boldsymbol{\iota}_3/(1-\phi_2)$. That is, the equilibrium effort of the remaining agents are not affected by the removal of agent 3 or 4 either. Although agents 1, 3 and 4 have different positions in the network, removing any one of them leads to the same amount of reduction in the aggregate effort level.

On the other hand, agent 2 has a very special position in the network, as the removal of agent 2 breaks the link between agent 1 and the other agents. Indeed, when agent 2 is removed, we have

$$\mathbf{G}^{[-2]*} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\alpha}^{[-2]} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_3 + \mathbf{x}_4 \\ \mathbf{x}_4 + \mathbf{x}_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

and $\mathbf{b}(g^{[-2]}, 0, \phi_2, \boldsymbol{\alpha}^{[-2]}) = \mathbf{M}(g^{[-2]}, 0, \phi_2)\boldsymbol{\alpha}^{[-2]} = (\mathbf{I} - \phi_2 \mathbf{G}^{[-2]*})^{-1}\boldsymbol{\alpha}^{[-2]} = [1, 2/(1-\phi_2), 2/(1-\phi_2)]'$. As $0 < \phi_2 < 1$, the equilibrium effort of agent 1 is less than $2/(1-\phi_2)$. Hence, removal of agent 2 leads to the largest reduction in the aggregate effort level and agent 2 is the key player.

From the above example, we can see that, in general, when the agents are *ex ante* homogeneous, which agent to remove from the network does not matter in terms of the aggregate effort level reduction, unless the agent holds a very special position in the network such that removing this agent generates isolated nodes in the network.

If the agents have different values of $\mathbf{x}_i$, the key-player problem for the local-average network game and the general network game with utility (1) does not have an analytical solution. The difficulty comes from the fact that the row-normalized adjacency matrix of network $g^{[-i]}$ is, in general, not a submatrix of the row-normalized adjacency matrix of network $g$ (as one can see that $\mathbf{G}^{[-1]*}$ is not a submatrix of $\mathbf{G}^*$ in the above example). Yet we can still determine the key player numerically using its definition given by (6) if we can estimate the unknown parameters in the best-response function (5). In the rest of paper, we present an empirical investigation of the key-player problem for the general network game with both local-aggregate and local-average peers effects.

# 4 Data

Our analysis is made possible by the use of a unique database on friendship networks from the National Longitudinal Survey of Adolescent Health (Add Health). The Add Health database has been designed to study the impact of the social environment (i.e. friends, family, neighborhood and school) on adolescents' behavior in the United States by collecting data on students in grades 7-12 from a nationally representative sample of roughly 130 private and public schools in years 1994-95. Every student attending the sampled schools on the interview day is asked to complete a questionnaire (*in-school survey*) containing questions on respondents' demographic and behavioral characteristics, education, family background and friendships. This sample contains information on roughly 90,000 students. A subset of adolescents selected from the rosters of the sampled schools, about 20,000 individuals, is then asked to complete a longer questionnaire containing more sensitive individual and household information (*in-home survey* and parental data). Our analysis only uses the in-school survey data.

For the purposes of our analysis, the most interesting aspect of the Add Health data is the information on friendships that is based upon actual friend nominations. Students were asked to identify their best friends from a school roster (up to five males and five females).[8] We assume friendship to be reciprocal so that the adjacency matrix $\mathbf{G} = [g_{ij}]$ is symmetric with $g_{ij} = 1$ if $i$ nominates $j$ as a friend or $j$ nominates $i$ as a friend.

The dependent variable in our analysis is an index of participation in delinquent activities. In the Add Health, the following question about delinquent activities is asked to each student in the in-school survey: "In the past 12 months, how often did you: (*i*) smoke cigarettes; (*ii*) drink beer, wine, or liquor; (*iii*) get drunk; (*iv*) race on a bike, on a skateboard or roller blades, or in a boat or car; (*v*) do something dangerous due to dare; (*vi*) lie to your parents or guardians; (*vii*) skip school without an excuse?" The answers are coded using an ordinal scale from 0 (never), 1 (once or twice), 2 (once a month or less), 3 (2 or 3 days a month), 4 (once or twice a week), 5 (3-5 days a week), and 6 (nearly everyday). The delinquency index $y_i$ is given by the average frequency of the above delinquent behaviors of student $i$.[9]

After removing isolated students and pairs (i.e. network with only two students) as well as students with missing information, the sample consists of 64,186 students distributed over 206 networks[10], with network size ranging from 3 to 1,786. Because the strength of peer effect may vary with network sizes (see Calvó-Armengol et al., 2009), we only consider networks with sizes under the 50th percentile of the network size distribution.[11]

Our selected sample consists of 3,786 students distributed over 103 networks, with network sizes ranging from 3 to 200. The median, mean and the standard deviation of network sizes are, respectively, 4, 36.75 and 54.60. Furthermore, in our sample, the average number of friends of a student is 5.27 with a standard deviation of 3.59. A summary of the data can be found in Table 1. Table 1 also shows that the sample we use does not lose representativeness of the Add Health data.

[*Insert Table* 1 *here*]

[8] The limit in the number of nominations is not binding (even by gender). Less than 1% of the students in our sample show a list of ten best friends.

[9] In some sense, this delinquency index measures more a *risky* behavior than a *criminal* behavior. In the sample considered, only about 5% of the students claimed they never participated in the above activities in the past 12 months. In our empirical analysis, we consider everybody, even those who declare not having participated to any delinquent activity.

[10] A network is defined as the largest set of students who are directly or indirectly connected through friend nomination. By definition, students from two different networks cannot be friends.

[11] Indeed, if we use all the networks in the estimation, then we need to assume that the strength of peer effect is the same in a network with 3 students and in a network with 1,786 students. This is a rather strong assumption. Hence, we only consider networks with sizes under the 50th percentile of the network size distribution to avoid extremely big networks. We show below that our sub-sample does not lose representativeness compared to the whole sample.

## 5 Peer effects and network centrality

### 5.1 Econometric model

To determine the key player, we need to estimate the vector of parameters $\boldsymbol{\theta} = (\phi_1, \phi_2, \boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ in the best-response function (3). Let $\bar{r}$ be the total number of networks, $n_r$ be the number of agents in network $g_r$, and $n = \sum_{r=1}^{\bar{r}} n_r$ be the total number of agents in the sample. The econometric model corresponding to the best-response function (3) of agent $i$ in network $g_r$ can be written as

$$y_{i,r} = \phi_1 \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + \phi_2 \sum_{j=1}^{n_r} g_{ij,r}^* y_{j,r} + \mathbf{x}_{i,r}' \boldsymbol{\beta}_1 + \sum_{j=1}^{n_r} g_{ij,r}^* \mathbf{x}_{j,r}' \boldsymbol{\beta}_2 + \eta_r + \epsilon_{i,r}, \tag{8}$$

where $\epsilon_{i,r} \sim i.i.d.(0, \sigma^2)$. Let $\mathbf{G}_r = [g_{ij,r}]$ and $\mathbf{G}_r^* = [g_{ij,r}^*]$ denote, respectively, the adjacency matrix and the row-normalized adjacency matrix for network $g_r$. Let $\mathbf{y}_r = (y_{1,r}, \cdots, y_{n_r,r})'$ denote an $n_r \times 1$ vector of observations on the dependent variable and $\mathbf{X}_r = (\mathbf{x}_{1,r}, \cdots, \mathbf{x}_{n_r,r})'$ denote an $n_r \times k_x$ matrix of observations on $k_x$ exogenous variables. In matrix form, model (8) can be written as

$$\mathbf{y}_r = \phi_1 \mathbf{G}_r \mathbf{y}_r + \phi_2 \mathbf{G}_r^* \mathbf{y}_r + \mathbf{X}_r \boldsymbol{\beta}_1 + \mathbf{G}_r^* \mathbf{X}_r \boldsymbol{\beta}_2 + \eta_r \boldsymbol{\iota}_{n_r} + \boldsymbol{\epsilon}_r.$$

For a sample with $\bar{r}$ networks, stack up the data by defining $\mathbf{y} = (\mathbf{y}_1', \cdots, \mathbf{y}_{\bar{r}}')'$, $\mathbf{X} = (\mathbf{X}_1', \cdots, \mathbf{X}_{\bar{r}}')'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \cdots, \boldsymbol{\epsilon}_{\bar{r}}')'$, $\mathbf{G} = \mathrm{diag}\{\mathbf{G}_r\}_{r=1}^{\bar{r}}$, $\mathbf{G}^* = \mathrm{diag}\{\mathbf{G}_r^*\}_{r=1}^{\bar{r}}$, $\mathbf{L} = \mathrm{diag}\{\boldsymbol{\iota}_{n_r}\}_{r=1}^{\bar{r}}$ and $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_{\bar{r}})'$, where $\mathrm{diag}\{\mathbf{A}_k\}$ denotes a "generalized" block diagonal matrix in which the diagonal blocks are $m_k \times n_k$ matrices $\mathbf{A}_k$'s. For the entire sample, the model is

$$\mathbf{y} = \phi_1 \mathbf{G}\mathbf{y} + \phi_2 \mathbf{G}^* \mathbf{y} + \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{G}^* \mathbf{X} \boldsymbol{\beta}_2 + \mathbf{L}\boldsymbol{\eta} + \boldsymbol{\epsilon}. \tag{9}$$

In model (9), $\phi_1$ represents the *local-aggregate endogenous peer effect*, where an agent's effort may depend on the aggregate effort level of her friends; $\phi_2$ represents the *local-average endogenous peer effect*, where an agent's effort may depend on the average effort level of her friends; and $\boldsymbol{\beta}_2$ represents *contextual effects*, where an agent's effort may depend on the exogenous characteristics of her friends. The vector of network fixed effects given by $\boldsymbol{\eta}$ captures *the correlated effect* where agents in the same network may behave similarly as they have similar unobserved individual characteristics or they face a similar (institutional) environment. The network fixed effect serves as a remedy for the selection bias that originates from the possible sorting of individuals with similar unobserved characteristics into a network. Since the seminal work of Manski (1993), identification of social network models in the presence of endogenous, contextual and correlated effects has attracted a lot of attention in the literature (see Blume et al., 2011 for an excellent review).

The network fixed effect also captures the deterrence effect on delinquency, i.e., the term $p\,f$ in the utility (1). Indeed, because networks are within schools in our data, the use of network fixed effects also accounts for differences in the strictness of anti-delinquency regulations across schools. Thus, instead of directly estimating deterrence effects (i.e. including in the model specification observable measures of deterrence, such as local police expenditures or the arrest rate in the local area), we focus our attention on the estimation of peer effects in delinquency, accounting for the deterrence effect.

In the econometric model, we allow network fixed effects $\boldsymbol{\eta}$ to depend on $\mathbf{G}, \mathbf{G}^*$ and $\mathbf{X}$ by treating $\boldsymbol{\eta}$ as a vector of unknown parameters (as in a fixed effect panel data model). When the number of groups $\bar{r}$ is large, we may have the incidental parameter problem. To avoid the incidental parameter problem, we transform (9) using the deviation from group mean projector $\mathbf{J} = \text{diag}\{\mathbf{J}_r\}_{r=1}^{\bar{r}}$, where $\mathbf{J}_r = \mathbf{I}_{n_r} - \frac{1}{n_r}\boldsymbol{\iota}_{n_r}\boldsymbol{\iota}'_{n_r}$. This transformation is analogous to the *within* transformation for the fixed effect panel data model. As $\mathbf{JL} = 0$, the transformed model is

$$\mathbf{Jy} = \phi_1\mathbf{JGy} + \phi_2\mathbf{JG}^*\mathbf{y} + \mathbf{JX}\boldsymbol{\beta}_1 + \mathbf{JG}^*\mathbf{X}\boldsymbol{\beta}_2 + \mathbf{J}\boldsymbol{\epsilon}. \tag{10}$$

### 5.2 Identification of peer effects

To understand the identification of the general network model (9), we first discuss the identification of two special cases, namely, the local-average model with $\phi_1 = 0$ and the local-aggregate model with $\phi_2 = 0$.

#### 5.2.1 Identification of the local-average network model

Bramoullé et al. (2009) and Lee et al. (2010) have considered the identification of the local-average network model given by

$$\mathbf{y} = \phi_2\mathbf{G}^*\mathbf{y} + \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{G}^*\mathbf{X}\boldsymbol{\beta}_2 + \mathbf{L}\boldsymbol{\eta} + \boldsymbol{\epsilon}. \tag{11}$$

To estimate the transformed local-average model

$$\mathbf{Jy} = \phi_2\mathbf{JG}^*\mathbf{y} + \mathbf{JX}\boldsymbol{\beta}_1 + \mathbf{JG}^*\mathbf{X}\boldsymbol{\beta}_2 + \mathbf{J}\boldsymbol{\epsilon}$$

using IV-based estimators, the conditional mean of the right-hand-side (RHS) variables, $[\text{E}(\mathbf{JG}^*\mathbf{y}|\mathbf{G}, \mathbf{X}), \mathbf{JX}, \mathbf{JG}^*\mathbf{X}]$, needs to have full column rank for large enough sample size.

When $(\mathbf{I} - \phi_2\mathbf{G}^*)$ is nonsingular, from the reduced-form equation of the local-average model, we have

$$\text{E}(\mathbf{JG}^*\mathbf{y}|\mathbf{G}, \mathbf{X}) = \mathbf{JG}^*\mathbf{X}\boldsymbol{\beta}_1 + (\mathbf{JG}^{*2}\mathbf{X} + \phi_2\mathbf{JG}^{*3}\mathbf{X} + \cdots)(\phi_2\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2). \tag{12}$$

If $\phi_2\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 = 0$, then $\text{E}(\mathbf{JG}^*\mathbf{y}|\mathbf{G}, \mathbf{X}) = \mathbf{JG}^*\mathbf{X}\boldsymbol{\beta}_1$. Thus, the model cannot be identified as $[\text{E}(\mathbf{JG}^*\mathbf{y}|\mathbf{G}, \mathbf{X}), \mathbf{JX}, \mathbf{JG}^*\mathbf{X}]$ does not have full column rank. On the other hand, if $\phi_2\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \neq 0$, then $\mathbf{JG}^{*2}\mathbf{X}$ can be used as IVs for the endogenous variable $\mathbf{JG}^*\mathbf{y}$. Observe that, in a social network, if individuals $i, j$ are friends and $j, k$ are friends, it does not necessarily imply that $i, k$ are also friends. Thus, *the intransitivity in social connections* provides an exclusion restriction so that the IV matrix $\mathbf{JG}^{*2}\mathbf{X}$ can be linearly independent of the exogenous regressors $[\mathbf{JX}, \mathbf{JG}^*\mathbf{X}]$.[12] Based on this important observation, Bramoullé et al. (2009) have argued that the local-average model is identified if $\mathbf{I}, \mathbf{G}^*, \mathbf{G}^{*2}, \mathbf{G}^{*3}$ are linearly independent.

[12] In a linear-in-means model, individuals are affected by all members in their group and by no one outside the group. Hence, the perfect collinearity of $\mathbf{JG}^{*2}\mathbf{X}$ and $[\mathbf{JX}, \mathbf{JG}^*\mathbf{X}]$ prevent the identification of the endogenous effect from the contextual effect (i.e. the so-called *reflection problem*, Manski, 1993).

### 5.2.2 Identification of the local-aggregate network model

If $\phi_2 = 0$, then the general network model (9) reduces to the local-average network model given by

$$\mathbf{y} = \phi_1 \mathbf{G}\mathbf{y} + \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{G}^*\mathbf{X}\boldsymbol{\beta}_2 + \mathbf{L}\boldsymbol{\eta} + \boldsymbol{\epsilon}. \tag{13}$$

Similar to the local-average model, IV-based estimation of this model requires the conditional mean of the RHS variables, $[\mathrm{E}(\mathbf{JGy}|\mathbf{G},\mathbf{X}), \mathbf{JX}, \mathbf{JG}^*\mathbf{X}]$, to have full column rank for large enough sample size.

When $(\mathbf{I} - \phi_1\mathbf{G})$ is nonsingular, from the reduced-form equation of (13), we have

$$\mathrm{E}(\mathbf{JGy}|\mathbf{G},\mathbf{X}) = \mathbf{J}(\mathbf{GX}+\phi_1\mathbf{G}^2\mathbf{X}+\cdots)\boldsymbol{\beta}_1+\mathbf{J}(\mathbf{GG}^*\mathbf{X}+\phi_1\mathbf{G}^2\mathbf{G}^*\mathbf{X}+\cdots)\boldsymbol{\beta}_2+\mathbf{J}(\mathbf{GL}+\phi_1\mathbf{G}^2\mathbf{L}+\cdots)\boldsymbol{\eta}. \tag{14}$$

So even if $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = 0$, $\mathrm{E}(\mathbf{JGy}|\mathbf{G},\mathbf{X}) = \mathbf{J}(\mathbf{GL} + \phi_1\mathbf{G}^2\mathbf{L} + \cdots)\boldsymbol{\eta}$ may still be linearly independent of the exogenous regressors $[\mathbf{JX}, \mathbf{JG}^*\mathbf{X}]$. Observe that, the matrix $(\mathbf{GL} + \phi_1\mathbf{G}^2\mathbf{L} + \cdots)$ collects the *Katz-Bonacich centrality* for all individuals, with the non-zero entries of the leading order matrix $\mathbf{GL} = \mathrm{diag}\{\mathbf{G}_r\boldsymbol{\iota}_{n_r}\}_{r=1}^{\bar{r}}$ being the number of friends of every student. Therefore, for the local-aggregate model, in addition to the exogenous characteristics of (indirect) friends, the Katz-Bonacich centrality can also be used as an IV to identify the endogenous peer effect.

Let $\mathbf{Z} = [\mathbf{Gy}, \mathbf{X}, \mathbf{G}^*\mathbf{X}]$. For identification of (13) through linear IV estimators, $\mathrm{E}(\mathbf{JZ}|\mathbf{G},\mathbf{X})$ needs to have full column rank for large enough sample size. Suppose $\mathbf{X}$ is a random column vector,[13] then the following proposition gives sufficient conditions for the rank condition. Henceforth, let $c$ (possibly with subscripts) denote a constant scalar that may take different values for different uses.

**Proposition 2** *Suppose* $\mathbf{X}$ *is a random column vector. When* $\mathbf{G}_r$ *has non-constant row sums for some network* $r$*,* $\mathrm{E}(\mathbf{JZ}|\mathbf{G},\mathbf{X})$ *has full column rank if:* (i) $\mathbf{I}_{n_r}, \mathbf{G}_r, \mathbf{G}_r^*, \mathbf{G}_r\mathbf{G}_r^*$ *are linearly independent and* $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \eta_r$ *are not all zeros; or* (ii) $\mathbf{G}_r\mathbf{G}_r^* = c_1\mathbf{I}_{n_r} + c_2\mathbf{G}_r + c_3\mathbf{G}_r^*$ *and* $\boldsymbol{\Lambda}_1$ *given by (16) has full rank.*
*When* $\mathbf{G}_r$ *has constant row sums such that* $\bar{g}_{i,r} = \bar{g}_r$ *for all* $r$*,* $\mathrm{E}(\mathbf{JZ}|\mathbf{G},\mathbf{X})$ *has full column rank if:* (iii) $\mathbf{I}, \mathbf{G}, \mathbf{G}^*, \mathbf{GG}^*, \mathbf{G}^{*2}, \mathbf{GG}^{*2}$ *are linearly independent and* $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ *are not both zeros;* (iv) $\mathbf{I}, \mathbf{G}, \mathbf{G}^*, \mathbf{GG}^*, \mathbf{G}^{*2}$ *are linearly independent,* $\mathbf{GG}^{*2} = c_1\mathbf{I} + c_2\mathbf{G} + c_3\mathbf{G}^* + c_4\mathbf{GG}^* + c_5\mathbf{G}^{*2}$*, and* $\boldsymbol{\Lambda}_2$ *given by (17) has full rank; or* (v) $\bar{g}_r = \bar{g}$ *for all* $r$*,* $\mathbf{I}, \mathbf{G}^*, \mathbf{G}^{*2}, \mathbf{G}^{*3}$ *are linearly independent, and* $\bar{g}\phi_1\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \neq 0$.

The proof of Proposition 2 is given in Appendix A. From Proposition 2, we can see that, for a given network, sometimes identification of the local-aggregate model is easier to achieve than the local-average model. Figure 2 gives some examples where identification is possible for the local-aggregate model but fails for the local-average model. First, consider a sample where each network is represented by graph (a) (a star-shaped network). The corresponding adjacency matrix $\mathbf{G}$ is a block-diagonal matrix with diagonal blocks $\mathbf{G}_r$ representing graph (a). For the row-normalized adjacency matrix $\mathbf{G}^*$, it is easy to see that $\mathbf{G}^{*3} = \mathbf{G}^*$. Therefore, it follows from Proposition 5 of Bramoullé et al. (2009) that the local-average model (11) is not identified. On the other hand, as

[13] This assumption is also assumed in Bramoullé et al. (2009).

$\mathbf{G}_r$ corresponding to the graph (a) has non-constant row sums and $\mathbf{I}_{n_r}, \mathbf{G}_r, \mathbf{G}_r^*, \mathbf{G}_r\mathbf{G}_r^*$ are linearly independent, it follows from our Proposition 2(i) that the local-aggregate model can be identified for this network.

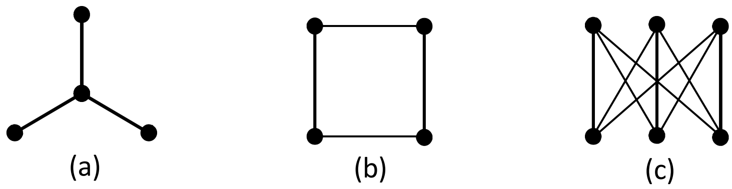



Figure 2: example networks for the identification of the local-aggregate model

Graphs (b) and (c) in Figure 2 provide another example where the local-average model cannot be identified while the local-aggregate model can be. Consider a sample that consists of two types of networks. The first $\bar{r}_1$ networks are represented by graph (b) (a regular or circle network). The rest $\bar{r}_2$ networks are represented by graph (c) (a bi-partite network). Suppose $\bar{r}_1 > 0$, $\bar{r}_2 > 0$ and $\bar{r}_1 + \bar{r}_2 = \bar{r}$. The corresponding adjacency matrix $\mathbf{G}$ is a block-diagonal matrix with the first $\bar{r}_1$ diagonal blocks $\mathbf{G}_{1r}$ representing graph (b) and the rest $\bar{r}_2$ diagonal blocks $\mathbf{G}_{2r}$ representing graph (c). For the row-normalized adjacency matrix $\mathbf{G}^*$, it is easy to see that $\mathbf{G}^{*3} = \mathbf{G}^*$. Therefore, it follows from Proposition 5 of Bramoullé et al. (2009) that the local-average model (11) is not identified. On the other hand, as $\mathbf{G}_{1r}$ and $\mathbf{G}_{2r}$ have different row sums, $\mathbf{I}, \mathbf{G}, \mathbf{G}^*, \mathbf{G}\mathbf{G}^*, \mathbf{G}^{*2}$ are linearly independent and $\mathbf{G}\mathbf{G}^{*2} = \mathbf{G}$. Therefore, the local-aggregate model can be identified by our Proposition 2(iv).

#### 5.2.3 Identification of the general network model

The identification of the general network model with both local-aggregate and local-average endogenous peer effects has been discussed in Liu et al. (2014). Intuitively, the identification of two types of endogenous peer effects relies on the row sum variation of $\mathbf{G}$. If $\mathbf{G}$ has constant row sums such that $\mathbf{G} = \bar{g}\mathbf{G}^*$ for some constant $\bar{g}$, then local-aggregate and local-average peer effects cannot be separately identified. If $\mathbf{G}$ does not have constant row sums so that local-aggregate and local-average peer effects can be separately identified, then, similar to the local-aggregate model, exogenous characteristics of indirect friends and the Katz-Bonacich centrality can be used as IVs to identify the local-aggregate and local-average peer effects from contextual effects.

### 5.3 Estimation methods and results

For the estimation of the (within-transformed) general network model (10) with both local-aggregate and local-average peer effects, we generalize the two-stage least squares (2SLS) and generalized method of moments (GMM) estimators developed in Liu and Lee (2010). Specifically, we consider the following estimators:

(a) "2SLS": a 2SLS estimator with the IV matrix $\mathbf{Q}_1 = \mathbf{J}[\mathbf{X}, \mathbf{G}^*\mathbf{X}, \mathbf{G}\mathbf{X}, \mathbf{G}^{*2}\mathbf{X}]$.

(b) "BC2SLS": a bias-corrected 2SLS estimator with the IV matrix $\mathbf{Q}_2 = [\mathbf{Q}_1, \mathbf{J}\mathbf{G}\mathbf{L}]$. The IVs $\mathbf{J}\mathbf{G}\mathbf{L}$ in $\mathbf{Q}_2$ correspond to the leading order term of the (within-transformed) *Katz-Bonacich* *centrality*. The additional IVs in $\mathbf{Q}_2$ improve asymptotic efficiency of the 2SLS estimator. Note that, the matrix $\mathbf{JGL}$ has $\bar{r}$ columns, where $\bar{r}$ is the number of networks in the data. Therefore, if there are many networks in the data, the 2SLS estimator with the IV matrix $\mathbf{Q}_2$ may have an asymptotic bias, which is known as the many-instrument bias. The "BC2SLS" estimator corrects the many-instrument bias by subtracting the estimated leading-order bias from the 2SLS estimator.

The 2SLS estimators are based on moment conditions that are *linear* in the estimation residuals. Lee (2007b) has suggested to generalize the 2SLS method to a GMM framework with additional *quadratic* moment conditions based on the covariance structure of the reduced form equation to improve estimation efficiency. The added quadratic moment conditions are especially helpful when the IVs are weak. We consider the following GMM estimators for the estimation of the empirical model:

(c) "GMM": an optimal GMM estimator using the IV matrix $\mathbf{Q}_1$ and additional quadratic moment conditions.

(d) "BCGMM": a bias-corrected optimal GMM estimator using the IV matrix $\mathbf{Q}_2$ and the same quadratic moment conditions as in "GMM". Similar to the corresponding 2SLS estimator with $\mathbf{Q}_2$, the IVs $\mathbf{JGL}$ may introduce a many-instrument bias into the GMM estimator. The "BCGMM" estimator corrects the many-instrument bias by subtracting the estimated leading-order bias from the GMM estimator.

The details of the 2SLS and GMM methods, including the explicit form of the quadratic moment conditions, are given in Appendix B.

Table 2 collects the estimation results of the general network model (10) where both the local-average and the local-aggregate effects are incorporated. The first stage partial F-statistics (see Stock et al., 2002 and Stock and Yogo, 2005) reveals that our instruments are not very informative. Hence, the GMM estimates are more reliable. Both the 2SLS and GMM estimates of $\phi_1, \phi_2$ are positive and statistically significant, and satisfy the condition $\bar{g}^{\max}\phi_1 + \phi_2 < 1$, which is needed to guarantee the existence of the Nash equilibrium of the network game presented in Section 3.1.

[*Insert Table* 2 *here*]

As described in Section 4, the adjacency matrix $\mathbf{G}$ in the general network model (10) in based on reciprocal nominations such that $g_{ij} = 1$ if $i$ nominates $j$ as a friend or $j$ nominates $i$ as a friend. As a robustness check of the possible misspecification of the adjacency matrix, we also consider an adjacency matrix $\mathbf{G}$ based on actual nominations such that $g_{ij} = 1$ if $i$ nominates $j$ as a friend. The estimation results with $\mathbf{G}$ based on actual nominations are reported in Table 3. The estimation results are similar to those in Table 2, suggesting the estimation is robust to possible misspecification of the adjacency matrix.

[*Insert Table* 3 *here*]

Table 4 reports the quasi-maximum-likelihood (QML) estimates of the local-average network model (11) (see Lee et al. 2010 for the QML estimator) and the BCGMM estimates of the local-aggregate model (13) with $\mathbf{G}$ based on reciprocal nominations. Each model is estimated separately. Compared with the BCGMM estimates reported in Table 2, both the estimate of $\phi_1$ in the local-aggregate model and the estimate of $\phi_2$ in the local-average model are upwards biased. This shows the importance of estimating the two models together.

[*Insert Table* 4 *here*]

### 5.4 Tests for the exogeneity of the adjacency matrix

The validity of our identification strategy and the moment conditions employed by the 2SLS and the GMM estimators rests on the (conditional) exogeneity of the adjacency matrix $\mathbf{G}$. Here, we provide two tests for the potential endogeneity of $\mathbf{G}$.

The first test is based on the over-identifying restrictions (OIR) test (Lee, 1992). Since the IVs used in the 2SLS are constructed based on $\mathbf{G}$, if the OIR test cannot reject the null hypothesis that the IVs are exogenous, then it provides evidence that $\mathbf{G}$ is uncorrelated with the error term, when $\mathbf{X}$, the contextual effects and $\boldsymbol{\eta}$, the network fixed effects are controlled for. As reported at the bottom of Tables 2 and 3, the $p$-value of the OIR test is larger than conventional significance levels, which provides evidence of the *exogeneity* of $\mathbf{G}$.

The second test is inspired by the network formation model proposed by Goldsmith-Pinkham and Imbens (2013). Assuming homophily, students with similar observed and/or unobserved characteristics are more likely to be friends. Suppose students $i$ and $j$ ($i < j$) in network $g_r$ are friends, i.e., $g_{ij,r} = 1$ if $d^*_{ij,r} > 0$, where

$$d^*_{ij,r} = \delta_0 + \delta_1 |\hat{\epsilon}_{i,r} - \hat{\epsilon}_{j,r}|^{-1} + \boldsymbol{\Delta}_{ij,r}\boldsymbol{\gamma} + u_{ij,r}. \tag{15}$$

In (15), $\hat{\epsilon}_{i,r}$ is the BCGMM estimation residual of the general network model (9), $\boldsymbol{\Delta}_{ij,r}$ is a $k_x \times 1$ vector with the $p$-th entry of $\boldsymbol{\Delta}_{ij,r}$ being 1 if the $p$-th entries $\mathbf{x}_{i,r}$ and $\mathbf{x}_{j,r}$ are the same and being 0 otherwise,[14] $u_{ij,r}$ is the error term that is assumed i.i.d. for all $(i, j)$ such that $i < j$. Note that $\boldsymbol{\Delta}_{ij,r}$ and $|\hat{\epsilon}_{i,r} - \hat{\epsilon}_{j,r}|^{-1}$ capture how similar students $i$ and $j$ are in their observed and unobserved characteristics.[15] Under the null hypothesis that $\mathbf{G}$ is exogenous, $\delta_1$ should be zero, which means the friendship formation does not depend on students' unobserved characteristics that affect delinquency activities.

The logit regression of model (15) is reported in Table 5. Students are more likely to be friends if they are of the same gender and race, if they were born in the same country, if they are in the same grade, if they have similar math scores, if they have similar family structure, and if their parents have the same type of jobs. On the other hand, the estimated $\delta_1$ is statistically insignificant, showing evidence of the exogeneity of $\mathbf{G}$.

[*Insert Table* 5 *here*]

[14] Note that all entries of $\mathbf{x}_{i,r}$ in the empirical application are discrete variables (most are binary indicator variables).

[15] In the spatial econometrics literature, the proximity given by the inversed distance between spatial units is often used to construct the spatial weights matrix.

## 6 Who is the key player?

With the BCGMM estimates of $\boldsymbol{\theta}$ reported in Table 2, the key player can be determined for each network according to (6).[16] First, we find that the key player is not necessarily the most *active* student in delinquent activities. Only in 19 networks out of the 103 networks in the sample, the key player is the most active student in delinquent activities. Second, the key player is not necessarily the student with the highest *Katz-Bonacich centrality* given by $(\mathbf{I}_{n_r} - \phi_1 \mathbf{G}_r)^{-1}\boldsymbol{\iota}_{n_r}$. In 63 networks out of the 103 networks in the sample, the key player is the student with the highest Katz-Bonacich centrality.

Once we identify the key player for each network, we can draw her "profile" by comparing the characteristics of the key players with those of the other students in the network. Table 6 lists the characteristics of the key player that is significantly different from those of the other students.

[*Insert Table* 6 *here*]

Compared to other students, the key players are older and in higher grades, are less likely to be white, have lower math scores, have lower self esteem, and are less likely to feel being part of school or being safe in school. They are also less likely to feel that their parents care about them, and are less likely to come from families where parents work as professionals. Also, key players are more likely to have friends, who are older, non-white, born outside of the U.S., less intelligent (lower math scores), and does not feel part of school or feel safe in school.

We also compare the aggregate-delinquency reductions by removing, respectively, the *key player*, the *most active delinquent*, and a *random delinquent.* In the 103 networks of our sample, on average, the aggregate-delinquency reductions by removing the key player, the most active delinquent, and a random delinquent are, respectively, 3.34, 2.14, and 1.96.[17] A paired two-sample $t$ test suggests that the aggregate-delinquency reduction by removing the key player is statistically significantly larger than that by removing the most active delinquent or a random delinquent. The median network size in our sample is 4. For a network with 4 students, the percentage reductions in aggregate delinquency by removing the key player, the most active delinquent, and a random delinquent are, respectively, 70.19%, 53.13%, and 50.71%. Hence, targeting key players is more effective than targeting the most active delinquent in reducing total delinquency.

## 7 Policy implications

Let us address the fundamental policy issues of the key player. First, we provide some evidence that there exist data on criminal networks and discuss the extent to which our analysis can be used in practice to address general policies against crime. Second, we show that the methodology developed in this paper can be used to address policies in other activities such as financial networks, R&D networks, social networks in developing economies and political networks.

[16] The key players remain largely unchanged when (6) is evaluated with other estimates reported in Table 2. For example, with the GMM estimates of $\boldsymbol{\theta}$, only in one network out of the 103 networks in the sample, the identity of the key player changes.

[17] Note that, on average, the delinquency index of a student in our sample is 1.03.

### 7.1 How the key-player policy can be useful in fighting crime

#### 7.1.1 Data about criminal networks

In order to apply the key-player policy one needs to collect detailed data on the social networks linking individuals and on their criminal activity. Let us now give a (selective) list of available network data:

(i) *Juvenile delinquency in schools (survey data):* There are data similar to the Add Health data for students in schools. For example, Weerman (2011) uses data from the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) "School Study", a study that focused on social networks and the role of peers in delinquency with two waves, conducted in the spring of 2002 and 2003. Students were provided with a numbered list of all students in the same grade in their school and were asked to fill in the numbers of those fellow students they spend time with at school ("with which of these students do you associate regularly?"), with a maximum of ten possible nominations. Students' delinquent behavior was measured using self-reports on a variety of offenses. The final sample consisted of 1,156 students in ten schools that participated in both waves.

(ii) *Adult crime (police data)*: the police has in fact quite a lot of information on criminal networks. For example, Lindquist and Zenou (2014) were able to construct the network of all criminals in Sweden for several years. The way a link is defined is as follows. Each time two (or more) persons are suspected of a crime (*co-offenders*), the police in Sweden registers this information. A link in a network is then created between individuals $i$ and $j$, i.e. $g_{ij} = g_{ji} = 1$, whenever individuals $i$ and $j$ are suspected of a crime together.[18] With this information at hand, we can match each individual's social security number with her characteristics (education, age, ethnicity, gender, etc.). In that case, Lindquist and Zenou (2014) were able to determine the key criminals in Sweden over a long period of time.[19]

This type of information can be obtained from the police in many other countries. Indeed, it is important to identify criminal networks in data resources readily available to investigators, such as police arrest data and court data. See, for example, Tayebi et al. (2011) who define a *co-offending network*, that is a network of offenders who have committed crimes together. In the United States, there are also similar data. For example, Coplink (Hauck et al., 2002) was one of the first large scale research projects in crime data mining, and an excellent work in criminal network analysis. It is remarkable in its practicality, being integrated with and used in the workflow of the Tucson Police Department. Coplink has information about the perpetrators' habits and close associations in crime to capture the *connections* between people, places, events, and vehicles, based on past crimes. Xu and Chen (2005) built on this when they created CrimeNet Explorer, a framework for criminal network knowledge discovery incorporating hierarchical clustering, Social Network Analysis (SNA) methods, and multidimensional scaling.

(iii) *Gang networks*: McGloin (2004, 2005) use data from the Newark portion of the North Jersey Gang Task Force, a regional problem analysis project that sought to define the local gang landscape

[18] The authors also check if the two suspected criminals have been eventually prosecuted and condemned to a prison sentence.

[19] The authors perform several exercises to test for measurement errors in links and show that even by adding new links the estimations are not affected.

in Northern New Jersey. These data came from the experiential knowledge of representatives of various criminal justice agencies, including the Newark Police Department, Essex County Sheriff's Office, Essex County Department of Parole, and Juvenile Justice Commission of New Jersey. In particular, groups of law enforcement officials from this variety of criminal justice agencies engaged in collective semi-structured interviews — 32 over the course of one year — that solicited information on the gang landscape. In particular, they provided information on known gang members, as well as the quantity and type of their respective associates. The classification of the profession of gang members in the questionnaire relied on New Jersey code, which defines a gang as three or more people who are, in fact, associated, that is people who have a common group name; identifying sign, tattoos, or other indices of association; and who have committed criminal offenses while engaged in gang-related activity.[20]

#### 7.1.2 How to implement a key-player policy

Once we have criminal network data, there are different ways of implementing a key-player policy. Indeed, once we have identified a key player in a network, one cannot put him/her in prison if he/she hasn't committed any crime. However, different policies can be implemented to reduce crime using a key-player approach.

(i) First, the police can offer to the key player(s) incentives to leave the gang or the criminal network. For example, the police can offer them a job or a conditional transfer (by asking them to move to another city) or monitor them more. These types of policies have been implemented in Canada where some gang members of criminal networks were persuaded to abandon gang life in return for needed employment training, educational training, and skills training (Tremblay et al., 1996).

(ii) Second, the police can target key players in a meaningful way. A very similar type of policy has actually been implemented in the US. Indeed, a recent innovation in policing that capitalizes on the growing evidence of the effectiveness of police deterrence strategies is the "focused deterrence" framework, which is often referred to as "pulling-levers policing" (Kennedy, 1998, 2008). This strategy was pioneered in Boston (*Boston Gun Project*) with the intent of understanding the purported nexus of rising youth violence and use of firearms. As part of its problem analysis, representatives of various criminal justice agencies defined and characterized problematic local gangs (Braga et al., 2001; Kennedy et al., 1996, 1997, 2001). This process included elaborating on the relationships among the street gangs, which Kennedy et al. (1996, 1997, 2001) translated into sociograms illustrating connections within the gang landscape. This seemingly simple information was invaluable for the problem analysis and construction of *Operation Ceasefire.* This latter policy combines a strong law enforcement response with a "pulling levers" deterrence effort aimed at chronic gang offenders. The key to the success is to use a "lever pulling" approach, which is a crime deterrence strategy that attempts to prevent violent behavior by using a *targeted individual or group's vulnerability* to law enforcement as a means of gaining their compliance. Operation Ceasefire was first launched in Boston and youth homicide fell by two-thirds after the Ceasefire strategy was put in place in

[20] See also Mastrobuoni and Patacchini (2012) who use a data set provided by the Federal Bureau of Narcotics on criminal profiles of 800 US Mafia members active in the 1950s and 1960s and on their connections within the Cosa Nostra network.

1996 (Kennedy, 1998). It was then implemented in Los Angeles in 2000: police beefed up patrols in the area, attempting to locate gang members who had *outstanding arrest warrants* or *had violated probation or parole regulations*. Gang members who had *violated public housing rules*, *failed to pay child support*, or were similarly vulnerable were also subjected to stringent enforcement (Tita et al., 2003).

(iii) Finally, a key-player policy can also help for related issues. For example, there is a lot of debate in the US on how to allocate under-age adolescents who have committed an offence into juvenile facilities (detention centers). We know that there is a lot of learning in crime in prisons (Bayer et al., 2009). If we can rank these adolescents by their key-player centralities, then our model predict that we should put together the delinquents with high key-player centralities while grouping together young delinquents with low key-player centralities.

### 7.2 Using our methodology to find key players for other types of networks and activities

Our paper can also be seen as a methodological one. Our methodology can be applied to other contexts where network data are much easier to obtain. Let us provide some examples.

**Financial networks**

There is an abundance of information available on financial networks where links are usually bank loans. For example, Cohen-Cole et al. (2011) use transaction level data on interbank lending from an electronic interbank market, the e-MID SPA (or e-MID), which was the reference marketplace for liquidity trading in the Euro area from January 2002 to December 2009. Boss et al. (2004) analyze the network of Austrian banks in the year 2008 where links in the network represent exposures between Austrian-domiciled banks on a non-consolidated basis (i.e. no exposures to foreign subsidiaries are included). In that case, the key player policy would be: Which bank should we bail out in order to reduce systemic risk or maximize total activity? This is an extremely important issue because the recent financial crisis has shown that the problem is not necessary "too big to fail" but "too interconnected to fail". Interestingly, using a sterling interbank network database from January 2006 to September 2010, Denbee et al. (2014) use our methodology to determine the key banks. They show that, during the years before and after the 2007-08 crisis, the key players vary and are not necessarily the largest borrowers.

**R&D networks**

There is also a lot of information on R&D networks. For example, García-Canal et al. (2008) use alliance data stemming from the Thomson SDC Platinum data base. Three types of alliances are reported in the SDC database: (1) alliances that imply the transmission of an existing technology from one partner to another or to the alliance; (2) alliances that imply the cross-transfer of existing technologies between two or more partners or between these and the alliance, and (3) alliances that include the undertaking of R&D activities.[21]

Using data on interfirm R&D collaborations stemming from the MERIT-CATI database, König et al. (2014a) apply our methodology to determine the key firms in R&D networks. They show that

[21] Another dataset that has been used is the one on interfirm collaborations from the NBER Patent Data File (Hall et al., 2001).

the key firms vary over time, even though the ranking is relatively stable. They show that General Motors, which was bailed out by president Obama in 2008, ranked first in 1990 and ninth in 2005. According to König et al. (2014a), if General Motors had gone bankrupt in 1990 and exited the market, the loss of total welfare for both firms and consumers would have been as high as 8.37%.

**Networks in developing economies**

There are many network data for developing countries (see e.g. Fafchamps and Lund, 2003, who conducted a survey in four villages in the Cordillera mountains of northern Philippines between July 1994 and March 1995; Krishnan and Sciubba, 2009, who use the second round of the Ethiopian Rural Household Survey, conducted in 1994). There is also a recent paper by Banerjee et al. (2013) which study a problem related to the key-player issue. Their data come from a survey on 75 rural villages in Karnataka, India, that the authors conducted to obtain information on network structure and various demographics. They look at the diffusion of a microfinance program in these villages and show that, if the bank in charge of this program had targeted individuals in the village with the highest eigenvector centrality (a measure related to the Katz-Bonacich centrality), the diffusion of the microfinance program (i.e. take-up rates) would have been much higher. More generally, in developing countries, one could apply the key player policy to the issue of adoption of a new technology since there is strong evidence of social learning (Conley and Udry, 2010) and take-up rates in microfinance programs.

**Political networks**

Another application of a key player policy could be the political world. There is evidence that personal connections amongst politicians have a significant impact on the voting behavior of U.S. politicians (Cohen and Malloy, 2014). There is also evidence on lobbying to persuade public opinion when members of the public influence each other's opinion (Lever, 2010). When people are deciding how to vote or which product to buy, they discuss their decision with people in their social environment. Competitions to persuade public opinion are the essence of political campaigns, but also occur in marketing between rival firms or in lobbying by interests groups on opposite sides of a legislation. Matching data on campaign contributions by lobby groups with data on co-sponsorship networks in the US House of Representatives, Lever (2010) finds that changes in both network in influence and pivot probabilities are significant predictors of changes in campaign contributions. Our key-player policy suggests that resources should be spent on *key voters* who have an influential position in the social network.

## 8 Concluding remarks

This paper presents a methodology for determining the key player whose exclusion from her network would result in the greatest impact on the outcome of interest (adolescent crime rates in the case of our Add Health data). We provide a structural estimation of the model, and a simulation describing how such a key player is identified, and furthermore, that this "greatest impact" key player cannot be accurately identified by the other measures prominent in the literature. This methodology provides a cost effective instrument for a wide variety of policy interventions. Implementation of a key-player based policy intervention in networks would result in drastic cost reductions by taking advantage of multiplier effects.

The key-player methodology has great scope for practical implementation, since it takes advantage of multiplier effects in naturally occurring networks. Given the right data (which in several instances already exist), it can be implemented in many contexts.[22] Consider the recent Obama-Romney US presidential election, which recorded the most extensive campaign spending in US history. Implementation of our key-player methodology would shift attention away from "swing states" and rather target "swing voters", who would have the greatest possible impact on voter decision within their social networks. Such an approach could easily have tempered the USD 2.3 billion total campaign cost.[23] The cost of gathering the necessary data could not be expected to be more than a fraction of this. Alternatively consider vaccination in developing countries, targeting individuals who would reduce the spread of infectious diseases the most would be highly cost effective, freeing up resources to make more vaccination projects viable, and reducing the overall infection rate. Even where data is not yet freely available, implementing the key-player policy could be cost-effective overall, if the cost of data gathering is recouped by the cost saving in the actual policy intervention that stems from targeting key-players rather than all individuals.

## References


[1] Akers, R.L. (1998), *Social Learning and Social Structure: A General Theory of Crime and Deviance*, Boston: Northeastern University Press.

[2] Ballester, C., Calvó-Armengol, A. and Y. Zenou (2006), "Who's who in networks. Wanted: the key player," *Econometrica* 74, 1403-1417.

[3] Ballester, C., Calvó-Armengol, A. and Y. Zenou (2010), "Delinquent networks," *Journal of the European Economic Association* 8, 34-61.

[4] Ballester, C. and Y. Zenou (2014), "Key player policies when contextual effects matter," *Journal of Mathematical Sociology* 38, 233-248.

[5] Banerjee, A., Chandrasekhar, A.G., Duflo, E. and M.O. Jackson (2013), "The diffusion of microfinance," *Science* 26, Vol. 341, no. 6144.

[6] Bayer, P., Hjalmarsson, R. and D. Pozen (2009), "Building criminal capital behind bars: Peer effects in juvenile corrections," *Quarterly Journal of Economics* 124, 105-147.

[7] Blume, L.E., Brock, W.A., Durlauf, S.N. and Y.M. Ioannides (2011), "Identification of social interactions," In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics*, Amsterdam: Elsevier Science, pp. 853-964.

[8] Bonacich, P. (1987), "Power and centrality: A family of measures," *American Journal of Sociology* 92, 1170-1182.


[22] For example, using data for the Second Congo (DRC) War, a conflict involving many groups and a complex network of alliances and rivalries, König et al. (2014b) apply our methodology to infer the extent to which the removal of each group involved in the conflict would reduce the conflict intensity. By collecting data from the UEFA Euro 2008 Tournament, Sarangi and Unlu (2014) use our methodology to evaluate a player's contribution to her team and relates her effort to her salary.

[23] See http://www.opensecrets.org/pres12/index.php#out.

[9] Borgatti, S.P. (2003), "The key player problem," In: R. Breiger, K. Carley and P. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, New York: National Academy of Sciences Press, pp. 241-252.

[10] Borgatti, S.P. (2006), "Identifying sets of key players in a network," *Computational, Mathematical and Organizational Theory* 12, 21-34.

[11] Boss, M., Elsinger, H., Summer, M. and S. Thurner (2004), "Network topology of the interbank market," *Quantitative Finance* 4, 677-684.

[12] Braga, A.A., Kennedy, D.M., Waring, E.J. and A.M. Piehl (2001), "Problem oriented policing, deterrence, and youth violence: An evaluation of Boston's Operation Ceasefire," *Journal of Research in Crime and Delinquency* 38, 195-225.

[13] Bramoullé, Y., Djebbari, H. and B. Fortin (2009), "Identification of peer effects through social networks," *Journal of Econometrics* 150, 41-55.

[14] Calvó-Armengol, A., Patacchini, E. and Y. Zenou (2009), "Peer effects and social networks in education," *Review of Economic Studies* 76, 1239-1267.

[15] Calvó-Armengol, A. and Y. Zenou (2004), "Social networks and crime decisions: The role of social structure in facilitating delinquent behavior," *International Economic Review* 45, 935-954.

[16] Cohen-Cole, E., Patacchini, E. and Y. Zenou (2011), "Systemic risk and network formation in the interbank market," CEPR Discussion Paper No. 8332.

[17] Cohen, L. and C. Malloy (2014), "Friends in high places," *American Economic Journal: Economic Policy*, forthcoming.

[18] Conley, T.J. and C.R. Udry (2010), "Learning about a new technology: Pineapple in Ghana," *American Economic Review* 100, 35-69.

[19] Cook, P.J. and J. Ludwig (2010), "Economical crime control," NBER Working Paper No. 16513.

[20] Denbee, E., Julliard, C., Li, Y. and K. Yuan (2014), "Network risk and key players: A structural analysis of interbank liquidity," Unpublished manuscript, London School of Economics,

[21] Durlauf, S.N. and Y.M. Ioannides (2010), "Social interactions," *Annual Review of Economics* 2, 451-478.

[22] Eck, J.E., Chainey, S., Cameron, J.G., Leitner, M. and R.E.Wilson (2005), *Mapping Crime: Understanding Hot Spots*, Washington, DC: Office of Justice Programs, U.S. Department of Justice.

[23] Fafchamps, M. and Lund, S. (2003), "Risk sharing networks in rural Philippines," *Journal of Development Economics* 71, 261-287.

[24] García-Canal, E., Valdés-Llaneza, A. and P. Sánchez-Lorda (2008), "Technological flows and choice of joint ventures in technology alliances," *Research Policy* 37, 97-114.

[25] Glaeser, E.L., Sacerdote, B. and J. Scheinkman (1996), "Crime and social interactions," *Quarterly Journal of Economics* 111, 508-548.

[26] Goldsmith-Pinkham, P. and G.W. Imbens (2013), "Social networks and the identification of peer effects," *Journal of Business and Economic Statistics* 31, 253-264.

[27] Goyal, S. (2007), *Connections: An Introduction to the Economics of Networks*, Princeton: Princeton University Press.

[28] Hall, B.H., Jaffe, A.B. and M. Trajtenberg (2001), "The NBER patent citations data file: Lessons, insights and methodological tools," NBER Working Paper No. 8498.

[29] Hansen, C., Hausman, J. and W. Newey (2008), "Estimation with many instrumental variables," *Journal of Business and Economic Statistics* 26, 398-422.

[30] Hauck, R.V., Atabakhsh, H., Ongvasith, P., Gupta, H. and H. Chen (2002), "Using Coplink to analyze criminal-justice data," *IEEE Computer* 35, 3: 3037.

[31] Haynie, D.L. (2001), "Delinquent peers revisited: Does network structure matter?" *American Journal of Sociology* 106, 1013-1057.

[32] Ioannides, Y.M. (2012), *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton: Princeton University Press.

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

[34] Jackson, M.O. and Y. Zenou (2014), "Games on networks," In: P. Young and S. Zamir (Eds.), *Handbook of Game Theory, Vol. 4*, Amsterdam: Elsevier, pp. 91-157.

[35] Katz, L. (1953), "A new status index derived from sociometric analysis," *Psychometrika* 18, 39-43.

[36] Kennedy, D.M. (1998), "Pulling levers: Getting deterrence right," *National Institute of Justice Journal* 236, 2-8.

[37] Kennedy, D.M. (2008), *Deterrence and Crime Prevention: Reconsidering the Prospect of Sanction*, New York: Routledge.

[38] Kennedy, D.M., Braga, A.A. and A.M. Piehl (1996), *Youth Gun Violence in Boston: Gun Markets, Serious Youth Offenders, and a Use Reduction Strategy*, Boston, MA: John F. Kennedy School of Government, Harvard University Press.

[39] Kennedy, D.M., Braga, A.A. and A.M. Piehl (1997), "The (un)known universe: Mapping gangs and gang violence in Boston," In: D. Weisburd and T. McEwen (Eds.), *Crime Mapping and Crime Prevention*, Monsey, N.Y: Criminal Justice Press.

[40] Kennedy, D.M., Braga, A.A., Piehl, A.M. and E.J. Waring (2001), *Reducing Gun Violence: The Boston Gun Project's Operation Ceasefire*, Washington, D.C: U.S. National Institute of Justice.

[41] Kleiman, M.A. (2009), *When Brute Force Fails. How to Have Less Crime and Less Punishment*, Princeton: Princeton University Press.

[42] Kling, J.R., Ludwig, J. and L.F. Katz (2005), "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment," *Quarterly Journal of Economics* 120, 87-130.

[43] König, M., Liu, X. and Y. Zenou (2014a), "R&D networks: Theory, empirics and policy implications", CEPR Discussion Paper No. 9872.

[44] König, M., Rohner, D., Thoenig, M. and F. Zilibotti (2014b), "Networks in conflict: Theory and evidence from the great war of Africa," Unpublished manuscript, University of Zurich.

[45] Krishnan, P. and E. Sciubba (2009), "Links and architecture in village networks," *Economic Journal* 119, 917-949.

[46] Lee, L.F. (1992), "Amemiya's generalized least squares and tests of overidentification in simultaneous equation models with qualitative or limited dependent variables," *Econometric Reviews* 11, 319-328.

[47] Lee, L.F. (2007a), "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics* 140, 333-374.

[48] Lee, L.F. (2007b), "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models," *Journal of Econometrics* 137, 489-514.

[49] Lee, L.F., Liu, X. and X. Lin (2010), "Specification and estimation of social interaction models with network structures," *Econometrics Journal* 13, 145-176.

[50] Lever, C. (2010), "Strategic spending in voting competitions with social networks," Banco de México Working Paper No. 2010-16.

[51] Lindquist, M. and Y. Zenou (2014), "Key players in co-offending networks," CEPR Discussion Paper No. 9889.

[52] Liu, X. and L.F. Lee (2010), "GMM estimation of social interaction models with centrality," *Journal of Econometrics* 159, 99-115.

[53] Liu, X., Patacchini, E. and Y. Zenou (2014), "Endogenous peer effects: Local aggregate or local average?" *Journal of Economic Behavior and Organization* 103, 39-59.

[54] Ludwig, J., Duncan, G.J. and Hirschfield, P. (2001), "Urban poverty and juvenile crime: Evidence from a randomized housing-mobility experiment," *Quarterly Journal of Economics* 116, 655-679.

[55] Manski, C.F. (1993), "Identification of endogenous effects: The reflection problem," *Review of Economic Studies* 60, 531-542.

[56] McGloin, J.M. (2004), *Associations among Criminal Gang Members as a Defining Factor of Organization and as a Predictor of Criminal Behavior: The Gang Landscape of Newark, New Jersey*, Ann Arbor, MI: University of Michigan Press.

[57] McGloin, J.M. (2005), "Policy and intervention considerations of a network analysis of street gangs," *Criminology and Public Policy* 4, 607-636.

[58] Mastrobuoni, G. and E. Patacchini (2012), "Organized crime networks: An application of network analysis techniques to the American mafia," *Review of Network Economics* 11(3), Article 10.

[59] Patacchini, E. and Y. Zenou (2012), "Juvenile delinquency and conformism," *Journal of Law, Economic, and Organization* 28, 1-31.

[60] Sarangi, S. and E. Unlu (2014), "Key players and key groups in teams: A network approach using soccer data," Unpublished manuscript, Louisiana State University.

[61] Sarnecki, J. (2001), *Delinquent Networks: Youth Co-Offending in Stockholm*, Cambridge: Cambridge University Press.

[62] Stock, J.H, Wright J.H. and M. Yogo (2002), "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business and Economic Statistics* 2, 518-529.

[63] Stock, J.H and M. Yogo (2005), "Testing for weak instruments in linear IV regression," In: D.W.K. Andrews and J.H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge: Cambridge University Press, pp. 80-108.

[64] Sutherland, E.H. (1947), *Principles of Criminology*, fourth edition, Chicago: J.B. Lippincott.

[65] Tayebi, M.A., Bakker, L., Glässer, U. and V. Dabbaghian (2011), "Organized crime structures in co-offending networks," *IEEE Ninth International Conference on Autonomic and Secure Computing* (DASC), pp. 846-853.

[66] Tita, G.K. Riley, J., Ridgeway, G., Grammich, C., Abrahamse, A. and P.W. Greenwood (2003), *Reducing Gun Violence: Results from an Intervention in East Los Angeles*, Santa Monica, CA: RAND Corporation.

[67] Thornberry, T.P., Krohn, M.D., Lizotte, A.J., Smith, C.A. and K. Tobin (2003), *Gangs and Delinquency in Developmental Perspective*, Cambridge: Cambridge University Press.

[68] Tremblay, R.E., Masse, L., Pagani, L. and F. Vitaro (1996), "From childhood physical aggression to adolescent maladjustment: The Montreal Prevention Experiment," In R.D. Peters and R.J. McMahon (Eds.), *Preventing Childhood Disorders, Substance Abuse, and Delinquency*, Thousand Oaks, CA: Sage Publications.

[69] United States Treasury Department (2007), *MAFIA: The Government's Secret File on Organized Crime*, Bureau of Narcotics, New York: Harper Collins Publishers.

[70] Verdier, T. and Y. Zenou (2004), "Racial beliefs, location and the causes of crime," *International Economic Review* 45, 727-756.

[71] Warr, M. (2002), *Companions in Crime: The Social Aspects of Criminal Conduct*, Cambridge: Cambridge University Press.

[72] Wasserman, S. and K. Faust (1994), *Social Network Analysis. Methods and Applications*, Cambridge: Cambridge University Press.

[73] Weerman, F.M. (2011), "Delinquent peers in context: A longitudinal network analysis of selection and influence effects," *Criminology* 49, 253-286.

[74] Xu, J.J. and H. Chen (2005), "CrimeNet Explorer: A framework for criminal network knowledge discovery," *ACM Transactions on Information Systems* 23, 201-226.

# Appendices

## A Identification of the local-aggregate model

For simplicity, we assume $k_x = 1$. Let

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \boldsymbol{\beta}_2 c_1 & 1 & -\phi_1 c_1 \\ \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 c_2 & -\phi_1 & -\phi_1 c_2 \\ \boldsymbol{\beta}_2 c_3 & 0 & 1 - \phi_1 c_3 \\ \eta_r & 0 & 0 \end{bmatrix} \tag{16}$$

and

$$\mathbf{\Lambda}_2 = \begin{bmatrix} -\boldsymbol{\beta}_2 c_1 & 1 & \phi_1 c_1 \\ \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 c_2 & -\phi_1 & \phi_1 c_2 \\ -\boldsymbol{\beta}_2 c_3 & -1 & \phi_1 c_3 + 1 \\ \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 c_4 & \phi_1 & \phi_1 c_4 - \phi_1 \\ -\boldsymbol{\beta}_2 c_5 & 0 & \phi_1 c_5 - 1 \end{bmatrix}. \tag{17}$$

**Proof of Proposition 2.** We follow the identification strategy as in Bramoullé et al. (2009) by investigating identification conditions which can be revealed via networks as functions on a vector of regressors to outcomes. First, we consider the case that, for some network $r$, $\mathbf{G}_r$ has non-constant row sums. In this case, $\mathrm{E}(\mathbf{JZ}|\mathbf{G},\mathbf{X}) = \mathbf{J}[\mathrm{E}(\mathbf{Gy}|\mathbf{G},\mathbf{X}), \mathbf{X}, \mathbf{G}^*\mathbf{X}]$ has full column rank if

$$\mathbf{J}_r[\mathrm{E}(\mathbf{G}_r\mathbf{y}_r|\mathbf{G},\mathbf{X})d_1 + \mathbf{X}_r d_2 + \mathbf{G}_r^*\mathbf{X}_r d_3] = 0 \tag{18}$$

implies $d_1 = d_2 = d_3 = 0$. As $\mathbf{J}_r = \mathbf{I}_{n_r} - \frac{1}{n_r}\boldsymbol{\iota}_{n_r}\boldsymbol{\iota}'_{n_r}$, (18) can be rewritten as

$$\mathrm{E}(\mathbf{G}_r\mathbf{y}_r|\mathbf{G},\mathbf{X})d_1 + \mathbf{X}_r d_2 + \mathbf{G}_r^*\mathbf{X}_r d_3 + \boldsymbol{\iota}_{n_r}\mu = 0, \tag{19}$$

where $\mu = -\frac{1}{n_r}\boldsymbol{\iota}'_{n_r}[\mathrm{E}(\mathbf{G}_r\mathbf{y}_r|\mathbf{G},\mathbf{X})d_1 + \mathbf{X}_r d_2 + \mathbf{G}_r^*\mathbf{X}_r d_3]$. As

$$(\mathbf{I}_{n_r} - \phi_1\mathbf{G}_r)\mathrm{E}(\mathbf{G}_r\mathbf{y}_r|\mathbf{G},\mathbf{X}) = \mathbf{G}_r\mathbf{X}_r\boldsymbol{\beta}_1 + \mathbf{G}_r\mathbf{G}_r^*\mathbf{X}_r\boldsymbol{\beta}_2 + \mathbf{G}_r\boldsymbol{\iota}_{n_r}\eta_r$$

from the reduced form equation, premultiplying (19) by $(\mathbf{I}_{n_r} - \phi_1\mathbf{G}_r)$ gives

$$\mathbf{X}_r d_2 + \mathbf{G}_r\mathbf{X}_r(\boldsymbol{\beta}_1 d_1 - \phi_1 d_2) + \mathbf{G}_r^*\mathbf{X}_r d_3 + \mathbf{G}_r\mathbf{G}_r^*\mathbf{X}_r(\boldsymbol{\beta}_2 d_1 - \phi_1 d_3) + \boldsymbol{\iota}_{n_r}\mu + \mathbf{G}_r\boldsymbol{\iota}_{n_r}(\eta_r d_1 - \phi_1\mu) = 0.$$

Note, as $\mathbf{G}_r = \mathbf{R}_r\mathbf{G}_r^*$, where $\mathbf{R}_r$ is a diagonal matrix with the $i$-th diagonal element being $\bar{g}_{i,r} = \sum_j g_{ij,r}$, $\mathbf{G}_r, \mathbf{G}_r^*$ are linearly independent if and only if rows sums of $\mathbf{G}_r$ are not constant. For $\mathbf{G}_r$ with non-constant row sums, we consider two cases. (i) $\mathbf{I}_{n_r}, \mathbf{G}_r, \mathbf{G}_r^*, \mathbf{G}_r\mathbf{G}_r^*$ are linearly independent. In this case, $[\mathbf{X}_r, \mathbf{G}_r\mathbf{X}_r, \mathbf{G}_r^*\mathbf{X}_r, \mathbf{G}_r\mathbf{G}_r^*\mathbf{X}_r]$ has full column rank. Thus, for a general $\mathbf{X}_r$, $[\mathbf{X}_r, \mathbf{G}_r\mathbf{X}_r, \mathbf{G}_r^*\mathbf{X}_r, \mathbf{G}_r\mathbf{G}_r^*\mathbf{X}_r, \boldsymbol{\iota}_{n_r}, \mathbf{G}_r\boldsymbol{\iota}_{n_r}]$ has full column rank, which implies $d_2 = \boldsymbol{\beta}_1 d_1 - \phi_1 d_2 = d_3 = \boldsymbol{\beta}_2 d_1 - \phi_1 d_3 = \mu = \eta_r d_1 - \phi_1\mu = 0$. Therefore, $d_1 = d_2 = d_3 = 0$ if $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \eta_r$ are not all zeros.

(ii) $\mathbf{G}_r\mathbf{G}_r^* = c_1\mathbf{I}_{n_r} + c_2\mathbf{G}_r + c_3\mathbf{G}_r^*$ for some constant scalars $c_1, c_2, c_3$. In this case,

$$\begin{aligned} 0 &= \mathbf{X}_r(\boldsymbol{\beta}_2 c_1 d_1 + d_2 - \phi_1 c_1 d_3) + \mathbf{G}_r\mathbf{X}_r[(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 c_2)d_1 - \phi_1 d_2 - \phi_1 c_2 d_3] \\ &\quad + \mathbf{G}_r^*\mathbf{X}_r[\boldsymbol{\beta}_2 c_3 d_1 + (1-\phi_1 c_3)d_3] + \boldsymbol{\iota}_{n_r}\mu + \mathbf{G}_r\boldsymbol{\iota}_{n_r}(\eta_r d_1 - \phi_1\mu), \end{aligned}$$

which implies $d_1 = d_2 = d_3 = 0$ if $\mathbf{\Lambda}_1$ given by (16) has full rank. When $\eta_r \neq 0$, a sufficient condition for $\mathbf{\Lambda}_1$ to have full rank is $|\phi_1 c_1 + c_2| + |1 - \phi_1 c_3| \neq 0$.

Next, consider the case that $\mathbf{G}_r$ has constant row sums such that $\bar{g}_{i,r} = \bar{g}_r$ for all $r$. In this case, $\mathbf{G}_r = g_r\mathbf{G}_r^*$. $\mathrm{E}(\mathbf{JZ}|\mathbf{G},\mathbf{X}) = \mathbf{J}[\mathrm{E}(\mathbf{Gy}|\mathbf{G},\mathbf{X}), \mathbf{X}, \mathbf{G}^*\mathbf{X}]$ has full column rank if

$$\mathbf{J}[\mathrm{E}(\mathbf{Gy}|\mathbf{G},\mathbf{X})d_1 + \mathbf{X}d_2 + \mathbf{G}^*\mathbf{X}d_3] = 0 \tag{20}$$

implies $d_1 = d_2 = d_3 = 0$. As $\mathbf{J}(\mathbf{I} - \phi_1\mathbf{G})^{-1}\mathbf{GL} = 0$, substitution of $\mathrm{E}(\mathbf{Gy}|\mathbf{G},\mathbf{X}) = (\mathbf{I} - \phi_1\mathbf{G})^{-1}(\mathbf{GX}\boldsymbol{\beta}_1 + \mathbf{GG}^*\mathbf{X}\boldsymbol{\beta}_2 + \mathbf{GL}\eta)$ into (20) gives

$$\mathbf{J}(\mathbf{I} - \phi_1\mathbf{G})^{-1}[\mathbf{X}d_2 + \mathbf{GX}(\boldsymbol{\beta}_1 d_1 - \phi_1 d_2) + \mathbf{G}^*\mathbf{X}d_3 + \mathbf{GG}^*\mathbf{X}(\boldsymbol{\beta}_2 d_1 - \phi_1 d_3)] = 0,$$

which implies

$$\mathbf{X}d_2 + \mathbf{GX}(\boldsymbol{\beta}_1 d_1 - \phi_1 d_2) + \mathbf{G}^*\mathbf{X}d_3 + \mathbf{GG}^*\mathbf{X}(\boldsymbol{\beta}_2 d_1 - \phi_1 d_3) = c\mathbf{L} \tag{21}$$

because $\mathbf{JL} = 0$. As $\mathbf{G}^*\mathbf{L} = \mathbf{L}$, premultiplying (21) by $\mathbf{G}^*$ gives

$$\mathbf{G}^*\mathbf{X}d_2 + \mathbf{GG}^*\mathbf{X}(\boldsymbol{\beta}_1 d_1 - \phi_1 d_2) + \mathbf{G}^{*2}\mathbf{X}d_3 + \mathbf{GG}^{*2}\mathbf{X}(\boldsymbol{\beta}_2 d_1 - \phi_1 d_3) = c\mathbf{L}. \tag{22}$$

From (21) and (22), when $\mathbf{I}, \mathbf{G}, \mathbf{G}^*, \mathbf{GG}^*, \mathbf{G}^{*2}, \mathbf{GG}^{*2}$ are linearly independent, $d_1 = d_2 = d_3 = 0$ if $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ are not both zeros. When $\mathbf{GG}^{*2} = c_1\mathbf{I} + c_2\mathbf{G} + c_3\mathbf{G}^* + c_4\mathbf{GG}^* + c_5\mathbf{G}^{*2}$ and $\mathbf{I}, \mathbf{G}, \mathbf{G}^*, \mathbf{GG}^*, \mathbf{G}^{*2}$ are linearly independent, $d_1 = d_2 = d_3 = 0$ if $\mathbf{\Lambda}_2$ given by (17) has full rank. On the other hand, if $\bar{g}_r = \bar{g}$ for all $r$, then $\mathbf{G} = g\mathbf{G}^*$. When $\mathbf{I}, \mathbf{G}^*, \mathbf{G}^{*2}, \mathbf{G}^{*3}$ are linearly independent, (21) and (22) imply $d_1 = d_2 = d_3 = 0$ if $\bar{g}\phi_1\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \neq 0$. ∎

# B Estimation of the general network model

We consider the 2SLS and GMM estimation of the (within-transformed) general network model (10). This appendix presents the derivation and asymptotic properties of the estimators.

For any $n \times n$ matrix $\mathbf{A} = [a_{ij}]$, let $\mathrm{vec}_D(\mathbf{A}) = (a_{11}, \cdots, a_{nn})'$, $\mathbf{A}^s = \mathbf{A} + \mathbf{A}'$, $\mathbf{A}^t = \mathbf{A} - \mathrm{tr}(\mathbf{A})\mathbf{J}/\mathrm{tr}(\mathbf{J})$, and $\mathbf{A}^-$ denote a generalized inverse of a square matrix $\mathbf{A}$. Let $\mu_3$ and $\mu_4$ denote, respectively, the third and fourth moments of the error term $\epsilon_{i,r}$.

## B.1 2SLS estimation

Let $\mathbf{M} = (\mathbf{I} - \phi_1\mathbf{G} - \phi_2\mathbf{G}^*)^{-1}$, $\mathbf{X}^* = [\mathbf{X}, \mathbf{G}^*\mathbf{X}]$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$. From the reduced form equation, $\mathrm{E}(\mathbf{y}|\mathbf{G},\mathbf{X}) = \mathbf{M}(\mathbf{X}^*\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\eta})$. For $\mathbf{Z} = [\mathbf{Gy}, \mathbf{G}^*\mathbf{y}, \mathbf{X}^*]$, the ideal IV matrix for the regressor matrix

$\mathbf{JZ}$ in (10) is given by

$$\mathbf{F} = \mathrm{E}(\mathbf{JZ}|\mathbf{G},\mathbf{X}) = \mathbf{J}[\mathrm{E}(\mathbf{Gy}|\mathbf{G},\mathbf{X}), \mathrm{E}(\mathbf{G}^*\mathbf{y}|\mathbf{G},\mathbf{X}), \mathbf{X}^*]. \tag{23}$$

However, this IV matrix is infeasible as it involves unknown parameters. Note that $\mathbf{F}$ can be considered as a linear combination of the IVs in the matrix $\mathbf{Q}_\infty = \mathbf{J}[\mathbf{GMX}^*, \mathbf{GML}, \mathbf{G}^*\mathbf{MX}^*, \mathbf{G}^*\mathbf{ML}, \mathbf{X}^*]$. As $\mathbf{L}$ has $\bar{r}$ columns, the number of IVs in $\mathbf{Q}_\infty$ increases as the number of groups $\bar{r}$ increases. Furthermore, if $\bar{g}^{\max}|\phi_1| + |\phi_2| < 1$, we have $\mathbf{M} = (\mathbf{I} - \phi_1\mathbf{G} - \phi_2\mathbf{G}^*)^{-1} = \sum_{j=0}^{\infty}(\phi_1\mathbf{G} + \phi_2^*\mathbf{G})^j$. Hence, $\mathbf{M}$ can be approximated by a linear combination of $[\mathbf{I}, \mathbf{G}, \mathbf{G}^*, \mathbf{G}^2, \mathbf{GG}^*, \mathbf{G}^*\mathbf{G}, \mathbf{G}^{*2}, \cdots]$.

To achieve asymptotic efficiency, we assume the number of IVs increases with the sample size so that the ideal IV matrix $\mathbf{F}$ can be approximated by a feasible IV matrix $\mathbf{Q}_K$ with an approximation error diminishing to zero. That is, for an $n \times K$ IV matrix $\mathbf{Q}_K$, there exists some conformable matrix $\boldsymbol{\xi}_K$ such that $||\mathbf{F} - \mathbf{Q}_K\boldsymbol{\xi}_K||_\infty \to 0$ as $n, K \to \infty$. Let $\mathbf{P}_K = \mathbf{Q}_K(\mathbf{Q}_K'\mathbf{Q}_K)^{-}\mathbf{Q}_K'$, the 2SLS estimator of $\boldsymbol{\theta} = (\phi_1, \phi_2, \boldsymbol{\beta}')'$ is given by $\hat{\boldsymbol{\theta}}_{2sls} = (\mathbf{Z}'\mathbf{P}_K\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_K\mathbf{y}$.

If $K/n \to 0$, then it follows by a similar argument as in Liu and Lee (2010) that $\sqrt{n}(\hat{\boldsymbol{\theta}}_{2sls} - \boldsymbol{\theta} - \mathbf{b}_{2sls}) \xrightarrow{d} N(0, \sigma^2\bar{\mathbf{H}}^{-1})$, where $\bar{\mathbf{H}} = \lim_{n\to\infty}\frac{1}{n}\mathbf{F}'\mathbf{F}$ and $\mathbf{b}_{2sls} = \sigma^2(\mathbf{Z}'\mathbf{P}_K\mathbf{Z})^{-1}[\mathrm{tr}(\mathbf{P}_K\mathbf{GM}), \mathrm{tr}(\mathbf{P}_K\mathbf{G}^*\mathbf{M}), \mathbf{0}]' = O_p(K/n)$. The 2SLS estimator has an asymptotic bias term due to the large number of IVs. When $K^2/n \to 0$, the leading order bias term $\sqrt{n}\mathbf{b}_{2sls}$ converges to zero and the proposed 2SLS estimator is the most efficient IV-based estimator as the variance matrix $\sigma^2\bar{\mathbf{H}}^{-1}$ attains the efficiency lower bound for the class of IV estimators.

To correct for the many-instrument bias in the 2SLS estimator, one can estimate the leading order bias term and adjust the 2SLS estimator by the estimated leading-order bias $\tilde{\mathbf{b}}_{2sls}$. With $\sqrt{n}$-consistent initial estimates $\tilde{\sigma}^2, \tilde{\phi}_1, \tilde{\phi}_2$, the bias-corrected 2SLS (BC2SLS) estimator is given by $\hat{\boldsymbol{\theta}}_{bc2sls} = \hat{\boldsymbol{\theta}}_{2sls} - \tilde{\mathbf{b}}_{2sls}$, where $\tilde{\mathbf{b}}_{2sls} = \tilde{\sigma}^2(\mathbf{Z}'\mathbf{P}_K\mathbf{Z})^{-1}[\mathrm{tr}(\mathbf{P}_K\mathbf{G}\tilde{\mathbf{M}}), \mathrm{tr}(\mathbf{P}_K\mathbf{G}^*\tilde{\mathbf{M}}), \mathbf{0}]'$ and $\tilde{\mathbf{M}} = (\mathbf{I} - \tilde{\phi}_1\mathbf{G} - \tilde{\phi}_2\mathbf{G}^*)^{-1}$. The BC2SLS estimator $\sqrt{n}(\hat{\boldsymbol{\theta}}_{bc2sls} - \boldsymbol{\theta}) \xrightarrow{d} N(0, \sigma^2\bar{\mathbf{H}}^{-1})$ when $K/n \to 0$.

### B.2 GMM estimation

The 2SLS estimator can be generalized to the GMM with additional quadratic moment equations. Let $\boldsymbol{\epsilon}(\boldsymbol{\theta}) = \mathbf{J}(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta})$. The IV moment conditions $\mathbf{Q}_K'\boldsymbol{\epsilon}(\boldsymbol{\theta}) = 0$ are linear in $\boldsymbol{\epsilon}$ at the true value of $\boldsymbol{\theta}$. As $\mathrm{E}(\boldsymbol{\epsilon}'\mathbf{U}_1\boldsymbol{\epsilon}|\mathbf{G})$ =$\mathrm{E}(\boldsymbol{\epsilon}'\mathbf{U}_2\boldsymbol{\epsilon}|\mathbf{G}) = 0$ for $\mathbf{U}_1 = (\mathbf{JGMJ})^t$ and $\mathbf{U}_2 = (\mathbf{JG}^*\mathbf{MJ})^t$, the quadratic moment conditions for estimation are given by $[\mathbf{U}_1\boldsymbol{\epsilon}(\boldsymbol{\theta}), \mathbf{U}_1\boldsymbol{\epsilon}(\boldsymbol{\theta})]'\boldsymbol{\epsilon}(\boldsymbol{\theta}) = \mathbf{0}$. The proposed quadratic moment conditions can be shown to be optimal (in terms of efficiency of the GMM estimator) under normality (see Lee and Liu, 2010). The vector of linear and quadratic empirical moments for the GMM estimation is given by $\mathbf{g}(\boldsymbol{\theta}) = [\mathbf{Q}_K, \mathbf{U}_1\boldsymbol{\epsilon}(\boldsymbol{\theta}), \mathbf{U}_1\boldsymbol{\epsilon}(\boldsymbol{\theta})]'\boldsymbol{\epsilon}(\boldsymbol{\theta})$.

In order for inference based on the following asymptotic results to be robust, we do not impose the normality assumption in the following analysis. The variance matrix of $\mathbf{g}(\boldsymbol{\theta})$ at the true value of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\Omega} = \mathrm{Var}[\mathbf{g}(\boldsymbol{\theta})|\mathbf{G},\mathbf{X}] = \begin{pmatrix} \sigma^2\mathbf{Q}_K'\mathbf{Q}_K & \mu_3\mathbf{Q}_K'\boldsymbol{\omega} \\ \mu_3\boldsymbol{\omega}'\mathbf{Q}_K & (\mu_4 - 3\sigma^4)\boldsymbol{\omega}'\boldsymbol{\omega} + \sigma^4\boldsymbol{\Upsilon} \end{pmatrix},$$

where $\boldsymbol{\omega} = [\mathrm{vec}_D(\mathbf{U}_1), \mathrm{vec}_D(\mathbf{U}_2)]$ and $\boldsymbol{\Upsilon} = \frac{1}{2}[\mathrm{vec}(\mathbf{U}_1^s), \mathrm{vec}(\mathbf{U}_2^s)]'[\mathrm{vec}(\mathbf{U}_1^s), \mathrm{vec}(\mathbf{U}_2^s)]$. By the generalized Schwarz inequality, the optimal GMM estimator is given by

$$\hat{\boldsymbol{\theta}}_{gmm} = \arg\min \mathbf{g}(\boldsymbol{\theta})'\boldsymbol{\Omega}^{-1}\mathbf{g}(\boldsymbol{\theta}). \tag{24}$$

Let $\mathbf{B}^{-1} = (\mu_4 - 3\sigma^4)\boldsymbol{\omega}'\boldsymbol{\omega} + \sigma^4\boldsymbol{\Upsilon} - \frac{\mu_3^2}{\sigma^2}\boldsymbol{\omega}'\mathbf{P}_K\boldsymbol{\omega}$,

$$\mathbf{D} = -\sigma^2 \begin{pmatrix} \operatorname{tr}(\mathbf{U}_1^s\mathbf{GM}) & \operatorname{tr}(\mathbf{U}_1^s\mathbf{G}^*\mathbf{M}) & \mathbf{0} \\ \operatorname{tr}(\mathbf{U}_2^s\mathbf{GM}) & \operatorname{tr}(\mathbf{U}_2^s\mathbf{G}^*\mathbf{M}) & \mathbf{0} \end{pmatrix},$$

$\bar{\mathbf{D}} = \mathbf{D} - \frac{\mu_3}{\sigma^2}\boldsymbol{\omega}'\mathbf{F}$, and $\check{\mathbf{D}} = \mathbf{D} - \frac{\mu_3}{\sigma^2}\boldsymbol{\omega}'\mathbf{P}_K\mathbf{Z}$. When $K^{3/2}/n \to 0$, the optimal GMM estimator[24] has the asymptotic distribution

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta} - \mathbf{b}_{gmm}) \xrightarrow{d} N(0, (\sigma^{-2}\bar{\mathbf{H}} + \lim_{n\to\infty}\frac{1}{n}\bar{\mathbf{D}}'\mathbf{B}\bar{\mathbf{D}})^{-1}), \tag{25}$$

where $\mathbf{b}_{gmm} = (\sigma^{-2}\mathbf{Z}'\mathbf{P}_K\mathbf{Z} + \check{\mathbf{D}}'\mathbf{B}\check{\mathbf{D}})^{-1}[\operatorname{tr}(\mathbf{P}_K\mathbf{GM}), \operatorname{tr}(\mathbf{P}_K\mathbf{G}^*\mathbf{M}), \mathbf{0}]' = O(K/n)$.

As the asymptotic bias $\sqrt{n}\mathbf{b}_{gmm}$ is $O(K/\sqrt{n})$, the asymptotic distribution of the GMM estimator $\hat{\boldsymbol{\theta}}_{gmm}$ will be centered at the true value of $\boldsymbol{\theta}$ only if $K^2/n \to 0$. With a consistently estimated leading order bias $\tilde{\mathbf{b}}_{gmm}$, the bias-corrected GMM (BCGMM) estimator $\hat{\boldsymbol{\theta}}_{bcgmm} = \hat{\boldsymbol{\theta}}_{gmm} - \tilde{\mathbf{b}}_{gmm}$ has a proper centered asymptotic normal distribution as given in (25) if $K^{3/2}/n \to 0$.

The asymptotic variance matrix of the many-IV GMM estimator can be compared with that of the many-IV 2SLS estimator. As $\bar{\mathbf{D}}'\mathbf{B}\bar{\mathbf{D}}$ is nonnegative definite, the asymptotic variance of the many-IV GMM estimator is smaller relative to that of the 2SLS estimator. Thus, the many-IV GMM estimator with additional quadratic moments improves efficiency upon the 2SLS estimator.

[24] The weighting matrices for quadratic moments $\mathbf{U}_1, \mathbf{U}_2$ and the optimal weighting matrix for the objective function $\boldsymbol{\Omega}^{-1}$ involves unknown parameters $\phi_1, \phi_2, \sigma^2, \mu_3$ and $\mu_4$. With consistent preliminary estimators of those unknown parameters, the feasible optimal GMM estimator can be shown to have the same asymptotic distribution given by (25).

Table 1: Data Summary

| Variable | Definition | Sample with networks 3-1786 ($n = 64186$) | | Sample with networks 3-200 ($n = 3785$) | |
|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD |
| delinquency | In the text | 1.16 | 1.02 | 1.03 | 0.95 |
| age | Age | 15.03 | 1.69 | 13.77 | 1.63 |
| female | 1 if female | 0.52 | 0.50 | 0.56 | 0.50 |
| white | 1 if white | 0.60 | 0.49 | 0.48 | 0.50 |
| born in the U.S. | 1 if born in the U.S. | 0.92 | 0.27 | 0.94 | 0.24 |
| grade | 7th-12th grades are coded as 1,…,6 | 3.67 | 1.59 | 2.36 | 1.50 |
| fitness | 1 if physically fit | 0.66 | 0.48 | 0.62 | 0.49 |
| math score | 1 if the recent math grade is A or B | 0.53 | 0.50 | 0.52 | 0.50 |
| self esteem | 1 if thinks himself/herself has a lot of good qualities | 0.79 | 0.40 | 0.73 | 0.44 |
| school attachment | 1 if feels like being part of the school | 0.57 | 0.50 | 0.58 | 0.49 |
| neighborhood safety | 1 if feels safe in the neighbourhood | 0.72 | 0.45 | 0.63 | 0.48 |
| school safety | 1 if feels safe at school | 0.61 | 0.49 | 0.60 | 0.49 |
| live with both parents | 1 if lives with both parents | 0.74 | 0.44 | 0.71 | 0.45 |
| parental care | 1 if parents care very much | 0.84 | 0.36 | 0.85 | 0.36 |
| *Parent Education* | | | | | |
| (less than HS) | 1 if parent's education is less than high school (HS) | 0.11 | 0.32 | 0.14 | 0.35 |
| HS grad | 1 if parent's education is HS or higher but no college degree | 0.45 | 0.50 | 0.46 | 0.50 |
| college grad | 1 if parent's education is college or higher | 0.32 | 0.47 | 0.24 | 0.42 |
| missing | 1 if parent's education information is missing | 0.11 | 0.32 | 0.17 | 0.37 |
| *Parent Job* | | | | | |
| (stay home) | 1 if parent is a homemaker, retired, or does not work | 0.09 | 0.29 | 0.12 | 0.32 |
| professional | 1 if parent's job is a doctor, lawyer, scientist, teacher, librarian, nurse, manager, executive, director | 0.28 | 0.45 | 0.19 | 0.39 |
| other jobs | 1 if parent's job is not "stay home" or "professional" | 0.54 | 0.50 | 0.57 | 0.49 |
| missing | 1 if parent's job information is missing | 0.09 | 0.28 | 0.12 | 0.33 |

The variable in the parentheses is the reference category.
If both parents are in the household, the education and job of the father is considered.

Table 2: Estimation of the General Network Model (Reciprocal Nominations)

| | 2SLS | BC2SLS | GMM | BCGMM |
|---|---|---|---|---|
| local-aggregate peer effect | 0.0112*** (0.0047) | 0.0118*** (0.0044) | 0.0144*** (0.0040) | 0.0158*** (0.0039) |
| local-average peer effect | 0.5761*** (0.1502) | 0.4153*** (0.0983) | 0.1260*** (0.0286) | 0.1082*** (0.0281) |
| age | 0.0717*** (0.0240) | 0.0744*** (0.0236) | 0.0794*** (0.0233) | 0.0796*** (0.0233) |
| female | -0.3314*** (0.0333) | -0.3299*** (0.0327) | -0.3273*** (0.0324) | -0.3273*** (0.0323) |
| white | -0.0521 (0.0436) | -0.0490 (0.0428) | -0.0452 (0.0423) | -0.0450 (0.0423) |
| born in the U.S. | 0.1123 (0.0683) | 0.1199* (0.0670) | 0.1341** (0.0661) | 0.1351** (0.0661) |
| grade | 0.0173 (0.0408) | 0.0120 (0.0400) | 0.0024 (0.0394) | 0.0013 (0.0394) |
| fitness | 0.0273 (0.0366) | 0.0267 (0.0360) | 0.0235 (0.0356) | 0.0233 (0.0356) |
| math score | -0.1403*** (0.0322) | -0.1475*** (0.0313) | -0.1602*** (0.0306) | -0.1615*** (0.0306) |
| self esteem | -0.0392 (0.0407) | -0.0445 (0.0399) | -0.0535 (0.0393) | -0.0548 (0.0393) |
| school attachment | -0.0892*** (0.0360) | -0.0950*** (0.0352) | -0.1065*** (0.0346) | -0.1079*** (0.0346) |
| neighborhood safety | -0.0846** (0.0386) | -0.0797** (0.0378) | -0.0724* (0.0373) | -0.0718* (0.0373) |
| school safety | -0.1271*** (0.0404) | -0.1325*** (0.0396) | -0.1414*** (0.0390) | -0.1423*** (0.0390) |
| live with both parents | 0.0281 (0.0363) | 0.0261 (0.0356) | 0.0233 (0.0352) | 0.0224 (0.0352) |
| parental care | -0.3182*** (0.0461) | -0.3261*** (0.0450) | -0.3407*** (0.0443) | -0.3418*** (0.0443) |
| parent education: HS grad | -0.0710 (0.0469) | -0.0740 (0.0461) | -0.0802* (0.0456) | -0.0814* (0.0456) |
| parent education: college grad | -0.1634*** (0.0560) | -0.1716*** (0.0548) | -0.1876*** (0.0540) | -0.1908*** (0.0540) |
| parent education: missing | -0.2322*** (0.0582) | -0.2425*** (0.0568) | -0.2611*** (0.0559) | -0.2637*** (0.0559) |
| parent job: professional | -0.0328 (0.0589) | -0.0379 (0.0578) | -0.0493 (0.0571) | -0.0496 (0.0571) |
| parent job: other | -0.0707 (0.0494) | -0.0670 (0.0485) | -0.0622 (0.0480) | -0.0610 (0.0479) |
| parent job: missing | -0.1529*** (0.0648) | -0.1517*** (0.0637) | -0.1507*** (0.0630) | -0.1500*** (0.0630) |
| contextual effects | Yes | Yes | Yes | Yes |
| network fixed effects | Yes | Yes | Yes | Yes |

Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.
2SLS First Stage F test statistic: 4.287; OIR test p-value: 0.219

Table 3: Estimation of the General Network Model (Actual Nominations)

| | 2SLS | BC2SLS | GMM | BCGMM |
|---|---|---|---|---|
| local-aggregate peer effect | 0.0074 | 0.0046 | 0.0147** | 0.0130** |
| | (0.0072) | (0.0071) | (0.0064) | (0.0063) |
| local-average peer effect | 0.3726*** | 0.2219*** | 0.1499*** | 0.1474*** |
| | (0.1318) | (0.0933) | (0.0317) | (0.0313) |
| age | 0.0796*** | 0.0844*** | 0.0849*** | 0.0852*** |
| | (0.0238) | (0.0235) | (0.0233) | (0.0233) |
| female | -0.3225*** | -0.3215*** | -0.3219*** | -0.3218*** |
| | (0.0330) | (0.0327) | (0.0327) | (0.0327) |
| white | -0.0355 | -0.0343 | -0.0348 | -0.0346 |
| | (0.0422) | (0.0418) | (0.0418) | (0.0418) |
| born in the U.S. | 0.1429** | 0.1422** | 0.1440** | 0.1435** |
| | (0.0669) | (0.0664) | (0.0663) | (0.0663) |
| grade | 0.0261 | 0.0216 | 0.0210 | 0.0206 |
| | (0.0321) | (0.0317) | (0.0316) | (0.0316) |
| fitness | 0.0398 | 0.0357 | 0.0322 | 0.0313 |
| | (0.0361) | (0.0357) | (0.0356) | (0.0356) |
| math score | -0.1465*** | -0.1532*** | -0.1564*** | -0.1573*** |
| | (0.0315) | (0.0310) | (0.0307) | (0.0307) |
| self esteem | -0.0589 | -0.0623 | -0.0634 | -0.0645 |
| | (0.0398) | (0.0395) | (0.0394) | (0.0394) |
| school attachment | -0.1038*** | -0.1048*** | -0.1103*** | -0.1086*** |
| | (0.0350) | (0.0347) | (0.0346) | (0.0346) |
| neighborhood safety | -0.0766** | -0.0720* | -0.0721* | -0.0712* |
| | (0.0379) | (0.0375) | (0.0374) | (0.0374) |
| school safety | -0.1374*** | -0.1399*** | -0.1399*** | -0.1402*** |
| | (0.0394) | (0.0391) | (0.0390) | (0.0390) |
| live with both parents | 0.0244 | 0.0234 | 0.0244 | 0.0241 |
| | (0.0356) | (0.0353) | (0.0352) | (0.0352) |
| parental care | -0.3223*** | -0.3295*** | -0.3325*** | -0.3332*** |
| | (0.0451) | (0.0445) | (0.0443) | (0.0443) |
| parent education: HS grad | -0.0685 | -0.0745 | -0.0794* | -0.0799* |
| | (0.0465) | (0.0460) | (0.0458) | (0.0458) |
| parent education: college grad | -0.1571*** | -0.1704*** | -0.1794*** | -0.1807*** |
| | (0.0558) | (0.0548) | (0.0542) | (0.0542) |
| parent education: missing | -0.2590*** | -0.2654*** | -0.2712*** | -0.2715*** |
| | (0.0567) | (0.0561) | (0.0560) | (0.0560) |
| parent job: professional | -0.0391 | -0.0410 | -0.0452 | -0.0450 |
| | (0.0578) | (0.0573) | (0.0572) | (0.0572) |
| parent job: other | -0.0613 | -0.0604 | -0.0624 | -0.0621 |
| | (0.0484) | (0.0480) | (0.0480) | (0.0480) |
| parent job: missing | -0.1257** | -0.1300** | -0.1314** | -0.1327** |
| | (0.0637) | (0.0631) | (0.0630) | (0.0630) |
| contextual effects | Yes | Yes | Yes | Yes |
| network fixed effects | Yes | Yes | Yes | Yes |

Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.
2SLS First Stage F test statistic: 4.498; OIR test p-value: 0.392

Table 4: Estimation of Local-aggregate and Local-average Models Separately (Reciprocal Nominations)

| | Local-aggregate Model | Local-average Model |
|---|---|---|
| | BCGMM | QML |
| local-aggregate peer effect | 0.0249*** | |
| | (0.0033) | |
| local-average peer effect | | 0.1677*** |
| | | (0.0242) |
| age | 0.0810*** | 0.0788*** |
| | (0.0233) | (0.0233) |
| female | -0.3291*** | -0.3240*** |
| | (0.0324) | (0.0324) |
| white | -0.0426 | -0.0436 |
| | (0.0424) | (0.0425) |
| born in the U.S. | 0.1393** | 0.1339** |
| | (0.0662) | (0.0663) |
| grade | -0.0011 | 0.0008 |
| | (0.0395) | (0.0395) |
| fitness | 0.0209 | 0.0294 |
| | (0.0356) | (0.0356) |
| math score | -0.1659*** | -0.1586*** |
| | (0.0306) | (0.0307) |
| self esteem | -0.0581 | -0.0558 |
| | (0.0394) | (0.0394) |
| school attachment | -0.1188*** | -0.0947*** |
| | (0.0345) | (0.0345) |
| neighborhood safety | -0.0705* | -0.0692* |
| | (0.0374) | (0.0374) |
| school safety | -0.1462*** | -0.1413*** |
| | (0.0391) | (0.0391) |
| live with both parents | 0.0208 | 0.0222 |
| | (0.0353) | (0.0353) |
| parental care | -0.3458*** | -0.3408*** |
| | (0.0444) | (0.0444) |
| parent education: HS grad | -0.0837* | -0.0800* |
| | (0.0456) | (0.0457) |
| parent education: college grad | -0.1981*** | -0.1858*** |
| | (0.0541) | (0.0542) |
| parent education: missing | -0.2682*** | -0.2651*** |
| | (0.0559) | (0.0560) |
| parent job: professional | -0.0566 | -0.0418 |
| | (0.0572) | (0.0573) |
| parent job: other | -0.0609 | -0.0566 |
| | (0.0480) | (0.0481) |
| parent job: missing | -0.1480*** | -0.1502*** |
| | (0.0631) | (0.0632) |
| contextual effects | Yes | Yes |
| network fixed effects | Yes | Yes |

Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.

Table 5: Logit Regression of the Link Formation Model

| | |
|---|---|
| BCGMM estimation residuals | -0.0003<br>(0.0006) |
| age | 0.2107<br>(0.1400) |
| gender | 0.4469**<br>(0.2110) |
| race | 0.9396***<br>(0.3284) |
| place of birth | 3.7119***<br>(0.7549) |
| grade | 1.6042***<br>(0.3014) |
| fitness | -0.1280<br>(0.1880) |
| math score | 0.5768**<br>(0.2389) |
| self esteem | 0.1085<br>(0.2151) |
| school attachment | -0.0440<br>(0.1918) |
| neighborhood safety | -0.0248<br>(0.1954) |
| school safety | -0.2204<br>(0.1920) |
| live with both parents | 0.1054<br>(0.3476) |
| parental care | -0.5231*<br>(0.2764) |
| parent education | -0.0835<br>(0.1099) |
| parent job | 0.3226**<br>(0.1453) |

Sample size: 219,671.

Standard errors in parentheses.

Statistical significance: ***p<0.01; **p<0.05; *p<0.1.

Table 6: Characteristics of the Key Player

| Variable | Key Players | | Other Students | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | p value |
| Own Characteristics | | | | | |
| age | 14.52 | 1.89 | 13.75 | 1.62 | 0.00 |
| white | 0.34 | 0.48 | 0.49 | 0.50 | 0.00 |
| grade | 2.92 | 1.60 | 2.34 | 1.50 | 0.00 |
| math score | 0.33 | 0.47 | 0.53 | 0.50 | 0.00 |
| self esteem | 0.59 | 0.49 | 0.74 | 0.44 | 0.00 |
| school attachment | 0.27 | 0.45 | 0.59 | 0.49 | 0.00 |
| school safety | 0.45 | 0.50 | 0.61 | 0.49 | 0.00 |
| parental care | 0.70 | 0.46 | 0.85 | 0.35 | 0.00 |
| parent job: professional | 0.50 | 0.50 | 0.57 | 0.49 | 0.09 |
| parent job: other jobs | 0.17 | 0.38 | 0.12 | 0.32 | 0.07 |
| Friends' Characteristics | | | | | |
| age | 14.44 | 1.66 | 13.73 | 1.47 | 0.00 |
| white | 0.36 | 0.42 | 0.49 | 0.42 | 0.00 |
| born in the U.S. | 0.89 | 0.25 | 0.94 | 0.18 | 0.03 |
| grade | 2.92 | 1.53 | 2.34 | 1.43 | 0.00 |
| math score | 0.42 | 0.31 | 0.54 | 0.33 | 0.00 |
| school attachment | 0.52 | 0.31 | 0.61 | 0.31 | 0.00 |
| neighborhood safety | 0.56 | 0.34 | 0.65 | 0.31 | 0.00 |
| school safety | 0.49 | 0.34 | 0.62 | 0.33 | 0.00 |

The two-sample t test p value is reported in the last column.