DISCUSSION PAPER SERIES

No. 8185

CRIMINAL NETWORKS: WHO IS THE KEY PLAYER?

Lung-Fei Lee, Xiaodong Liu, Eleonora Patacchini and Yves Zenou

LABOUR ECONOMICS and PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP8185.asp**

ISSN 0265-8003

CRIMINAL NETWORKS: WHO IS THE KEY PLAYER?

Lung-Fei Lee, Ohio State University
Xiaodong Liu, University of Colorado, Boulder
Eleonora Patacchini, Università di Roma La Sapienza and CEPR
Yves Zenou, Stockholm University and CEPR

Discussion Paper No. 8185
January 2011

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **LABOUR ECONOMICS and PUBLIC POLICY**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Lung-Fei Lee, Xiaodong Liu, Eleonora Patacchini and Yves Zenou

CEPR Discussion Paper No. 8185

January 2011

ABSTRACT

Criminal Networks: Who is the Key Player?*

We analyze delinquent networks of adolescents in the United States. We develop a theoretical model showing who the key player is, i.e. the criminal who once removed generates the highest possible reduction in aggregate crime level. We also show that key players are not necessary the most active criminals in a network. We then test our model using data on criminal behaviors of adolescents in the United States (AddHealth data). Compared to other criminals, key players are more likely to be a male, have less educated parents, are less attached to religion and feel socially more excluded. They also feel that adults care less about them, are less attached to their school and have more troubles getting along with the teachers. We also find that, even though some criminals are not very active in criminal activities, they can be key players because they have a crucial position in the network in terms of betweenness centrality.

JEL Classification: A14, D85, K42 and Z13
Keywords: betweenness centrality, Bonacich centrality, crime, crime policies and network characteristics

Lung-Fei Lee
Department of Economics
Ohio State University
475 Arps Hall,1945 N High Street
Columbus OH 43210
USA

Email:
lflee@ecolan.sbs.ohio-state.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=154441

Xiaodong Liu
Department of Economics
University of Colorado at Boulder
256 UCB
Boulder, Colorado 80309-0256
USA

Email:
xiaodong.liu@colorado.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=173030

Eleonora Patacchini
Faculty of Statistics
Universita di Roma 'La Sapienza'
Box 83 - RM62, P.le Aldo Moro 5
00185 Roma
ITALY

Email:
eleonora.patacchini@uniroma1.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=158584

Yves Zenou
Department of Economics
Stockholm University
10691 Stockholm
SWEDEN

Email:
yves.zenou@ne.su.se

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=126027

* We would like to thank the participants of the Einaudi Institute of Economics and Finance lunch seminar and of the Iowa State University departmental seminar for helpful comments, in particular, David Levine, Marcus Mobius and Tanya Rosenblat.

Submitted 17 December 2010

1 Introduction

There are 2.3 million people behind bars at any moment of 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] Even as the prisoner head count continues to rise, crime has stopped falling, and poor people and minorities still bear the brunt of both crime and punishment. We need to cut both crime and the prison population in half within a decade.

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 those actions cost money, currently about $200 billion per year nationwide. This "brute force" policy does not seem to work well since, for example, the cost of prison in California is higher than the cost of education[2] and crime rates do not seem to decrease.

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 criminals", i.e. criminals who once removed generate the highest possible reduction in aggregate crime level in a network, can have large effects on crime because of the feedback effects or "social multipliers" at work (see, in particular, Sah, 1991; Kleiman, 1993, 2009; Glaeser et al., 1996; Rasmussen, 1996; Schrag and Scotchmer, 1997; Verdier and Zenou, 2004). That is, as the fraction of individuals participating in a criminal behavior increases, the impact on others is multiplied through social networks. Thus, criminal behaviors can be magnified, and interventions can become more effective. The impacts from social networks may also be particularly important for adolescents because this developmental pe-

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

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

riod 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 is, to some extent, a group phenomenon, and the source of crime and delinquency is located in the intimate social networks of individuals (see e.g. Sutherland, 1947; Sarnecki, 2001; Warr, 2002; Haynie, 2001; Patacchini and Zenou, 2008; 2011). Indeed, 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 crimes. In fact, not only friends but also the *structure* of social networks matters in explaining 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.

Following Ballester et al. (2006, 2010), we propose a theoretical model of criminal networks. Building on the Beckerian incentives approach to delinquency, we develop a model where peer effects matter so that criminals are directly influenced by their friends. Individuals decide non-cooperatively their crime effort and we show that, in equilibrium, each criminal effort is equal to his/her Katz-Bonacich centrality.[3] The Katz-Bonacich centrality measure is an index of *connectivity* that not only takes into account the number of direct links a given delinquent has but also all his indirect connections. In our delinquency game, the network payoff interdependence is restricted to direct network mates. But, because clusters of direct friends overlap, this local payoff interdependence spreads all over the network. In equilibrium, individual decisions emanate from all the existing network chains of direct and indirect contacts stemming from each player, a feature characteristic of Katz-Bonacich centrality.

We then consider different policies that aim at reducing the total crime activity in a delinquent network. The standard policy tool to reduce aggregate delinquency relies on the deterrence effects of punishment (Becker, 1968). By uniformly hardening the punishment costs borne by all delinquents, the distribution of delinquency efforts shifts to the left and the average (and aggregate) delinquency level decreases. This homogeneous policy tackles average behavior explicitly and does not discriminate among delinquents depending on their relative contribution to the aggregate delinquency level. To this "brute force" policy, we propose a targeted policy that discriminates among delinquents depending on their relative network location, and removes a few suitably selected targets from this network, alters the

[3] Due to Katz (1953) and extended by Bonacich (1987).

whole distribution of delinquency efforts, not just shifting it. To characterize the network optimal targets, we use a new measure of network centrality, the *intercentrality measure*, proposed by Ballester et al. (2006). This measure solves the planner's problem that consists in finding and getting rid of the *key player*, i.e., the delinquent who, once removed, leads to the highest aggregate delinquency reduction. We show that the key player is, precisely, the individual with the highest intercentrality in the network.

Using the AddHealth data of adolescents in the United States, we then test the results of our theoretical analysis. We first test whether or not there are peer effects in crime. While the potential benefits of leveraging social networks to reduce criminal behaviors are substantial, so too are the empirical difficulties of uncovering how social networks form, operate and the strength of network effects on outcomes. These difficulties are partly due to the lack of theoretical models that can help us understand the way these feedback effects operate. They are also due to the lack of network data, as well as to the fact that social networks are formed purposefully and connected individuals share environmental influences. These features of social networks complicate the estimation of causal impacts of networks and reduce the ability to suggest policies to reduce bad behaviors and encourage good behaviors. It is often difficult to disentangle whether the observation of two friends skipping school or smoking with other adolescents is due to both facing low punishment regimes, or because they influence each other to pursue risky behaviors, or because they choose to be friends based on their common interest in pursuing risky behaviors.

In order to suggest policies that can leverage social networks to reduce risky behaviors, researchers must be able to disentangle these mechanisms. For example, policy makers may want to increase randomly punishments, or target both friends simultaneously with interventions, or recruit one friend into an intervention program and rely on spillover effects to reduce both friends' bad behaviors, or seek to connect those who pursue risky behaviors with friends who do not pursue these behaviors. It is difficult to know which type of policy to suggest without knowing the mechanism underlying the observation that friends often make similar choices.

We tackle the econometric issues in the estimation of peer effects in crime by extending the recent method of Liu and Lee (2010). Using an instrumental variable approach as well as network fixed effects, we estimate the first-order conditions of our theoretical model to evaluate the intensity of peer effects as well as the role of centrality in crime. We find that a standard deviation increase in the aggregate level of delinquent activity of the peers translate into a roughly 11 percent increase of a standard deviation in the individual level of activity.

Finally, we test the second prediction of the theoretical model, the key player policy. We determine for each network the key player (i.e., the delinquent who, once removed, leads to the highest aggregate delinquency reduction), analyze his/her main characteristics and examine to what extent the Katz-Bonacich centrality of each individual is related to his/her intercentrality measure. Compared to other criminals, we find that key players are more likely to be a male, have less educated parents, are less attached to religion and feel socially more excluded. They also feel that adults care less about them, are less attached to their school and have more troubles getting along with the teachers. From our empirical analysis, we also find that, even though some criminals are not very active in criminal activities, they can be key players because they have a crucial position in the network in terms of betweenness centrality.

The rest of the paper unfolds as follows. In the next section, we discuss the related literature and explain our contribution. In Section 3, we present our theoretical framework, that is both the Nash equilibrium and the key-player policy. Our data are described in Section 4 while the estimation and empirical results of the impact of peer effects on crime are provided in Section 5. Section 6 details the empirical analysis of the key player and gives the results. Finally, in Section 7, we conclude and discuss some policy implications of our results.

2 Related literature

Theory There is a growing theoretical literature on the social aspects of crime. In Sah (1991), the social setting affects the individual perception of the costs of crime, and is thus conducive to a higher or a lower sense of impunity. In Glaeser et al. (1996), criminal interconnections act as a social multiplier on aggregate crime. Calvó-Armengol and Zenou (2004), Ballester et al. (2006, 2010), Patacchini and Zenou (2008, 2011) develop more general models by studying the effect of the structure of the social network on crime. They show that the location in the social network is crucial to understand crime and that not only direct friends but also friends of friends of friends, etc. have an impact of criminal activities and the decision to become a criminal.[4]

[4]Linking social interactions with crime has also been done in dynamic general equilibrium models (İmrohoroğlu et al., 2000, and Lochner 2004) and in search-theoretic frameworks (Burdett et al., 2004, and Huang et al., 2004). Other related contributions on the social aspects of crime include Silverman (2004), Verdier and Zenou (2004), Calvó-Armengol et al. (2007), Ferrer (2010).

Empirics There is a also growing empirical literature in economics suggesting that peer effects are very strong in criminal decisions. Case and Katz (1991), using data from the 1989 NBER survey of youths living in low-income Boston neighborhoods, find that the behaviors of neighborhood peers appear to substantially affect criminal activities of youth behaviors. They find that the direct effect of moving a youth with given family and personal characteristics to a neighborhood where 10 percent more of the youths are involved in crime than in his or her initial neighborhood is to raise the probability the youth will become involved in crime by 2.3 percent. Ludwig et al. (2001) and Kling et al. (2005) explore this last result by using data from the Moving to Opportunity (MTO) experiment that relocates families from high- to low-poverty neighborhoods. They find that this policy reduces juvenile arrests for violent offences by 30 to 50 percent for the control group. This also suggests very strong social interactions in crime behaviors. Patacchini and Zenou (2008, 2011) find that peer effects in crime are strong, especially for petty crimes.

Damm and Dustmann (2008) investigate the following question: Does growing up in a neighborhood in which a relatively high share of youth has committed crime increase the individual's probability of committing crime later on? To answer this question, Damm and Dustmann exploit a Danish natural experiment that randomly allocates parents of young children to neighborhoods with different shares of youth criminals. With area fixed effects, their key results are that one standard deviation increase in the share of youth criminals in the municipality of initial assignment increases the probability of being charge with an offense at the age 18-21 by 8 percentages point (or 23 percent) for men. This neighborhood crime effect is mainly driven by property crime. 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 with that crime.[5]

This paper's contributions Compared to these literatures, we have the following main contributions:

(i) We provide an explicit crime model where individuals are ex ante heterogeneous, derive the key-player policy and propose a simple model that can explain the link formation in our specific context;

[5]Building on the binary choice model of Brock and Durlauf (2001), Sirakaya (2006) identifies social interactions as the primary source of recidivist behavior in the United States.

(ii) We improve the identification strategy of peer effects proposed by Bramoullé et al. (2009) and Lee et al. (2010) by addressing the case of a non-row-normalized matrix of social interactions adopted from Liu and Lee (2010);

(iii) For both undirected and directed networks, we provide estimates of criminal outcomes that separate peer effects from contextual and correlated effects;

(iv) We contrast the importance of a weighted Katz-Bonacich centrality measure (i.e. most active criminals) and the intercentrality measure in criminal activities (i.e. key players);

(v) Using a counterfactual analysis, we identify the characteristics of the key player in each network of criminals in the AddHealth data, study the significant differences between key players and criminals and see if other measures of centrality can explain why some key players are not the most active criminals in a network.

3 Theoretical framework

3.1 The model

We develop a network model of peer effects, where the network reflects the collection of active bilateral influences.

The network $N_r = \{1, \ldots, n_r\}$ is a finite set of agents in network g_r ($r = 1, ..., \overline{r}$), where $\overline{r}$ is the total number of networks. We keep track of social connections by a delinquency network g_r, where $g_{ij,r} = 1$ if i and j are direct friends, and $g_{ij,r} = 0$, otherwise. Friendship are reciprocal so that $g_{ij,r} = g_{ji,r}$. All our results hold for non-symmetric networks but, for the ease of the presentation, we focus on symmetric networks in the theoretical model (which is more relevant for friendship networks). We also set $g_{ii,r} = 0$.[6]

Preferences[7] Delinquents in network g_r decide how much effort to exert. We denote by $y_{i,r}$ the delinquency effort level of delinquent i in network g_r and by $\mathbf{y}_r = (y_{1,r}, ..., y_{n,r})'$ the population delinquency profile in network g_r. Each agent i selects an effort $y_{i,r} \geq 0$, and obtains a payoff $u_{i,r}(\mathbf{y}_r, g_r)$ that depends on the effort profile $\mathbf{y}_r$ and on the underlying

[6] See Goyal (2007) and Jackson (2008) for overviews on network theory. See Ioannides and Loury (2004) for an overview on social networks and the labor market.

[7] Matrices and vectors are in bold while scalars are in normal letters.

network g_r, in the following way:

$$u_{i,r}(\mathbf{y}_r, g_r) = \underbrace{(a_{i,r} + \eta_r + \varepsilon_{i,r})\ y_{i,r}}_{\text{Proceeds}} - \underbrace{\frac{1}{2} y_{i,r}^2}_{\text{moral cost of crime}} - \underbrace{p_r\, f_r\, y_{i,r}}_{\text{cost of being caught}} + \underbrace{\phi \sum_{j=1}^{n} g_{ij,r} y_{i,r} y_{j,r}}_{\text{positive peer effects}} \quad (1)$$

where $\phi > 0$. This utility has a standard cost/benefit structure (as in Becker, 1968). The proceeds from crime are given by $(a_{i,r} + \eta_r + \varepsilon_{i,r})\ y_{i,r}$ and are increasing in own effort $y_{i,r}$. The costs of committing crime are captured by the probability to be caught $0 < p_r < 1$ times the fine $f_r\, y_{i,r}$, which increases with own effort $y_{i,r}$, as the severity of the punishment increases with one's involvement in crime. Also, as it now quite standard (see e.g. Verdier and Zenou, 2004; Conley and Wang, 2006), individuals have a *moral* cost of committing crime equals to $\frac{1}{2} y_{i,r}^2$, which is also increasing in own crime effort $y_{i,r}$. Finally, the new element in this utility function is the last term $\phi \sum_{j=1}^{n_r} g_{ij,r} y_{i,r} y_{j,r}$, which reflects the influence of friends' behavior on own action. The peer effect component can also be heterogeneous, and this *endogenous heterogeneity* reflects the different locations of individuals in the friendship network g_r and the resulting effort levels. More precisely, bilateral influences are captured by the following cross derivatives, for $i \neq j$:

$$\frac{\partial^2 u_{i,r}(\mathbf{y}_r, g_r)}{\partial y_{i,r} \partial y_{j,r}} = \phi g_{ij,r} \geq 0. \quad (2)$$

When i and j are direct friends, the cross derivative is $\phi > 0$ and reflects strategic complementarity in efforts. When i and j are not direct friends, this cross derivative is zero. In the context of crime, $\phi > 0$ means that if two students are friends, i.e. $g_{ij,r} = 1$, and if j increases her crime effort, then i will experience an increase in her (marginal) utility if she also increases her crime effort.

Let us now comment in more detail this utility function. In (1), η_r denotes the the unobservable network characteristics, e.g., the prosperous level of the neighborhood/network g_r (i.e. more prosperous neighborhoods lead to higher proceeds from crime) and $\varepsilon_{i,r}$ is an error term, meaning that there is some uncertainty in the proceeds from crime. Both η_r and $\varepsilon_{i,r}$ are observed by the delinquents but not by the econometrician. Also, in (1), $a_{i,r}$ denotes the *exogenous heterogeneity* that captures the *observable* differences between individuals. In this model, $a_{i,r}$ captures the fact that individuals differ in their ability (or productivity) of committing crime. Indeed, for a given effort level $y_{i,r}$, the higher $a_{i,r}$, the higher the productivity and thus the higher the proceeds from crime $a_{i,r} y_{i,r}$. Observe that $a_{i,r}$ is assumed to be deterministic, perfectly *observable* by all individuals in the network and corresponds to the observable characteristics of individual i (e.g. sex, race, age, parental education, etc.)

To summarize, the utility function can be written as:

$$u_{i,r}(\mathbf{y}_r, g_r) = [a_{i,r} - p_r\, f_r + \eta_r + \varepsilon_{i,r}]\; y_{i,r} - \frac{1}{2}y_{i,r}^2 + \phi \sum_{j=1}^{n_r} g_{ij,r}y_{i,r}y_{j,r}$$

So when a delinquent i exerts some effort in crime, the proceeds from crime depends on ability $a_{i,r}$, the expected marginal cost of being caught $p_r f_r$, how prosperous is the neighborhood/network η_r and on some random element $\varepsilon_{i,r}$, which is specific to individual i. In other words, $a_{i,r}$ is the observable part (by the econometrician) of i's characteristics while $\varepsilon_{i,r}$ captures the unobservable characteristics of individual i. Note that the utility (1) is concave in own decisions, and displays decreasing marginal returns in own effort levels.

From now on, since we focus only on one network, when there is ambiguity we will drop the subscript r in the theoretical section.

The Bonacich network centrality To each network g, we associate its adjacency matrix $\mathbf{G} = [g_{ij}]$. This is a symmetric zero-diagonal square matrix that keeps track of the direct connections in g.

The kth power $\mathbf{G}^k = \mathbf{G}^{(k \overset{times}{\ldots})}\mathbf{G}$ of the adjacency matrix $\mathbf{G}$ keeps track of indirect connections in g. More precisely, the coefficient $g_{ij}^{[k]}$ in the (i,j) cell of $\mathbf{G}^k$ gives the number of paths of length k in g between i and j. In particular, $\mathbf{G}^0 = \mathbf{I}$. Note that, by definition, a path between i and j needs not to follow the shortest possible route between those agents. For instance, when $g_{ij} = 1$, the sequence $ij \rightarrow ji \rightarrow ij$ constitutes a path of length three in g between i and j.

Definition 1 (Katz, 1953; Bonacich, 1987) *Given a vector* $\mathbf{u} \in \mathbb{R}_+^n$, *and* $\phi \geq 0$ *a small enough scalar, the vector of Bonacich centralities of parameter* ϕ *in network* g *is defined as:*

$$\mathbf{b}_\mathbf{u}(g,\phi) = (\mathbf{I} - \phi\mathbf{G})^{-1}\,\mathbf{u} = \sum_{p=0}^{+\infty}\phi^p\mathbf{G}^p\mathbf{u}. \tag{3}$$

Nash equilibrium

We now characterize the Nash equilibrium of the game where agents choose their effort level $y_i \geq 0$ simultaneously. At equilibrium, each agent maximizes her utility (1). The corresponding first-order conditions are:

$$\frac{\partial u_i(\mathbf{y},g)}{\partial y_i} = a_i - pf + \eta + \varepsilon_i - y_i + \phi\sum_{j=1}^{n_\kappa} g_{ij}y_j = 0.$$

Therefore, we obtain the following best-reply function for each $i = 1, ..., n$:

$$y_i = \phi \sum_{j=1}^{n} g_{ij} y_j + a_i - pf + \eta + \varepsilon_i, \tag{4}$$

Denote by $\mu_1(\mathbf{G})$ the spectral radius of $\mathbf{G}$, by $\alpha_i = a_i - pf + \eta + \varepsilon_i$, with corresponding non-negative vector $\boldsymbol{\alpha}$, we have:

Proposition 1 *If $\phi\mu_1(\mathbf{G}) < 1$, the peer effect game with payoffs (1) has a unique Nash equilibrium in pure strategies given by:*

$$\mathbf{y}^* = \mathbf{b}_{\boldsymbol{\alpha}}(g, \phi) \tag{5}$$

Proof. Apply Theorem 1, part b, in Calvó-Armengol et al. (2009) to our problem. ∎

This results shows that the Bonacich centrality is the right network index to account for equilibrium behavior when the utility functions are linear-quadratic. In (1), the local payoff interdependence is restricted to direct network contacts. At equilibrium, though, this local payoff interdependence spreads all over the network through the overlap of direct friendship clusters. The Bonacich centrality precisely reflects how individual decisions feed into each other along any direct and indirect network path. Furthermore, the condition $\phi\mu_1(\mathbf{G}) < 1$ stipulates that local complementarities must be small enough than own concavity, which prevents multiple equilibria to emerge and, in the same time, rules out corner solutions (i.e., negative or zero solutions).[8] This condition also guarantees that $(\mathbf{I} - \phi\mathbf{G})$ is invertible and its series expansion well defined. Observe that

$$\mathbf{b}_{\boldsymbol{\alpha}}(g, \phi) = (\mathbf{I} - \phi\mathbf{G})^{-1}\boldsymbol{\alpha} = \sum_{p=0}^{+\infty} \phi^p \mathbf{G}^p \boldsymbol{\alpha} \tag{6}$$

where

$$\boldsymbol{\alpha} = \mathbf{a} + \boldsymbol{\varepsilon} - pf\mathbf{l_n} + \eta\mathbf{l_n}$$

and where $\mathbf{l_n}$ is an n-dimensional vector of ones. In particular, Proposition 1 states that for each delinquent i, we have:

$$y_i^* = b_{\alpha,i}(g, \phi)$$

3.2 Finding the key player

We would like now to expose the "key player" policy. The planner's objective to find the key player is to generate the highest possible reduction in aggregate delinquency level by picking

[8] See Ballester et al. (2006) for a formal proof of this result.

the appropriate delinquent. Formally, the planner's problem is the following:

$$\max\{y^*(g) - y^*(g^{[-i]}) \mid i = 1, ..., n\},$$

where $y^*(g) = \sum_i y_i^*(g)$ is the total level of crime in network g and $g^{[-i]}$ is network g without individual i. When the original delinquency network g is fixed, this is equivalent to:

$$\min\{y^*(g^{[-i]}) \mid i = 1, ..., n\} \tag{7}$$

From Ballester et al. (2006), we now define a new network centrality measure $d(g, \phi)$ that will happen to solve this compromise. Define $\mathbf{M}(g, \phi) = (\mathbf{I} - \phi\mathbf{G})^{-1}$ a non-negative matrix. Its coefficients $m_{ij}(g, \phi) = \sum_{k=0}^{+\infty} \phi^k g_{ij}^k$ count the number of walks in g starting from i and ending at j, where walks of length k are weighted by ϕ^k.

The Bonacich centrality of node i is $b_{\alpha,i}(g, \phi) = \sum_{j=1}^{n} \alpha_j m_{ij}(g, \phi)$, and counts the *total* number of paths in g starting from i weighted by the α_j of each linked node j.

Let $b_{\alpha,i}(g, \phi)$ be the centrality of i in network g, $b_\alpha(g, \phi)$ the *total* centrality in network g (i.e. $b_\alpha(g, \phi) = \sum_{i=1}^{n} b_{\alpha,i}(g, \phi)$) and $b_\alpha(g^{[-i]}, \phi)$ the *total* centrality in $g^{[-i]}$.

Definition 2 *For all networks g and for all i, the intercentrality measure of delinquent i is:*

$$d_i(g, \phi) = b_\alpha(g, \phi) - b_\alpha(g^{[-i]}, \phi) = \frac{b_{\alpha,i}(g, \phi) \sum_{j=1}^{n} m_{ji}(g, \phi)}{m_{ii}(g, \phi)} \tag{8}$$

Proof. Apply Lemma 1 in Ballester et al. (2006) to this problem.

Observe that, in (8), $b_{\alpha,i}(g, \phi)$ is the *weighted* Bonacich (out−) centrality of delinquent i where the weights are in terms of the α's, $\sum_{j=1}^{j=n} m_{ji}(g, \phi)$ is the *unweighted* (in−) centrality of player delinquent i (i.e. it counts the total number of paths in g that end at i) and $m_{ii}(g, \phi)$ is *unweighted* and counts the total number of paths in g from i to i itself where walks of length k are weighted by ϕ^k.

The intercentrality measure $d_i(g, \phi)$ of delinquent i is the sum of i's centrality measures in g, and i's contribution to the centrality measure of every other delinquent $j \neq i$ also in g. It accounts both for one's exposure to the rest of the group and for one's contribution to every other exposure.

The following result establishes that intercentrality captures, in an meaningful way, the two dimensions of the removal of a delinquent from a network, namely, the direct effect on delinquency and the indirect effect on others' delinquency involvement.[9]

[9] As in Ballester et al. (2010), we could also identify a *key group* that reduces the most aggregate delinquency in each network by characterizing the *optimal group removal* from the network. Because in the empirical analysis our networks have relatively small sizes (see Section 4), the key group policy is less relevant and, therefore, we will mainly focus on the key player policy.

Proposition 2 *A player i^* is the key player that solves (7) if and only if i^* is a delinquent with the highest intercentrality in g, that is, $d_{i^*}(g, \phi) \geq d_i(g, \phi)$, for all $i = 1, ..., n$.*

Proof. Theorem 3 in Ballester et al. (2006).

Observe that this result is true for both *undirected* networks (*symmetric* adjacency matrix) and *directed* networks (*asymmetric* adjacency matrix). It is also true for adjacency matrices with weights (i.e. values different than 0 and 1) and self-loops (delinquents have a link with themselves).

To illustrate Proposition 2, consider the following symmetric undirected network with four delinquents (i.e. $n = 4$):

Figure 1: A network with 4 criminals

The adjacency matrix is then given by:

$$\mathbf{G} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

We assume $\phi = 0.3$.[10] and that $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.1, 0.2, 0.3, 0.4)$. It is then straightforward to see that, using Proposition 1, we obtain:

$$\begin{pmatrix} y_1^* \\ y_2^* \\ y_3^* \\ y_4^* \end{pmatrix} = \begin{pmatrix} b_{\alpha,1}(g, \phi) \\ b_{\alpha,2}(g, \phi) \\ b_{\alpha,3}(g, \phi) \\ b_{\alpha,4}(g, \phi) \end{pmatrix} = \begin{pmatrix} 0.66521 \\ 0.60377 \\ 0.68068 \\ 0.59958 \end{pmatrix}$$

[10]The spectral radius of this graph is: 2.17 and thus the condition $\phi\mu_1(\mathbf{G}) < 1$ is satisfied since $2.17 \times 0.3 = 0.651 < 1$.

so that the total activity level is given by:

$$y^*(g) = y_1^*(g) + y_2^*(g) + y_3^*(g) + y_4^*(g) = b_\alpha(g, \phi) = 2.549$$

Individual 3 has the highest weighted Bonacich and thus provides the highest crime effort. If we look at the formula in Definition 2, it says that the delinquent that the planner wants to remove is:

$$d_{i^*}(g, \phi) = b_\alpha(g, \phi) - b_\alpha(g^{[-i]}, \phi)$$

Let us remove delinquent 1. The network becomes:

Figure 2: The network when criminal 1 has been removed

Using the same decay factor, $\phi = 0.3$, we obtain:

$$\begin{pmatrix} y_2^* \\ y_3^* \\ y_4^* \end{pmatrix} = \begin{pmatrix} b_{\alpha,2}(g^{[-1]}, \phi) \\ b_{\alpha,3}(g^{[-1]}, \phi) \\ b_{\alpha,4}(g^{[-1]}, \phi) \end{pmatrix} = \begin{pmatrix} 0.31868 \\ 0.3956 \\ 0.4 \end{pmatrix}$$

so that the total effort is now given by:

$$y^*(g^{[-1]}) = y_2^*(g^{[-1]}) + y_3^*(g^{[-1]}) + y_4^*(g^{[-1]}) = b_\alpha(g^{[-1]}, \phi) = 1.114$$

Thus, player 1's contribution is

$$b_\alpha(g, \phi) - b_\alpha(g^{[-1]}, \phi) = 2.549 - 1.114 = 1.435 \tag{9}$$

Doing the similar exercise for individuals 2, 3, 4, we obtain:

$$b_\alpha(g, \phi) - b_\alpha(g^{[-2]}, \phi) = 1.244$$

$$b_\alpha(g, \phi) - b_\alpha(g^{[-3]}, \phi) = 1.146$$

$$b_\alpha(g,\phi) - b_\alpha(g^{[-4]},\phi) = 0.988$$

Criminal 1 is the key player since his/her contribution to total crime is the highest one.

Let us now check if the formula (8) works, i.e.,

$$d_1(g,\phi) = b_\alpha(g,\phi) - b_\alpha(g^{[-1]},\phi) = 1.435$$

From (8), we have:

$$d_1(g,\phi) = \frac{b_{\alpha,1}(g,\phi)\sum_{j=1}^{4} m_{j1}(g,\phi)}{m_{11}(g,\phi)}$$

Let us go back to the initial network with four individuals. It is easily verified that (with $\phi = 0.3$):

$$\mathbf{M} = (\mathbf{I} - \phi\mathbf{G})^{-1} = \begin{pmatrix} 1.5317 & 0.65646 & 0.65646 & 0.45952 \\ 0.65646 & 1.3802 & 0.61101 & 0.19694 \\ 0.65646 & 0.61101 & 1.3802 & 0.19694 \\ 0.45952 & 0.19694 & 0.19694 & 1.1379 \end{pmatrix}$$

so that

$$m_{11}(g,\phi) = 1.5317$$

and

$$\begin{aligned} \sum_{j=1}^{4} m_{j1}(g,\phi) &= m_{11}(g,\phi) + m_{21}(g,\phi) + m_{31}(g,\phi) + m_{41}(g,\phi) \\ &= 1.5317 + 0.65646 + 0.65646 + 0.45952 \\ &= 3.3041 \end{aligned}$$

Therefore,

$$\begin{aligned} d_1(g,\phi) &= \frac{b_{\alpha,1}\sum_{j=1}^{3} m_{j1}(g,\phi)}{m_{11}(g,\phi)} \\ &= \frac{0.66521 \times 3.3041}{1.5317} \\ &= 1.435 \end{aligned} \tag{10}$$

When comparing (9) and (10), we see that the values are the same and thus:

$$d_1(g,\phi) = b_\alpha(g,\phi) - b_\alpha(g^{[-1]},\phi) = 1.435$$

3.3 The invariant assumption on $g^{[-i]}$: Theoretical issues

In our theoretical framework, when the key player i is removed from network g, the remaining network becomes $g^{[-i]}$ where the ith row and ith column in $\mathbf{G}$ has been removed. In other words, we have an *invariant assumption* on the reduced network $g^{[-i]}$, i.e. we assume that, when the key player is removed, the other criminals in the network do not form new links. Also $\mathbf{G}$ is exogenous, which means that $\mathbf{G}$ is not correlated with the error term ϵ. However, in our framework, $\mathbf{G}$ is allowed to be correlated with $\mathbf{x}$ ($\mathbf{x} = (x_1, \cdots, x_n)'$ is a vector of individuals' characteristics) and the network-specific fixed effect η. The invariant assumption can be justified by using some models of network formation. The formation of links $\mathbf{G} = [g_{ij}]$ can depend on $\mathbf{x}$ in the following way:

$$\begin{aligned} P_{ij} &= f(x_i, x_j) + e_{ij}, \\ g_{ij} &= \begin{cases} 1 & \text{if } P_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}, \end{aligned}$$

where P_{ij} is the propensity to form link ij, $f(x_i, x_j)$ is a function of x_i and x_j (where x_i' and x_j' are the ith and jth rows of $\mathbf{x}$) and e_{ij} is an error term. A possible parametric specification of $f(x_i, x_j)$ can be $f(x_i, x_j) = a + b|x_i - x_j|$. If the estimated b is negative, it implies a link is likely to form between i and j if they share similar observable characteristics (say, family, income, etc.).

The proposed key player theory, i.e., the invariant property of $g^{[-i]}$, holds if this network formation process is at work so that the link of i and j depends only on the characteristics of individuals i and j, but not on others such as a $k \neq i, j$. In this model, the formation of a link is based on mutual consent (as in Jackson and Wolinsky, 1996) and is not affected by other individuals in the network. In other words, *each link formed by two individuals only depends on the characteristics of these two individuals but not on any other one.* Indeed, when a key player i is removed, all his/her links are also removed, but since the formation of link is created pairwise there is no reason for the remaining individuals to create new links. They would have done it before. As a result, the invariant assumption of $\mathbf{G}$ is justified in this framework. This way of modelling link formation would correspond to what Bramoullé and Fortin (2009) called *pairwise independent link formulation*, i.e. separable utility framework in pairs.[11] As a result, in the case of pairwise independence, the invariance property of $\mathbf{G}$ could be justified by this setting of utility. We will provide a diagnostic check of this model

[11] For a directed network, this means that $u_i(g) = \sum_j v_i(g_{ij})$. If the network is undirected, one needs to impose an additional symmetry assumption (Bramoullé and Fortin, 2009).

in Section 6.2 below.

3.4 Is the key player always the more active criminal?

Definition 2 specifies a clear relationship between the Bonacich centrality and the inter-centrality measures. Holding $b_i(g, \phi)$ fixed, the intercentrality $d_i(g, \phi)$ of player i decreases with $m_{ii}(g, \phi)$ of i's Bonacich centrality due to self-loops, and increases with the fraction of i's centrality amenable to out-walks. As a result, it should be clear from Definition 2 that the key player is very likely to be the criminal with the highest Bonacich centrality (i.e. the most active criminal in the network) but not necessary. In the example provided in Section 3.2, the key player was criminal 1 but was not the most active criminal, i.e. the criminal with the highest Bonacich centrality. Criminal 3 was in fact the most active criminal. The result was mainly due to the fact that, ex ante, criminal had a higher heterogeneity than criminal 1, i.e., $\alpha_3 = 0.3 > 0.1 = \alpha_1$. We would like now to provide an example where, even if the αs are identical for all individuals, there can still be key players (highest intercentrality measures) who are not the most active criminals (highest Katz-Bonacich centrality measures).

Consider the network g in the following figure with eleven criminals.

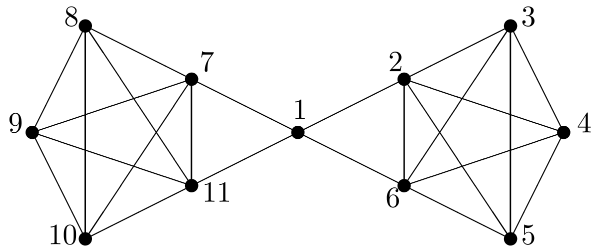

Figure 3: A bridge network with 11 criminals

We distinguish three different types of equivalent actors in this network, which are the following:

Type	Criminals
1	1
2	2, 6, 7 and 11
3	3, 4, 5, 8, 9 and 10

From a macro-structural perspective, type−1 and type−3 criminals are identical: they all have four direct links, while type −2 criminals have five direct links each. From a micro-structural perspective, though, criminal 1 plays a critical role by bridging together two closed-knit (fully intraconnected) communities of five criminal each. By removing delinquent 1, the network is maximally disrupted as these two communities become totally disconnected, while by removing any of the type-2 criminals, the resulting network has the lowest aggregate number of network links.

We identify the key player in this network of criminals. If the choice of the key player were solely governed by the *direct* effect of criminal removal on aggregate crime, type-2 criminals would be the natural candidates. Indeed, these are the ones with the highest number of direct connections. But the choice of the key player needs also to take into account the *indirect* effect on aggregate delinquency reduction induced by the network restructuring that follows the removal of one delinquent from the original network. Because of his communities' bridging role, criminal 1 is also a possible candidate for the preferred policy target.

In order to focus on the role of location in the network, in this example, we assume that criminals are ex identical so that $\boldsymbol{\alpha} = \mathbf{l_n}$ and thus $\mathbf{b_1}(g, \phi) = (\mathbf{I} - \phi \mathbf{G})^{-1} \mathbf{l_n}$ and $y_i^* = b_i(g, \phi)$[12] while $d_i(g, \phi) = b_i(g, \phi) - b_i(g^{[-i]}, \phi)$. We take $\phi = 0.2$. The following table computes, for criminals of types 1, 2 and 3, the value of delinquency centrality measures $b_i(g, \phi)$ (or equivalently efforts y_i^*) and intercentrality measures $d_i(g, \phi)$ for different values of ϕ. In each column, a variable with a star identifies the highest value.[13]

Table 1a: Key player versus Bonacich centrality in a bridge network

Player Type	1	2	3
$y_i = b_i$	8.33	9.17*	7.78
d_i	41.67*	40.33	32.67

First note that type-2 delinquents display the highest Bonacich centrality measure. These delinquents have the highest number of direct connections. Besides, they are directly connected to the bridge delinquent 1, which gives them access to a very wide and diversified span of indirect connections. Altogether, they are the most central delinquents (in terms of Bonacich centrality). Second, the most active delinquents are not the key players. Because indirect effects matter a lot ($\phi = 0.2$), eliminating delinquent 1 has the highest joint direct and indirect effect on aggregate delinquency reduction. Indeed, when ϕ is not too low, delinquents spread their know-how further away in the network and establishing synergies with delinquents located in distant parts of the social setting. In this case, the optimal

[12] Since all individuals have the same $\alpha = 1$, we denote the total crime in the network by $b_1(g, \phi)$ instead of $b_{1,1}(g, \phi)$.

[13] We can compute the highest possible value for ϕ compatible with our definition of centrality measure (i.e. the inverse of the largest eigenvalue of G), which is equal to $\widehat{\phi} = \frac{2}{3+\sqrt{41}} \simeq 0,213$.

targeted policy is the one that maximally disrupts the delinquency network, thus harming the most its know-how transferring ability.

In Table 1a, we have shown that the key player is not the most active criminal (i.e. does have the highest Bonacich centrality). To further understand this result, let us analyze the characteristics of all criminals in terms of network position, as well as those of the network described in Figure 3. For that, we will first use some measures of centrality other than Bonacich. Indeed, over the past years, social network theorists have proposed a number of centrality measures to account for the variability in network location across agents (Wasserman and Faust, 1994).[14] While these measures are mainly geometric in nature, our theory provides a behavioral foundation to the Bonacich centrality measure (and only this one) that coincides with the unique Nash equilibrium of a non-cooperative peer effects game on a social network. Let us now calculate for the network given in Figure 3 the other individual centrality measures, namely: degree, closeness, betweenness centralities as well as the clustering coefficient. Their mathematical definitions are given in Appendix 4. We obtain:

Table 1b: Characteristics of criminals in a network where the most active criminal is not the key player

Player type	1	2	3
Degree centrality	0.4	0.5	0.4
Closeness centrality	0.625	0.555	0.416
Betweenness centrality	0.555	0.2	0
Clustering coefficient	0.33	0.7	1

Even if player 1 is not the most active criminal (she has the lowest degree centrality and the lowest clustering coefficient), it is now even easier to understand why she is the key player: she has the highest closeness and betweenness centralities. Observe that criminal 3 has a betweenness centrality equals to zero because there are no shortest path between two criminals that go through her.

Let us now examine the characteristics of the network described in Figure 3 where the key player is not the most active criminal. We will consider standard network characteristics, which are all defined in Appendix 4. We obtain the following results:

[14] See Borgatti (2003) for a discussion on the lack of a systematic criterium to pick up the "right" network centrality measure for each particular situation.

Table 1c: Characteristics of the network
in which the most active criminal is not the key player

Network Characteristics	
Average Distance	2.11
Average Degree	4.36
Diameter	4
Density	0.211
Asymmetry	0.125
Clustering	0.805
Degree centrality	7.78×10^{-3}
Closeness centrality	0.323
Betweenness Centrality	0.47556
Assortativity	-3.49×10^{-16}

We see from Table 1c that the network described in Figure 3 has a low average distance and low diameter (small-world properties), a very high clustering (0.805) and a weak dissortativity. Furthermore, it is not very dense nor asymmetric while having average values of centralities measures.

To summarize, the individual Nash equilibrium efforts of the delinquency-network game are proportional to the *equilibrium Bonacich centrality* network measures, while the key player is the delinquent with the highest *intercentrality measure*. As the previous example illustrates, these two measures need not to coincide. This is not surprising, as both measures differ substantially in their foundation. Whereas the equilibrium-Bonacich centrality index derives from strategic individual considerations, the intercentrality measure solves the planner's optimality collective concerns. In particular, the equilibrium Bonacich centrality measure fails to internalize all the network payoff externalities delinquents exert on each other, while the intercentrality measure internalizes them all. More formally, the measure $d_\alpha(g, \phi)$ goes beyond the measure $\mathbf{b}_{\boldsymbol{\alpha}}(g, \phi)$ by keeping track of all the cross-contributions that arise between its coordinates $b_{\alpha,1}(g, \phi), ..., b_{\alpha,n}(g, \phi)$.

4 Data description

Our analysis is made possible by the use of a unique database on friendship networks from the National Longitudinal Survey of Adolescent Health (AddHealth).[15]

The AddHealth 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 pupil attending the sampled schools on the interview day is asked to compile a questionnaire (in-school data) containing questions on respondents' demographic and behavioral characteristics, education, family background and friendship. 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 compile a longer questionnaire containing more sensitive individual and household information (in-home and parental data). Those subjects of the subset are interviewed again in 1995–96 (wave II), in 2001–2 (wave III), and again in 2007-2008 (wave IV).[16] For the purposes of our analysis, we focus on wave I because the network information is only available in the first wave.

From a network perspective, the most interesting aspect of the AddHealth data is the information on friendships. Indeed, the friendship information is based upon actual friends nominations. Pupils were asked to identify their best friends from a school roster (up to five males and five females).[17] We assume that friendship relationships are reciprocal, i.e. a link exists between two friends if at least one of the two individuals has identified the other as his/her best friend.[18] By matching the identification numbers of the friendship nominations to respondents' identification numbers, one can obtain information on the characteristics of nominated friends. More importantly, one can reconstruct the whole geometric structure

[15]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.

[16]The AddHealth website describes survey design and data in details. http://www.cpc.unc.edu/projects/addhealth

[17]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.

[18]We considered non-reciprocal friendship networks below.

of the friendship networks. For each school, we thus obtain all the network components of (best) friends.[19]

The in-home questionnaire contains an extensive set of questions on juvenile delinquency, that are used to construct our dependent variable. Specifically, the AddHealth contains information on 15 delinquency items.[20] The survey asks students how often they participate in each of these activities during the past year.[21] Each response is coded using an ordinal scale ranging from 0 (i.e. never participate) to 1 (i.e. participate 1 or 2 times), 2 (participate 3 or 4 times) up to 3 (i.e. participate 5 or more times). To derive quantitative information on a topic using qualitative answers to a battery of related questions, we calculate an index of delinquency involvement for each respondent.[22] It ranges between 0.09 and 9.63, with mean equal to 0.94 and standard deviation to 1.09.

Because of the theoretical model (Section 3), we focus only on networks of delinquents, thus excluding the individuals who report never participating in any delinquent activity (roughly 40% of the total). Also, we do not consider networks at the extremes of the network size distribution to avoid the possibility that in these edge networks the strength of peer effects as well as the removal of the key player can have extreme values (too low or too high) that may be a matter of concern. Excluding individuals with non valid information, we obtain a final sample of 1,297 criminals distributed over 150 networks. The minimum number of individuals in a delinquent network is 4 while its maximum is 77. The mean and the standard deviation of network size are roughly 9 and 12 pupils, respectively.[23]

[19] Note that, when an individual i identifies a best friend j who does not belong to the same school, the database does not include j in the network of i; it provides no information about j. Fortunately, in the large majority of cases (more than 93%), best friends tend to be in the same school and thus are systematically included in the network.

[20] Namely, paint graffiti or signs on someone else's property or in a public place; deliberately damage property that didn't belong to you; lie to your parents or guardians about where you had been or whom you were with; take something from a store without paying for it; get into a serious physical fight; hurt someone badly enough to need bandages or care from a doctor or nurse; run away from home; drive a car without its owner's permission; steal something worth more than $50; go into a house or building to steal something; use or threaten to use a weapon to get something from someone; sell marijuana or other drugs; steal something worth less than $50; take part in a fight where a group of your friends was against another group; act loud, rowdy, or unruly in a public place.

[21] Respondents listened to pre-recorded questions through earphones and then they entered their answers directly on laptop computers. This administration of the survey for sensitive topics minimizes the potential for interview and parental influence, while maintaining data security.

[22] This is a standard factor analysis, where the factor loadings of the different variables are used to derive the total score.

[23] On average, delinquents declare having 2.26 delinquent friends with a standard deviation of 1.52.

Table A.1 in Appendix 1 provides the descriptive statistics and definitions of the variables used in our study.[24] Among the adolescents selected in our sample of delinquents, 32% are female and 19% are blacks. An average, criminal adolescent feel that adults care about them but have some trouble getting along with the teachers. Slightly less than 70% of our adolescents live in a household with two married parents, although about 30% come from a single parent family. The most popular occupation of the father is a manual one (roughly 30%) and 17% of them have parents who works in a professional/technical occupation. The average parental education is high school graduate. Almost 40% of our adolescents live in suburban areas. The performance at school, as measured by the mean mathematics score is slightly above the average. On average, our criminals consider themselves slightly more intelligent than their peers and their level of physical development appear to be slightly higher when compared to other boys/girls of the same age.[25] Our analysis in the following sections will shed lights on the characteristics of the most harmful individuals, that is on those pupils that, if removed, would lead to the highest crime reduction in their own groups.

5 Peer effects and network centrality

Let us now begin the test of our theoretical framework (Section 3) by providing an appropriate estimate of peer effects in crime ($\widehat{\phi}$). We first present our empirical model and estimation strategy. We use the architecture of networks to identify peer effects as described in Bramoullé et al. (2009) and Lee et al. (2010) but we consider the case of a non-row-normalized $\mathbf{G}$ and we highlight the methodological improvements that are achieved in our context. Our estimation method follows the 2SLS and GMM strategies proposed by Lee (2007) and refined by Liu and Lee (2010) to capture the impact of centrality in networks. To be more specific, we will begin by explaining the empirical issues than hinder the identification of peer effects and show to what extent it is possible to tackle each of these issues with the AddHealth dataset.

[24] Information at the school level, such as school quality and teacher/pupil ratio, is unnecessary given our fixed effects estimation strategy.

[25] When reading these summary information, one need to keep in mind that we deal here with juvenile delinquency, where some of the offences recorded as crimes (such as paint graffiti or lie to the parents) are quite minor.

5.1 Empirical model

Let $\bar{r}$ be the total number of networks in the sample (150 in our dataset), n_r be the number of individuals in the rth network, and $n = \sum_{r=1}^{\bar{r}} n_r$ be the total number of sample observations. Defining the ex ante heterogeneity $a_{i,r}$ of each individual in network r as

$$a_{i,r} = x'_{i,r}\beta + \frac{1}{g_{i,r}} \sum_{j=1}^{n_r} g_{ij,r} x'_{j,r}\gamma,$$

the empirical model corresponding to (4) can be written as:

$$y_{i,r} = \phi \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + x'_{i,r}\beta + \frac{1}{g_{i,r}} \sum_{j=1}^{n_r} g_{ij,r} x'_{j,r}\gamma + \eta_r^* + \epsilon_{i,r}, \tag{11}$$

for $i = 1, \cdots, n_r$ and $r = 1, \cdots, \bar{r}$, where $x_{i,r} = (x_{i,r}^1, \cdots, x_{i,r}^m)'$, $\eta_r^* = \eta_r - pf$, $g_{i,r} = \sum_{j=1}^{n_r} g_{ij,r}$ and $\epsilon_{i,r}$'s are i.i.d. innovations with zero mean and variance σ^2 for all i and r.

5.2 Identification strategy

The identification of peer effects (ϕ in model (11)) raises different challenges.

Reflection problem In linear-in-means models, simultaneity in behavior of interacting agents introduces a perfect collinearity between the expected mean outcome of the group and its mean characteristics. Therefore, it is difficult to differentiate between the effect of peers' choice of effort and peers' characteristics that do impact on their effort choice (the so-called *reflection problem*; see Manski, 1993). Basically, the reflection problem arises because, in the standard approach, individuals interact in groups, that is individuals are affected by all individuals belonging to their group and by nobody outside the group. In other words, groups completely overlap. In the case of social networks, instead, this is nearly never true since the reference group has individual-level variation. Take individuals i and k such that $g_{ik} = 1$. Then, individual i is directly influenced by $\sum_{j=1}^{n} g_{ij} y_j$ while individual k is directly influenced by $\sum_{j=1}^{n} g_{kj} y_j$, and there is little chance for these two values to be the same unless the network is complete (i.e. everybody is linked with everybody). Formally, as shown by Bramoullé et al. (2009), social effects are identified (i.e. no reflection problem) if $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{G}^2$ are linearly independent, where $\mathbf{G}^2$ keeps track of indirect connections of length 2 in g.[26]

[26] For example, complete networks do not satisfy this condition. In our dataset, where 150 networks are considered (see above in the data section), many of them have different sizes but none of them are complete and all satisfy the condition that guarantees the identification of social effects. Note that, even when networks are all complete, Lee (2007) shows that identification can be achieved by exploring strengths of interactions across networks of different sizes.

In other words, if i and j are friends and j and k are friends, it does not necessarily imply that i and k are also friends. Because of these intransitivities, $\mathbf{G}^2\mathbf{x}$, $\mathbf{G}^3\mathbf{x}$, etc. are not collinear with $\mathbf{Gx}$ and they act as valid instruments for $\mathbf{Gy}$ (under the situation that $\mathbf{x}$ is relevant). Intuitively, $\mathbf{G}^2\mathbf{x}$ represents the vector of the friends' friends attributes of each agent in the network. The architecture of social networks implies that these attributes will affect her outcome only through their effect on her friends' outcomes. Even in linear-in-means models the Manski's (1993) reflection problem is thus eluded.[27] Peer effects in social networks are thus identified and can be estimated using 2SLS (Lee 2007; Lin, 2010). In Appendix 2 we detail in a more technical way the identification of model (11). In particular, we highlight the difference between the case with row-normalized $\mathbf{G}$ (Bramoullé et al., 2009) and our case with non-row-normalized $\mathbf{G}$.

Endogenous network formation/correlated effects Although this setting allows us to solve the reflection problem, the estimation results might still be flawed because of the presence of unobservable factors affecting both individual and peer behavior. It is indeed difficult to disentangle the endogenous peer effects from the correlated effects, i.e. from effects arising from the fact that individuals in the same network tend to behave similarly because they face a common environment. If individuals are not randomly assigned into networks, this problem might originate from the possible sorting of agents. If the variables that drive this process of selection are not fully observable, potential correlations between (unobserved) network-specific factors and the target regressors are major sources of bias. Observe that our particularly large information on individual (observed) variables should reasonably explain the process of selection into groups. However, a number of papers have treated the estimation of peer effects with correlated effects (e.g., Clark and Loheac 2007; Lee 2007; Lin 2010; Lee et al. 2010). This approach is based on the use of *network fixed effects* and extends Lee (2003) 2SLS methodology after the removal of network fixed effects. Network fixed effects can be interpreted as originating from a two-step model of link formation where agents self-select into different networks in a first step with selection bias due to specific network characteristics and, then, in a second step, link formation takes place within networks based on observable individual characteristics only. An estimation procedure alike to a panel within group estimator is thus able to control for these correlated effects. One can get rid of the

[27] These results are formally derived in Bramoullé et al. (2009) (see, in particular, their Proposition 3) and used in Calvó-Armengol et al. (2009) and Lin (2010). Cohen-Cole (2006) presents a similar argument, i.e. the use of out-group effects, to achieve the identification of the endogenous group effect in the linear-in-means model (see also Weinberg et al., 2004; Laschever, 2009). See Durlauf and Ioannides (2010) and Blume et al. (2011) for an overview on these issues.

network fixed effects by subtracting the network average from the individual-level variables.[28] As detailed in the next section, this paper follows this approach.

Specific individual and contextual effects In this respect, the richness of the information provided by the AddHealth questionnaire on adolescents' behavior allow us to find observable individual variables as well as proxies for typically unobserved individual characteristics that may be correlated with our variable of interest. Specifically, to control for differences in leadership propensity across adolescents, we include an indicator of self-esteem and an indicator of the level of physical development compared to the peers, and we use mathematics score as an indicator of ability. Also, we attempt to capture differences in attitude towards education, parenting and more general social influences by including indicators of the student's school attachment, relationship with teachers, parental care and social inclusion.

To summarize, our identification strategy is based on the assumption that any troubling source of heterogeneity (if any), which is left unexplained by our unusually large set of observed characteristics can be captured at the network level, and thus taken into account by the inclusion of network fixed effects.

To be more precise, we allow link formation (as captured by our matrix $\mathbf{G}$) to be correlated with observed individual characteristics,[29] contextual effects ($\mathbf{G}^*\mathbf{x}$, where $\mathbf{G}^*$ is row-normalized from $\mathbf{G}$) and unobserved network characteristics (captured by the network fixed effects). The presence of other remaining unobserved effect is very unlikely in our case given our set of controls that includes behavioral factors and, most importantly, because we deal with quite small networks (see Section 4).

Deterrence effects So far, we have dealt with issues that are common to the identification of any kind of peer effects. There is, however, something that is specific to crime: How deterrence effects (pf in our theoretical model) are measured? The identification of deterrence effects on crime is an equally difficult empirical exercise because of the well-known potential simultaneity and reverse causality issues (Levitt, 1997), which cannot be totally solved using our network-based empirical strategy. Network fixed effects also prove useful in this respect. Because in our sample, networks are within schools, the use of network fixed effects also accounts for differences in the strictness of anti-crime regulations across schools

[28] Bramoullé et al. (2009) also deal with this problem in the case of a row-normalized $\mathbf{G}$ matrix. In their Proposition 5, they show that if the matrices $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ and $\mathbf{G}^3$ are linearly independent, then by subtracting from the variables the network average social effects are again identified and one can disentangle endogenous effects from correlated effects. In our dataset this condition of linear independence is always satisfied.

[29] As long as the link formation process between two individuals does not involve the characteristics of any third individual (see Sections 3.3). This assumption is subject to a diagnostic test below (Section 6.2).

(i.e. differences in the expected punishment for a student who is caught possessing illegal drug, stealing school property, verbally abusing a teacher, etc.). As mentioned above, they account for any kind of school level heterogeneity. As a result, instead of directly estimating deterrence effects (i.e. to include 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 crime, accounting for network fixed effects.

5.3 Econometric methodology

Let $\mathbf{y_r} = (y_{1,r}, \cdots, y_{n_r,r})'$, $\mathbf{x_r} = (x_{1,r}, \cdots, x_{n_r,r})'$, and $\boldsymbol{\epsilon}_\mathbf{r} = (\epsilon_{1,r}, \cdots, \epsilon_{n_r,r})'$. Denote the $n_r \times n_r$ sociomatrix by $\mathbf{G_r} = [g_{ij,r}]$, the row-normalized $\mathbf{G_r}$ by $\mathbf{G_r^*}$, and an n_r-dimensional vector of ones by $\mathbf{l_{n_r}}$. Then model (11) can be written in matrix form as:

$$\mathbf{y_r} = \phi \mathbf{G_r y_r} + \mathbf{x_r^*}\delta + \eta_r^* \mathbf{l_{n_r}} + \boldsymbol{\epsilon}_\mathbf{r},$$

where $\mathbf{x_r^*} = (\mathbf{x_r}, \mathbf{G_r^* x_r})$ and $\delta = (\beta', \gamma')'$.

For a sample with $\bar{r}$ networks, stack up the data by defining $\mathbf{y} = (\mathbf{Y_1'}, \cdots, \mathbf{Y_{\bar{r}}'})'$, $\mathbf{x}^* = (\mathbf{x_1^{*\prime}}, \cdots, \mathbf{x_{\bar{r}}^{*\prime}})'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_\mathbf{1}', \cdots, \boldsymbol{\epsilon}_\mathbf{\bar{r}}')'$, $\mathbf{G} = \mathrm{D}(\mathbf{G_1}, \cdots, \mathbf{G_{\bar{r}}})$, $\boldsymbol{\iota} = \mathrm{D}(\mathbf{l_{n_1}}, \cdots, \mathbf{l_{n_{\bar{r}}}})$ and $\boldsymbol{\eta}^* = (\eta_1^*, \cdots, \eta_{\bar{r}}^*)'$, where $\mathrm{D}(\mathbf{A_1}, \cdots, \mathbf{A_K})$ is a 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

$$\mathbf{y} = \mathbf{z}\theta + \boldsymbol{\iota} \cdot \boldsymbol{\eta}^* + \boldsymbol{\epsilon}, \tag{12}$$

where $\mathbf{z} = (\mathbf{Gy}, \mathbf{x}^*)$ and $\theta = (\phi, \delta')'$.

We treat $\boldsymbol{\eta}^*$ as a vector of unknown parameters. When the number of networks $\bar{r}$ is large, we have the incidental parameter problem. Let $\mathbf{J} = \mathrm{D}(\mathbf{J_1}, \cdots, \mathbf{J_{\bar{r}}})$, where $\mathbf{J_r} = \mathbf{I}_{n_r} - \frac{1}{n_r}\mathbf{l_{n_r}}\mathbf{l_{n_r}'}$. The network fixed effect can be eliminated by the transformation $\mathbf{J}$ such that

$$\mathbf{Jy} = \mathbf{Jz}\theta + \mathbf{J}\boldsymbol{\epsilon}. \tag{13}$$

Let $\mathbf{M} = (\mathbf{I} - \phi\mathbf{G})^{-1}$. The equilibrium outcome vector $\mathbf{y}$ in (12) is given by the reduced form equation

$$\mathbf{y} = \mathbf{M}(\mathbf{x}^*\delta + \boldsymbol{\iota} \cdot \boldsymbol{\eta}^*) + \mathbf{M}\boldsymbol{\epsilon}. \tag{14}$$

It follows that $\mathbf{Gy} = \mathbf{GMx}^*\delta + \mathbf{GM}\boldsymbol{\iota}\boldsymbol{\eta}^* + \mathbf{GM}\boldsymbol{\epsilon}$. $\mathbf{Gy}$ is correlated with ϵ because $\mathrm{E}[(\mathbf{GM}\boldsymbol{\epsilon})'\boldsymbol{\epsilon}] = \sigma^2 \mathrm{tr}(\mathbf{GM}) \neq 0$. Hence, in general, (13) cannot be consistently estimated by OLS.[30] If $\mathbf{G}$

[30] Lee (2002) has shown the OLS estimator can be consistent in the spatial scenario where each spatial unit is influenced by many neighbors whose influences are uniformly small. However, in the current data, the number of neighbors are limited, and hence that result does not apply.

is row-normalized such that $\mathbf{G} \cdot \mathbf{l_n} = \mathbf{l_n}$, the endogenous social interaction effect can be interpreted as an average effect. With a row-normalized $\mathbf{G}$, Lee et al. (2010) have proposed a partial-likelihood approach for the estimation based on the transformed model (13). However, for this empirical study, we are interested in the aggregate endogenous effect instead of the average effect. Hence, row-normalization is not appropriate. Furthermore, we are also interested in the centrality of networks that are captured by the variation in row sums (out-degrees) in the adjacency matrix $\mathbf{G}$. Row-normalization could eliminate such information. However, as $\mathbf{G}$ is not row-normalized in this empirical study, the (partial) likelihood function for (13) could not be derived, and alternative estimation approaches need to be considered.

In this paper, we estimate (13) by the 2SLS and generalized method of moments (GMM) approaches proposed by Liu and Lee (2010). The conventional instrumental matrix for the estimation of (13) is $\mathbf{Q_1} = \mathbf{J}(\mathbf{Gx}^*, \mathbf{x}^*)$ (*finite-IVs 2SLS*). For the case that the adjacency matrix $\mathbf{G}$ is not row-normalized, Liu and Lee (2010) have proposed to use additional instruments (IVs) $\mathbf{JG}\boldsymbol{\iota}$ so that $\mathbf{Q_K} = (\mathbf{Q_1}, \mathbf{JG}\boldsymbol{\iota})$ (*many-IVs 2SLS*). The additional IVs $\mathbf{JG}\boldsymbol{\iota}$ are based on the row sums of $\mathbf{G}$ and thus use the information on centrality of a network. Those additional IVs could help model identification when the conventional IVs are weak and improve upon the estimation efficiency of the conventional 2SLS estimator based on $\mathbf{Q_1}$. The number of such instruments depends on the number of networks. If the number of networks grows with the sample size, so does the number of IVs. The 2SLS could be asymptotic biased when the number of IVs increases too fast relative to the sample size (see, e.g., Bekker, 1994; Bekker and van der Ploeg, 2005; Hansen et al., 2008). Liu and Lee (2010) have shown that the proposed many-IV 2SLS estimator has a properly-centered asymptotic normal distribution when the average group size needs to be large relative to the number of networks in the sample. As detailed in Section 4, in this empirical study, we have a number of small networks. Liu and Lee (2010) have proposed a bias-correction procedure based on the estimated leading-order many-IV bias. The bias-corrected many-IV 2SLS estimator (*bias-corrected 2SLS*) is properly centered, asymptotically normally distributed, and efficient when the average group size is sufficiently large. It is thus the more appropriate estimator in our case study.

The 2SLS approach can be generalized to the GMM with additional quadratic moment equations (*finite-IVs GMM, many-IVs GMM*). While the IV moments use the information of the main regression function of (14) for estimation, the quadratic moments explore the correlation structure of the reduced form disturbances. Liu and Lee (2010) have shown that the many-IV GMM estimators can be consistent, asymptotically normal, and efficient when the sample size grows fast enough relative to the number of networks. Liu and Lee

(2010) have also suggested a bias-correction procedure for the many-IV GMM estimator based on the estimated leading order many-instrument bias. The bias-corrected many-IV GMM estimator (*bias-corrected GMM*) is shown to be more efficient than the corresponding 2SLS estimator. Appendix 3 details the derivation and asymptotic properties of both the 2SLS and GMM estimators.

5.4 Estimation results

Table 2a collects the estimation results of model (11) when using the different estimators discussed in the previous section.

As explained above, for the estimation of ϕ, we pool all the networks together by constructing a block-diagonal network matrix with the adjacency matrices from each network on the diagonal block. Hence we implicitly assume that the ϕ in the empirical model is the same for all networks. The difference between networks is controlled for by network fixed effects. Indeed, the estimation of ϕ for each network might be difficult (in terms of precision) for the small networks. Furthermore, it is a crucial empirical concern to control for unobserved network heterogeneity by using network fixed effects.

For equation (6) to be well-defined, ϕ needs to be in absolute value smaller than the inverse of the largest eigenvalue of the block-diagonal network matrix G (Proposition 1). In our case, the largest eigenvalue of $\mathbf{G}$ is 5.59. Furthermore our theoretical model postulates that $\phi \geq 0$. As a result, we can accept values within the range $[0, 0.179)$. Table 2 shows that all our estimates of ϕ are within this parameter space. As explained above, in our case study with small networks in the sample, the preferred estimator is the bias-corrected one. The GMM generalization improves upon the precision of the 2SLS estimates. Let us thus focus on the bias-corrected many-IV GMM estimator and interpret the results in terms of magnitude. We find that a standard deviation increase in the aggregate level of delinquent activity of the peers translate into a roughly 11 percent increase of a standard deviation in the individual level of activity. This is a strong effect, especially given our long list of controls.

[*Insert Table 2a here*]

5.5 Directed networks

So far, we have only considered *undirected* networks, i.e. we have assumed that friendship relationships are reciprocal, $g_{ij,r} = g_{ji,r}$. Our data, however, make it possible to know exactly who nominates whom in a network. Indeed, 20 percent of relationships in our dataset are not reciprocal.

In order to see how robust is our analysis, we now exploit the directed nature of the network data. Of course, the interpretation of centrality is now different since centrality contributions only flow in one direction on the directed links. We would like to see if our results change significantly under such a specification.

We follow the approach of Wasserman and Faust (1994, pages 205-210) who define the Katz-Bonacich centrality measure for directed networks. As they put it: "Centrality indices for directional relations generally focus on choices made".

In the language of graph theory, in a directed graph, a link has two distinct ends: a head (the end with an arrow) and a tail. Each end is counted separately. The sum of head endpoints count toward the *indegree* and the sum of tail endpoints count toward the *outdegree*. Formally, we denote a link from i to j as $g_{ij,r} = 1$ if j has nominated i as his/her friend, and $g_{ij,r} = 0$, otherwise. The indegree of student i, denoted by $g_{i,r}^{+}$, is the number of nominations student i receives from other students, that is $g_{i,r}^{+} = \sum_j g_{ij,r}$. The outdegree of student i, denoted by $g_{i,r}^{-}$, is the number of friends student i nominates, that is $g_{i,r}^{-} = \sum_j g_{ji,r}$. We consider only the indegree to define the Katz-Bonacich centrality measure. Observe that, by definition, the adjacency matrix $\mathbf{G}_r = [g_{ij,r}]$ is now *asymmetric*.

In the empirical analysis, we use *outdegrees* because if individual i nominates j but j does not, it is then very possible that j is a role model for i. In other words, i is learning from j even though j does not consider i as his/her best friend. In this context, y_j^*, the criminal activities of j, influences y_i^*.

From a theoretical point of view, the symmetry of $\mathbf{G}$ does not play any explicit role for the result established in Proposition 1. We can therefore define the Katz-Bonacich centrality measure $\mathbf{b}_{\boldsymbol{\alpha}}(g, \phi)$ exactly as in (3).

Turning to the empirical analysis, Table 2b reports the results of the estimation of model (11) when the directed nature of the network data is taken into account (i.e., with this alternative specification of $\mathbf{G}$). The parameter space $[0, 1/\mu_1(\mathbf{G}))$ is $[0, 0.322)$. Table 2b shows that the estimates of ϕ are all within this range. They are still statistically significant and only slightly higher in magnitude. Therefore, the results do not change substantially.

[*Insert Table 2b here*]

6 Who is the key player? Counterfactural Study

Let us now calculate empirically who is the key player in each our real-world networks. We set out a counterfactual study, which is now described.

6.1 Description of the procedure

With the estimates obtained from the bias-corrected many-IV GMM estimation procedure, for a network g_r, $\boldsymbol{\alpha}_\mathbf{r} = \mathbf{G}^*_\mathbf{r}\mathbf{x}_\mathbf{r}\gamma + \mathbf{x}_\mathbf{r}\beta + \boldsymbol{\eta}^*_\mathbf{r}\mathbf{l}_{\mathbf{n}_r} + \boldsymbol{\epsilon}_\mathbf{r}$ can be estimated by

$$\hat{\boldsymbol{\alpha}}_\mathbf{r} = (\mathbf{I}_{\mathbf{n}_r} - \hat{\boldsymbol{\phi}}\mathbf{G}_\mathbf{r})\mathbf{y}_\mathbf{r}$$

As $\mathbf{b}_{\boldsymbol{\alpha}_\mathbf{r}}(g,\phi) = (\mathbf{I}_{\mathbf{n}_r} - \hat{\phi}\mathbf{G}_\mathbf{r})^{-1}\hat{\boldsymbol{\alpha}}_\mathbf{r} = \mathbf{y}_\mathbf{r}$, the $n_r \times 1$ vector of Bonacich centrality of network g_r is given by $\mathbf{y}_\mathbf{r}$. As a result, the initial level of aggregate crime effort is given by:

$$\mathbf{b}^\mathbf{s}_{\boldsymbol{\alpha}_\mathbf{r}}(g,\phi) = \mathbf{l}'_{\mathbf{n}_r}(\mathbf{I}_{\mathbf{n}_r} - \hat{\phi}\mathbf{G}_\mathbf{r})^{-1}\hat{\boldsymbol{\alpha}}_\mathbf{r} = \mathbf{l}'_{\mathbf{n}_r}\mathbf{y}_\mathbf{r}$$

To identify the key player, we proceed exactly as in the theoretical model (see Section 3.2). For that, we calculate the crime reduction for removal of each player, one at a time, in the network. The key player is the one associated with the largest crime reduction. Let $\hat{\mathbf{e}}_\mathbf{r} = (\mathbf{I}_{\mathbf{n}_r} - \hat{\phi}\mathbf{G}_\mathbf{r})\mathbf{y}_\mathbf{r} - \mathbf{G}^*_\mathbf{r}\mathbf{x}_\mathbf{r}\hat{\gamma}$. When a player i is removed, we drop the ith row of $\mathbf{x}_\mathbf{r}$ and $\hat{\mathbf{e}}_\mathbf{r}$ to get $\widetilde{\mathbf{x}}_\mathbf{r}$ and $\tilde{\mathbf{e}}_\mathbf{r}$, and drop the ith row and column of $\mathbf{G}_\mathbf{r}$ to get $\tilde{\mathbf{G}}_\mathbf{r}$. Let $\tilde{\mathbf{G}}^*_\mathbf{r}$ be the row-normalized $\tilde{\mathbf{G}}_\mathbf{r}$. Then the aggregate crime effort with a player i being removed is

$$\mathbf{b}^\mathbf{s}_{\boldsymbol{\alpha}_\mathbf{r}}(g^{[-i]},\phi) = \mathbf{l}'_{\mathbf{n}_r}(\mathbf{I}_{\mathbf{n}_r-\mathbf{1}} - \hat{\phi}\tilde{\mathbf{G}}_\mathbf{r})^{-1}(\tilde{\mathbf{G}}^*_\mathbf{r}\widetilde{\mathbf{x}}_\mathbf{r}\hat{\gamma} + \tilde{\mathbf{e}}_\mathbf{r}) = \mathbf{l}'_{\mathbf{n}_r}\widetilde{\mathbf{y}}_\mathbf{r}$$

where $\widetilde{\mathbf{y}}_\mathbf{r}$ is the vector of criminal activities in network g_r when the criminal i has been removed.[31] As in the theoretical model (see (8)), the key player i is given by:

$$\arg\max_i(\mathbf{b}^\mathbf{s}_{\boldsymbol{\alpha}_\mathbf{r}}(g,\phi) - \mathbf{b}^\mathbf{s}_{\boldsymbol{\alpha}_\mathbf{r}}(g^{[-i]},\phi)) = \arg\min_i \mathbf{b}^\mathbf{s}_{\boldsymbol{\alpha}_\mathbf{r}}(g^{[-i]},\phi) \tag{15}$$

6.2 The invariant assumption on $g^{[-i]}$: Empirical issues

As observed in Section 3.2, in the calculation of the key player (in the formula (8) or, equivalently, in the simulations (15)), it is assumed that, when the key player is removed,

[31] Note that in this exercise the predicted Bonacich centralities and crime rates are the same because the definition of α_r in equation (6) ($\alpha_r = a_r - p_r f_r + \eta_r + \varepsilon_r$) includes the fixed-network effects ($\eta_r - p_r f_r$) and the error term ε_r. A less tractable set up where the equality is not necessarily true would imply to replace α_r by a_r in equation (6).

the other criminals in the network do not form new links (i.e. invariance of $g^{[-i]}$, i.e. network $g^{[-i]}$ has adjacency matrix $\mathbf{G}^{[-i]}$ where the ith row and the ith column have been removed from $\mathbf{G}$). In Section 3.3, we propose a simple network formation model that could justify this assumption. In this model, the link between i and j in network r only depends in the observable characteristics of i and j but not on the characteristics of the other criminals in the network (including i's friends other than j). In this section, we would like to test this model with our data.

Let us first consider undirected networks. For a network r with n_r criminals, if $\mathbf{G}_r$ is *undirected*, we have $n_r(n_r-1)/2$ distinct links in the network. Consider the following model:

$$g_{ij,r} = |x_{i,r} - x_{j,r}|\,\beta + \left(\min_{k\neq i,j} |x_{i,r} - x_{k,r}|\right)\gamma_1 + \left(\min_{k\neq i,j} |x_{j,r} - x_{k,r}|\right)\gamma_2 + \eta_r^* + \epsilon_{i,r} \qquad (16)$$

for $i = 1, \cdots, n_r - 1$, $j = i+1, ..., n_r$ and $r = 1, \cdots, \bar{r}$, and where the notations are the same as for model (11). Our aim is to test the hypothesis that $\gamma_1 = \gamma_2 = 0$, that is the link between i and j does not depend on individual k (whether k is a direct friend of i or not).

For directed networks, for a network r with n_r criminals, if $\mathbf{G}_r$ is *directed*, we have $n_r(n_r-1)$ distinct links in the network and we test the following model:

$$g_{ij,r} = |x_{i,r} - x_{j,r}|\,\beta + \left(\min_{k\neq i,j} |x_{i,r} - x_{k,r}|\right)\gamma + \eta_r^* + \epsilon_{i,r} \qquad (17)$$

for $i, j = 1, \cdots, n_r$, $i \neq j$ and $r = 1, \cdots, \bar{r}$, and where the notations are the same as for model (11). We will test here the hypothesis that $\gamma = 0$.

Here, we do not claim any causality. We are just looking at correlations and see if the network formation model proposed in Section 3.3 would not be rejected by the data. This is just a diagnostic check.

A linear probability model is estimated via least squares with network fixed effects. Tables 3a and 3b display the estimation results for the undirected and directed networks, respectively. It is clear from these tables that, for most variables, the formation of a link (i.e. friendship) between two criminals i and j is primarily affected by the observable characteristics of i and j but not by the characteristics of any other criminal $k \neq i, j$ belonging to the same network, that is, β is significant while γ (or γ_1 and γ_2 in the case of undirected networks) is not. Furthermore, since the sign of β is nearly always negative, there seems to be *homophily* in the friendship formation in these criminal networks, that is the closer two persons are in terms of characteristics, the more likely they will be friends.

[*Insert Table* 3a *and* 3b *here*]

6.3 Individual characteristics of key players

Once we have identified the key player for each network, we can draw his/her "profile" by comparing the characteristics of these key players with those of the other criminals in the network.[32] Table 4 displays the results only for the variables whose differences in means between these two samples are statistically significant. Compared to other criminals, "key" criminals belong to families whose parents are less educated and have the perception of being socially more excluded. They also feel that parents care less about them and have more troubles getting along with the teachers. Furthermore, the typical key player is more likely to be a male and have friends who are older and less attached to religion than other criminals. He/she is also more likely to come from residential areas with industrial properties of various types, although her/his friends are less likely to come from these kind of neighborhoods. Table A.2 in Appendix 4 contains the summary statistics of all the characteristics of the key players, as well as the ones of their best friends.

[*Insert Table* 4 *here*]

An interesting and important question that we seek to investigated empirically is whether the key player is always the player with the highest crime level (or equivalently with the highest Bonacich centrality in the network). We have shown in theoretical section that, in some cases, it is not the case (see Section 3.4) because the two measures (Bonacich versus inter-centrality) differ substantially in their foundation. Whereas the equilibrium-Bonacich centrality index (defined in (3)) derives from strategic individual considerations, the inter-centrality measure (defined in (8)) solves the planner's optimality collective concerns. In particular, the equilibrium Bonacich centrality measure fails to internalize all the network payoff externalities delinquents exert on each other, while the intercentrality measure internalizes them all.

For each of our 150 networks, we investigate whether the key player is also the most active criminal in the network (i.e. has the highest Bonacich centrality). We find that in 40 out of 150 networks (27%), it is not the case. This interesting (and unexpected) result is important for policy purposes since it means that, in some cases, we should not always target the most active criminals in a network.

In Table 5, we compare the characteristics of key players who are the most active criminals in the network with key players who are not. As in Table 4, Table 5 only shows variables

[32] Since the results on key players for directed networks are relatively similar, we will not discuss them. They are available upon request.

whose differences in means between these two samples are statistically significant. As compared to other key players, we find that key players who are *not* the most active criminals are more attached to religion,[33] are less likely to have single parents, to be socially integrated and have less troubles getting along with your teachers. They are also less physically developed and are less "able" individuals (to the extent to which mathematics score is a good proxy for individual ability). Now, looking at differences in the characteristics of their friends, it appears that key players who are not the most active criminals in the network have friends who have parents with higher education, who are less likely to be in manual occupations and a higher proportion of their friends feel that parents care very much about them. Compared to the key players who are the most active criminals in their network, their friends are also more religious. Our findings suggest that differences in the family background of the friends of key players may be important factors explaining the observed differences in crime between different types of key players. They can act as important protective factors. On the other hand, we are not able to detect relevant differences in neighborhood attributes to conclude that also neighborhood quality acts as a protective factor.

[*Insert Table* 5 *here*]

6.4 Petty versus serious crimes

We would like now to investigate whether the characteristics of key players differ by types of crime. For that, we split the reported offences between *petty crimes* and *more serious crimes*. The first group (*type-1 crimes or petty crimes*) encompasses the following offences: (*i*) paint graffiti or sign on someone else's property or in a public place; (*ii*) lie to the parents or guardians about where or with whom having been; (*iii*) run away from home; (*iv*) act loud, rowdy, or unruly in a public place; (*v*) take part in a group fight; (*vi*) damage properties that do not belong to you; (*vii*) steal something worth less than $50. The second group (*type-2 crimes or more serious crimes*) consists of (*i*): taking something from a store without paying for it; (*ii*) hurting someone badly enough to need bandages or care from a doctor or nurse; (*iii*) driving a car without its owner's permission; (*iv*) stealing something worth more than $50; (*v*) going into a house or building to steal something; (*iv*) using or threatening to use a weapon to get something from someone; (*vii*) selling marijuana or other drugs; (*viii*) getting into a serious physical fight.

We obtain a sample of 1099 petty criminals distributed over 132 networks and a sample

[33] Note that a higher value of the variable "Religion practice" means in fact that the person is less religious.

of 545 more serious criminals distributed over 75 networks. Petty crime networks have a minimum of 4 individuals and a maximum of 73 (with mean equals to 8.33 and standard deviation equals to 10.74), whereas the range for more serious crime networks is between 4 and 38 (with mean equals to 7.27 and standard deviation equals to 6.64).

We estimate model (11) for different crime types, which provides type of crime-specific peer effects. The results for *undirected networks* are contained in the last two columns of Table 2a. All estimates are within the acceptable parameter space $[0, 0.180)$ for type-1 crimes and $[0, 0.219)$ for type-2 crimes. In terms of magnitude, it appears that the impact of peer effects on crime are much higher for more serious crimes than for petty crimes. Indeed, we find that a standard deviation increase in the aggregate level of delinquent activity of the peers translate into a roughly 8 percent and 14.5 increase of a standard deviation in the individual level of activity for petty crimes and more serious crimes, respectively. The results for *directed networks* are contained in the last two columns of Table 2b. They are not qualitatively different (only slightly higher in magnitude). All the estimates are within the parameter space $[0, 1/\mu_1(\mathbf{G}))$, which is $[0, 0.322)$ and $[0, 0.423)$ for type-1 crimes and type-2 crimes, respectively.

We then repeat our counterfactual studies for key players for different types of crimes. Although the results of this exercise need to be taken with caution because of the small sample size of students committing the more serious offences, we report our findings in Tables $6-10$.[34]

Table 6 and 7 have the same structure as Table 4 but draw a profile of the key player for petty and more serious crimes. As compared to other criminals, a key player committing petty crimes is more likely to be a male, less likely to be black, is more able than other criminals, more likely to feel that parents do not care very much about him/her and has troubles getting along with teachers. His/her friends have parents who are less likely to be office or sales workers and in the farm or fishery sector. They come more frequently from suburban areas and have less troubles getting along with the teachers than friends of other types of criminals.

The portrait of a key player committing serious crimes has different features. Even though he/she is more likely to be a male, he/she is more physically developed compared to the boys of his/her age, feels to be part of the school but has troubles getting along with teachers. He/she is also more likely to reside in suburban areas and less likely in urban residential areas. Neither key players committing serious crime nor their friends have parents

[34] Also, in this case, we expose our results undirected networks only. The evidence for directed networks is similar. These results are also available upon request.

working in the military or security sector. The typical key player committing serious crimes has friends who are older, less likely to come from industrial residential areas and his/her parents are less likely to be in the farm or fishery sector.

[*Insert Tables* 6 *and* 7 *here*]

Table 8 instead compares the *characteristics* of key players for different types of crimes. We find only few significant differences in individual characteristics while differences in their friends' characteristics seem to be more important. Key players committing serious crime are more likely to be black and have friends who are blacks and feel that adults care about them more than key players committing petty crimes. This last aspect is not true for their friends. Also, as compared to friends of key players committing petty crimes, key players' friends in serious crimes are more able, have more troubles with teachers, are more likely to have parents in office or sales occupations and less likely to have parents occupied in the military or security sector. A larger proportion of them comes from urban residential areas.

[*Insert Table* 8 *here*]

Finally, Tables 9 and 10 have the same structure as Table 5 and uncover the difference in characteristics for different types of crimes between key players who are the most criminals and those who are not. When the offences are differentiated by the type of crime committed, we still find some networks where the most "harmful" criminal is not the individual with the highest Bonacich centrality (30% for petty crimes and 25% for serious crimes). Key players committing petty crimes who are not the most active criminals are less physically developed compared to more active key players, more religious and have less troubles getting along with the teachers. Interestingly, they have also less educated parents and their parents are less likely to be managers.

For serious crimes, the picture takes again different aspects. Key players in serious crimes who are not the most active criminals are individuals coming from households with more numerous members and are less likely to be of a race different from white and black. They also come from better quality neighborhoods, are more likely to live in suburban residential areas while their friends are more likely to reside in urban residential areas. For serious crimes, we also find that differences in parenting between friend groups might be important. Although less active key players have friends who are less likely to have two married parents and more likely to come from single parent families, almost all of these friends feel that parents care about them very much.

[*Insert Tables* 9 *and* 10 *here*]

6.5 Key players and network topology

As in Section 3.4, let us now investigate the characteristics of these key players in terms of other network centrality measures (i.e. other than Bonacich centrality).

So far, we have used the Bonacich centrality measure to capture the importance of network structure. The reason is that this measure has a precise behavioral foundation, as it stems from our theoretical model. However, it counts the number of any path connecting one node to the others, not the optimal ones. Let us then consider two other traditional measures of centrality in network analysis that are based on optimal paths, i.e. *closeness* and *betweenness* centralities, and a measure of cohesion of the sub-network around each node, i.e. the *clustering coefficient.* They are all defined in Appendix 4 and used in the theory section (3.4.)

Table 11 provides information on the distributions of these measures for the key players in our networks and compares them with the Bonacich centrality (which is equal to the crime level of each individual). Looking at the first measure, betweenness centrality, one can first notice that at least 50% of our key players has a betweenness centrality equal to zero (i.e. the median is equal to 0), meaning that there are few shortest paths that go through them. However, if we consider the upper tail of this distributions, that is we look at the key players with the highest betweenness centrality, we see that a larger portion of them are key players who are not the most active criminals. Indeed, above the 90th percentile of the distribution of the whole sample, 10% of the key players are not the most active criminals while it is 4.5% for the key players who are the most active criminals. This finding suggests that, even though some criminals do not commit much crime, they can be key players because they have a crucial position in the network in terms of betweenness centrality (for example, in the network described in Section 3.4, individual 1 who bridges two otherwise separated networks is not the most active criminal but is the key player and has the highest betweenness centrality). When looking at the closeness centrality, the results are quite different. Indeed, plenty of key players are quite central (median equals to 0.5). We also find that more active key players tend to be more concentrated in the upper tail of the closeness distribution than less active key players (11% in the upper 90% tail versus 5%). Finally, the results on the clustering coefficient suggest that the most active criminals are more likely to operate in tighter networks of best friends (4.5% in the upper 95% tail versus 2.5% for key players who are not the most active criminals).

[*Insert Table* 11 *here*]

In Tables 12 and 13, we perform the same analysis for petty and serious crimes, respectively. Interestingly, we find that, for serious crimes, key players have high betweenness centrality while this is less the case for petty crimes. Indeed, if we look at *p75* (lower 75% of the distribution), we see that among key players, at least 75% of them has a betweenness centrality less than 0.05 for petty crime while, for more serious crimes, this value is 0.67. Moreover, for petty crimes, the most active key players seem also to be the more central ones in terms of all measures considered while, for serious crimes, the most central players in terms of betweenness tend to be the less criminal ones. This evidence suggests that, for serious crimes, network position is an important determinant of key players.

[*Insert Tables* 12 *and* 13 *here*]

Finally, in Tables 14, 15 and 16, we investigate the role of network characteristics[35] in explaining the differences between key players who are the most active criminals and those who are not. In terms of statistical significance, the differences are not pronounced. We only find that, for serious crimes, the average degree is significantly higher for most-active key players. If we only look at the qualitative evidence, then we see that, for all crimes, the network diameter, network betweenness and the average distance are smaller for most-active key players. An interesting suggestive result is that networks tend to be dissortative ("popular" criminals are associated with less "popular" ones) for petty crimes while assortative ("popular" criminals are associated with "popular" ones) for most-active key players committing more serious crimes.

[*Insert Tables* 14, 15 *and* 16 *here*]

7 Policy implications

Identifying peer effects in crime We would like to discuss now some policy implications of our results. As noted by Manski (1993, 2000) and Moffitt (2001), it is important to *separately* identify peer or endogenous effects from contextual or exogenous effects. This is because endogenous effects generate a *social multiplier* while contextual effect don't. In

[35] They are all defined in Section 3.4.

the context of crime, this means that a special program targeting some individuals will have multiplier effects: the individual affected by the program will reduce its criminal activities and will influence the criminal activities of his/her peers, which, in turn, will affect the criminal activities of his/her peers, and so on. On the other hand, if only contextual effects are present, then there will be no social multiplier effects from any policy affecting only the "context" (for example, improving the quality of the teachers at school). Therefore, the identification of these two effects is of paramount importance for policy purposes. Another important policy issue in the estimation of social interactions is the separation between peer effects and confounding effects. Indeed, the formation of peer group is not random and individuals do select into groups of friends. It is therefore important to separate the endogenous peer effects from the correlated effects (Manski, 1993), i.e. the same criminal activities may be due to common unobservable variables (such as, for example, the fact that individuals from the same network like bowling together) faced by individuals belonging to the same network rather than peer effects. This is also very important for crime policies since, for example, if the high-crime rates are due to the fact that teenagers like to bowling together, then obviously the implications are very different than if it is due to peer effects.

The first aim of this paper was to clearly identify the peer effects from the contextual affects and from the correlated effects. For that, we first developed a theoretical model where all these effects were clearly separated. We then estimated the results of the model by using an econometric techniques, which utilizes the structure of the network as well as network fixed effects to identify each of these effects. We find that, indeed, peer effects are important in criminal activities for teenagers in the United States, indicating that any policy targeting some criminals will have multiplier effects.

Implementing key-player policies: Theoretical issues Once this has been showed, policy issues can be seen from a different perspective. Indeed, in the standard crime literature without social interactions (Becker, 1968; Garoupa, 1997; Polinsky and Shavell, 2000), punishment is seen as an effective tool for reducing crime. But punishment is random and not targeted to individuals that generate the highest multiplier effects. To address this issue, we have developed a theoretical framework where a key player has been identified. A key player is someone that needs to be removed in order to reduce as much as possible total crime in the network. The way a key player is calculated is precisely using the multiplier effects due to endogenous peer effects. Consider the key player removal policy. Indeed, when a delinquent is removed from network r, the intercentrality measures of all the delinquents that remain active are reduced, that is, $d_{j,r}(g_r^{[-i^*]}, \phi) \leq d_{j,r}(g_r, \phi)$, for all $j \neq i^*$, which triggers a decrease in delinquency involvement for all of them. Moreover, when delinquent i^* is removed from the delinquency network, the corresponding ratio of aggregate delinquency reduction with respect to the network centrality reduction is an increasing function of the intercentrality measure $d_{i,r}(g,\phi)$ of this delinquent i in network r (Ballester et al., 2010). This means that *the target policy displays amplifying effects*, and the gains following the judicious choice of the key player (the one with highest intercentrality measure) go beyond the differences in intercentrality measures between this player and any other delinquent in the network.

To fully assess the relevance of the key player delinquency policy, let us compare the relative returns of a network targeted policy with that of a random target policy. For each criminal i in the crime network g_i, define:

$$\xi_{i,r}(g_r) = n_r \frac{y_r^*(g_r) - y_r^*(g_r^{[-i]})}{\sum_{j=1}^{n_r} \left[y_r^*(g_r) - y_r^*(g_r^{[-j]}) \right]}.$$

This is the ratio of returns (in delinquency reduction) when i is the selected target versus a random selection with uniform probability for all delinquents in the network.

Denote by $\overline{d}_r(g_r,\phi)$ the average of the intercentrality measures in network r, and by $\sigma_{\mathbf{d}_r}(g_r,\phi)$ the standard deviation of the distribution of this intercentrality measures. It can be shown that (Ballester et al., 2010):

$$\xi_{i^*,r}(g_r) \geq 1 + \frac{\sigma_{\mathbf{d}_r}(g_r,\phi)}{\overline{d}_r(g_r,\phi)}.$$

where i^* is the key player in g_r for a given ϕ. The relative gains from targeting the key player instead of operating a selection at random in the delinquency network increase with the variability in intercentrality measures across delinquents as captured by $\sigma_{\mathbf{d}_r}(g_r,\phi)$. In other words, the key player prescription is particularly well-suited for networks that display stark location asymmetries across nodes. In these cases, it is more likely than the relative gains from implementing such a policy compensate for its relative costs.

The second aim of this paper was precisely to determine the key player in each of our adolescent networks. Because of its multiplier effects, it is important to know what are his/her characteristics, to which network does he/she belongs and if he/she has a different profile for different types of crime. We find that, compared to other criminals, key players are more likely to be a male, have less educated parents, are less attached to religion and feel socially more excluded. They also feel that adults care less about them, are less attached to their school and have more troubles getting along with the teachers. We also find that, even though some criminals are not very active in criminal activities, they can be key players because they have a crucial position in the network in terms of betweenness centrality.

Implementing key-player policies: Real-world issues How can we implement a key-player policy? There is a small literature that discussed and tested policies aiming at "neutralizing" disruptive kids because of negative peer interaction effects they have on other kids. Lazear (2001) proposed a model showing that class size can be an issue if some kids are disruptive. Indeed, classroom education has public good aspects. The technology is such that when one student disrupts the class, learning is reduced for all other students. Neither the student nor the classmates can learn much when the student is misbehaving, causing the teacher to allocate her time to him. The model implies that better students are optimally placed in the large classes, despite the reduced teacher-student ratio. Using our model, we could define the key player as the most disruptive student in a classroom, i.e., the student who once removed generates the highest possible increase in total education activity (as measured by the grades of the students). "Removing" the key player would mean here to put this student in another class or investing special resources (like having an extra teacher) on him. If we follow Lazear's theory, it would be optimal to have "key players" in classes of smaller size. Our theory helps determine who is the key player by using our intercentrality measure while our empirical analysis helps identify the characteristics of the key player.

It is often suggested that one way to reduce juvenile crime is to lengthen the school day or school year and/or to provide activities for young people when school is not in session. The implicit notion behind such program-oriented solutions to juvenile crime is a belief in the importance of *incapacitation*−that, as Jacob and Lefgren (2003) put it: "idle hands are the devil's workshop" and that keeping kids busy will keep them out of trouble. Advocates of after-school and other youth programs frequently claim that juvenile violence peaks in the after-school hours on school days and in the evenings on nonschool days. Using exogenous policy changes and other events that effectively force students to stay in school or take extra days off (e.g. changes in compulsory schooling laws, teacher in-service days and strikes), a few recent studies have shown that school attendance affects crime in rich and complex ways. Forcing some students to stay in school an extra year or two reduces both violent and property crime substantially (Anderson 2009).[36] Yet, day-to-day changes in school attendance have

[36] Anderson's estimates for total arrest rates imply that a compulsory schooling age of 17 significantly reduces age 17 arrests by about 8% (5.4 arrests per 1,000 youth) compared to a compulsory schooling age of 16 or less. Similarly, an age 18 compulsory schooling age significantly reduces arrests by 9.7- 11.5% at ages 16-18. Overall, the estimates generally suggest that forcing youth to spend an extra year or two in high school significantly reduces their arrest rates over that period.

opposing effects on violent and property crime. An extra day of school appears to reduce property crime while increasing violent crime (Jacob and Lefgren, 2003; Luallen 2006).[37] The latter most likely reflects social interaction effects from bringing together hundreds of adolescents and letting them all loose at the same time.[38]

All the potential effects of school attendance on crime are likely to be relevant to changes in compulsory schooling, while the effects of in-service days and teacher strikes are likely to be limited to incapacitation and social interactions (Lochner, 2011). Any social interaction effects are likely to be magnified in the latter cases due to the universal nature of the policy. Using our framework, we could recommend the same policies to reduce juvenile crime (i.e. lengthen the school day or school year and/or to provide activities for young people when school is not in session) by targeting "key players" instead of encompassing everybody. In their conclusion, Jacob and Lefgren (2003) suggest that summer youth employment programs or smaller, neighborhood-based after-school programs, that provide structured activities for adolescents but do not substantially increase their concentration, may be the best way to reduce juvenile crime. We could apply the same type of programs to "key players" that we could identify using our framework. Targeting these "key players", i.e. delinquents who once removed generate the highest possible reduction in aggregate delinquent level in a network, can have large effects on crime because of the feedback effects or "social multipliers" at work.

There is other strong evidence that key players matter. Two recent papers use exogenous shocks to test the impact of "key players" on outcomes. Azoulay et al. (2010) look at the impact of the sudden (i.e. prematurely and unexpected) death of 112 academic "superstars" on the productivity of their co-authors. Waldinger (2010) analyzes the effect of the expulsion

[37] Jacob and Lefgren (2003) examine the effects of single day changes in school-wide attendance on juvenile crime and arrest rates in 29 large American cities from 1995 to 1999. Exploiting teacher in-service days across jurisdictions over time as an exogenous source of variation in school days, they essentially compare local juvenile crime rates on days when school is not in session to those when it is. Their findings suggest that an additional day of school reduces serious juvenile property crime by about 14% that day while it increases serious juvenile violent crime by 28%. These results are consistent with an *incapacitation effect* of school that limits participation in property crime. However, the increased level of interaction among adolescents facilitated through schools may raise the likelihood of violent conflicts (and other minor delinquency) after school. Luallen (2006) follows a similar approach, using teacher strikes (typically lasting about 5 days) rather than in-service days as an exogenous source of school days. He finds that the incapacitation and social interaction effects appear to be particularly strong in urban areas and negligible elsewhere.

[38] Kline (2010) evaluates the effectiveness of curfew ordinances by comparing the arrest behavior of various age groups within a city before and after curfew enactment. The evidence suggests that curfews are effective at reducing both violent and property crimes committed by juveniles below the statutory curfew age. Curfews do not appear to be effective at influencing the criminal behavior of youth just above the curfew age, suggesting that the choice of statutory curfew age is important in crafting policy.

of mathematics professors in Nazi Germany on PhD student outcomes. Both studies find strong effects.

We believe that our key-player policy has more general policy implications and can be applied to contexts other than crime and education. For example, the financial market is very connected and can be considered as a network where links could be loans between banks (Leitner, 2005). A key-player policy would be to identify the key bank that needs to be bailed out in order for the system to resist a financial crisis. We could also apply the key player policy to the issue of adoption of a new technology in developing countries. There is indeed strong evidence of social learning (Conley and Udry, 2010). One could therefore identify key players and target them so that their influence on others will be crucial in the adoption of a new technology. 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, 2010). One could identify "key politicians" who should be promoted within the party because they would have a significant impact on election outcomes.

In the real-world, howeever, we do not always have the exact information on the social network of each individual. In that case, it could be quite difficult to determine the key player in a given network since the computation of the intercentrality measures relies on the knowledge of the adjacency matrix of the delinquency network. This matrix is obtained from sociometric data that identifies the network links between delinquents.[39] Our empirical framework can help determine the key player by identifying the key and significant characteristics highlighted in tables 4 to 16.

Finally, we hope that the results obtained in this paper will allow policy makers to think differently about crime and that, in the context of financial crisis and budget deficits, there is an alternative to "brute force" since targeting some specific individuals can have dramatic effects on crime reduction.

[39] Costebander and Valente (2003) show that centrality measures based on connectivity (rather than betweenness), such as intercentrality measures, are robust to mispecifications in sociometric data, and thus open the door to estimations of centrality measures with incomplete samples of network data. This, obviously, reduces the cost of identifying the key player. The idea behind these results is that these measures take into account *all* walks in the network. Thus, generally the centrality of an individual is not determined only by his/her direct links but by the complete structure of the network. In this sense, the probability that a missing link affects the choice of the most central/intercentral player is smaller than with other type of measures.

References

[1] Anderson, M. (2009), "In school and out of trouble? The minimum dropout age and juvenile crime," Unpublished manuscript.

[2] Azoulay, P., J.S. Graff Zivin and J. Wang (2010), "Superstar extinction," *Quarterly Journal of Economics* 125, 549-589.

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

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

[5] 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.

[6] Becker, G. (1968) "Crime and punishment: An economic approach," *Journal of Political Economy* 76, 169-217.

[7] Bekker, P. (1994), "Alternative approximations to the distributions of instrumental variable estimators," *Econometrica* 62, 657-681.

[8] Bekker, P. and J. van der Ploeg (2005), "Instrumental variable estimation based on grouped data," *Statistica Neerlandica* 59, 239-267.

[9] 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, forthcoming.

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

[11] 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*, Committee on Human Factors, National Research Council, pp. 241-252.

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

[13] Bramoullé, Y. and B. Fortin (2009), "The econometrics of social networks," In: S. Durlauf and L. Blume (Eds.), *The New Palgrave Dictionary of Economics*.

[14] Brock, W. and S. Durlauf (2001) "Discrete choice with social interactions," *Review of Economic Studies* 68, 235-260.

[15] Burdett, K., Lagos, R., and R. Wright (2003) "Crime, Inequality and Unemployment," *American Economic Review* 93, 1764-1777.

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

[17] Calvó-Armengol, A., Verdier, T. and Y. Zenou (2007), "Strong and weak ties in employment and crime," *Journal of Public Economics* 91, 203-233.

[18] 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.

[19] Clark, A.E. and Y. Loheac (2007), "It wasn't me, it was them! Social influence in risky behavior by adolescents," *Journal of Health Economics* 26, 763-784.

[20] Cohen-Cole, E. (2006), "Multiple groups identification in the linear-in-means model," *Economics Letters* 92, 157-162.

[21] Cohen, L. and C. Malloy (2010), "Friends in high places," NBER Working Paper No. 16437.

[22] Conley, J.P. and P. Wang (2006), "Crime and ethics," *Journal of Urban Economics* 60, 107-123.

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

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

[25] Costebander, E. and T.W. Valente (2003), "The stability of centrality measures when networks are sampled," *Social Networks* 25, 283-307.

[26] Damm, A.P. and C. Dustmann (2008), "Do young people learn criminal behavior? Quasi-experimental evidence," Unpublished manuscript, University College London.

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

[28] Ferrer, R. (2010), "Breaking the law when others do: A model of law enforcement with neighborhood externalities," *European Economic Review* 54, 163-180.

[29] Garoupa, N. (1997), "The theory of optimal law enforcement," *Journal of Economic Surveys* 11, 267-295.

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

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

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

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

[34] Huang, C.-C., Laing, D., and P. Wang (2004) "Crime and Poverty: A Search-Theoric Approach," *International Economic Review* 45, 909-938.

[35] Ioannides, Y.M. and L.D. Loury (2004), "Job information networks, neighborhood effects, and inequality," *Journal of Economic Literature* 42, 1056-1093.

[36] İmrohoroğlu, A., Merlo, A., and P. Rupert (2000), "On the Political Economy of Income Redistribution and Crime," *International Economic Review* 41, 1-25.

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

[38] Jackson, M.O. and A. Wolinsky (1996), "A strategic model of social and economic networks," *Journal of Economic Theory* 71, 44-74.

[39] Jacob, B. and L. Lefgren (2003), "Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime," *American Economic Review* 93, 1560-1577.

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

[41] Kleiman, M.A. (1993), "Enforcement swamping: A positive-feedback mechanism in rates of illicit activity," *Mathematical and Computer Modeling* 17, 65-75.

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

[43] Kline, P. (2010), "The impact of juvenile curfew laws," Unpublished manuscript, University of California, Berkeley.

[44] 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.

[45] Laschever, R. (2009), "The doughboys networks: Social interactions and labor market outcomes of World War I veterans," Unpusblished manuscript, University of Illinois at Urbana-Champaign.

[46] Lazear, E.P. (2001), "Educational production," *Quarterly Journal of Economics* 116, 777-803.

[47] Lee, L.F. (2002), "Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models," *Econometric Theory* 18, 252-277.

[48] Lee, L.F. (2003), "Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances," *Econometric Reviews* 22, 307-335.

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

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

[51] Leitner, Y. (2005), "Financial networks: Contagion, commitment, and private sector bailouts," *Journal of Finance* 6, 2925-2953.

[52] Levitt, S.D. (1997), "Using electoral cycles in police hiring to estimate the effect of police on crime," *American Economic Review* 87, 270-90.

[53] Lin, X. (2010), "Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables," *Journal of Labor Economics* 28, 825-860.

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

[55] Lochner, L. (2004), "Education, work, and crime: A human capital approach," *International Economic Review* 45, 811-843.

[56] Lochner, L. (2011), "Education policy and crime," In P.J. Cook, Ludwig, J. and J. McCrary (Eds.), *Controlling Crime: Strategies and Tradeoffs*, Chicago: University of Chicago Press, forthcoming.

[57] Luallen, J. (2006), "School's out ... forever: A study of juvenile crime, at-risk youths and teacher strikes," *Journal of Urban Economics* 59, 75-103.

[58] 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.

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

[60] Manski, C.F. (2000), "Economic analysis of social interactions," *Journal of Economic Perspectives* 14, 115-136.

[61] Moffitt, R. (2001), "Policy interventions low-level equilibria, and social interactions," In: S. Durlauf and P. Young (Eds.), *Social Dynamics*, Cambridge, MA: MIT Press, pp. 45-82.

[62] Patacchini, E. and Y. Zenou (2008), "The strength of weak ties in crime," *European Economic Review* 52, 209-236.

[63] Patacchini, E. and Y. Zenou (2011), "Juvenile delinquency and conformism," *Journal of Law, Economic, and Organization* forthcoming.

[64] Polinsky, M.A. and S. Shavell (2000), "The economic theory of public enforcement of law," *Journal of Economic Literature* 38, 45-76.

[65] Rasmussen, E. (1996), "Stigma and self-fulfilling expectations of criminality," *Journal of Law and Economics* 39 , 519-543.

[66] Sah, R. (1991), "Social osmosis and patterns of crime," *Journal of Political Economy* 99, 1272-1295.

[67] Schrag, J. and S. Scotchmer (1997), "The self-reinforcing nature of crime," *International Review of Law and Economics* 17, 325-335.

[68] Silverman, D. (2004), "Street crime and street culture," *International Economic Review* 45, 761-786.

[69] Sirakaya, S. (2006) "Recidivism and Social Interactions," *Journal of the American Statistical Association* 101, 863-877.

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

[71] 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.

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

[73] Waldinger, F. (2010), "Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany," *Journal of Political Economy* 118, 787-831.

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

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

[76] Weinberg, B., P.B. Reagan, and J.J. Yankow (2004), "Do neighborhoods affect work behavior? Evidence from the NLSY 79," *Journal of Labor Economics* 22, 891-924.

Appendix 1: Data appendix

Table A.1: Description of Data (1,297 individuals, 150 networks)

	Variable definition	Mean	St.dev	Min	Max
Delinquency index	In the text	0.94	1.09	0.09	9.63
Delinquency index of best friends	Aggregate value of the delinquency index over direct friends.	2.17	2.30	0.09	20.23
Delinquency index (type-1 crime)	In the text	1.15	1.15	0.20	7.31
Delinquency index of best friends (type-1 crime)	Aggregate value of the delinquency index over direct friends.	2.65	2.55	0.20	20.47
Delinquency index (type-2 crime)	In the text	1.25	1.36	0.28	12.55
Delinquency index of best friends (type-2 crime)	Aggregate value of the delinquency index over direct friends.	2.57	2.48	0.28	17.52
Individual socio-demographic variables					
Female	Dummy variable taking value one if the respondent is female.	0.32	0.47	0	1
Religion practice[1]	Response to the question: "In the past 12 months, how often did you attend religious services", coded as 4= never, 3= less than once a month, 2= once a month or more, but less than once a week, 1= once a week or more. Coded as 5 if the previous is skipped because of response "none" to the question: "What is your religion?"	2.35	1.48	1	5
Student grade	Grade of student in the current year.	9.15	1.59	7	12
Black or African American	Race dummies. "White" is the reference group.	0.19	0.39	0	1
Other races	"	0.07	0.25	0	1
Mathematics score[2]	Score in mathematics at the most recent grading period, coded as4= D or lower, 3= C, 2=B, 1=A.	2.29	1.10	1	4
Self esteem	Response to the question: "Compared with other people your age, how intelligent are you", coded as 1= moderately below average, 2= slightly below average, 3= about average, 4= slightly above average, 5= moderately above average, 6= extremely above average.	3.97	1.16	1	6
Physical development	Response to the question: "How advanced is your physical development compared to other boys/girls your age", coded as 1= I look younger than most, 2= I look younger than some, 3= I look about average, 4= I look older than some, 5= I look older than most	3.45	1.23	1	5
Family background variables					
Household size	Number of people living in the household.	4.46	1.26	2	11
Two married parent family	Dummy taking value one if the respondent lives in a household with two parents (both biological and non biological) that are married.	0.66	0.48	0	1
Single parent family	Dummy taking value one if the respondent lives in a household with only one parent (both biological and non biological).	0.27	0.44	0	1
Parent education	Schooling level of the (biological or non-biological) parent who is living with the child, distinguishing between "never went to school", "not graduate from high school", "high school graduate", "graduated from college or a university", "professional training beyond a four-year college", coded as 1 to 5.	3.08	1.17	0	5

[1] A higher value means here less religious.
[2] A higjher value means here lower grade.

	We consider only the education of the father if both parents are in the household.				
Parent occupation manager	Parent occupation dummies. Closest description of the job of (biological or non-biological) parent that is living with the child is manager. If both parents are in the household, the occupation of the father is considered. "none" is the reference group	0.15	0.36	0	1
Parent occupation professional/technical	”	0.17	0.38	0	1
Parent occupation office or sales worker	”	0.11	0.31	0	1
Parent occupation manual	”	0.29	0.45	0	1
Parent occupation military or security	”	0.01	0.08	0	1
Parent occupation farm or fishery	”	0.01	0.12	0	1
Parent occupation other	”	0.17	0.37	0	1
Protective factors					
School attachment[3]	Response to the question: "You feel like you are part of your school coded as 1= strongly agree, 2= agree, 3=neither agree nor disagree, 4= disagree, 5= strongly disagree.	2.10	0.94	1	5
Trouble relationship with teachers	Response to the question: "How often have you had trouble getting along with your teachers?" 0= never, 1= just a few times, 2= about once a week, 3= almost everyday, 4=everyday	1.38	1.03	0	4
Social inclusion	Response to the question: "How much do you feel that adults care about you, coded as 5= very much, 4= quite a bit, 3= somewhat, 2= very little, 1= not at all	4.28	0.82	2	5
Parental care	Dummy taking value one if the respondent reports that the (biological or non-biological) parent that is living with her/him or at least one of the parents if both are in the household cares very much about her/him	0.83	0.37	0	1
Residential neighborhood variables					
Residential building quality[4]	Interviewer response to the question "How well kept is the building in which the respondent lives", coded as 4= very poorly kept (needs major repairs), 3= poorly kept (needs minor repairs), 2= fairly well kept (needs cosmetic work), 1= very well kept.	1.60	0.87	1	4
Residential area suburban	Residential area type dummies: interviewer's description of the immediate area or street (one block, both sides) where the respondent lives. "Rural area" is the reference group.	0.38	0.49	0	1
Residential area urban - residential only	”	0.23	0.42	0	1
Residential area industrial properties - mostly wholesale	”	0.00	0.00	0	1
Residential area other type	”	0.05	0.21	0	1

[3] A higher value means here less school attachment.
[4] A higher value means here lower building quality.

Appendix 2: Identification of network models with non-row-normalized adjacency matrices

Consider the following model

$$\begin{aligned}\mathbf{y_r} &= \phi_0 \mathbf{G_r y_r} + \mathbf{x_r}\beta_0 + \gamma_0 \mathbf{G_r^* x_r} + \eta_r \mathbf{l}_{\mathbf{n}_r} + \boldsymbol{\epsilon}_\mathbf{r} \\ &= [\mathbf{G_r y_r}, \mathbf{x_r}, \mathbf{G_r^* x_r}, \mathbf{l}_{\mathbf{n}_r}]\vartheta + \boldsymbol{\epsilon}_\mathbf{r}, \end{aligned} \tag{18}$$

where $\mathbf{G_r^*}$ is the row-normalized $\mathbf{G_r}$ and $\vartheta = (\phi_0, \beta_0', \gamma_0', \eta_r)'$. To achieve model identification (based only on the reduced form regression equation),, we needs that the deterministic part of the right hand side variables, $[\mathrm{E}(\mathbf{G_r y_r}), \mathbf{x_r}, \mathbf{G_r^* x_r}, \mathbf{l}_{\mathbf{n}_r}]$, have full column rank, where

$$\mathrm{E}(\mathbf{G_r y_r}) = \mathbf{G_r x_r}\beta_0 + \phi_0 \mathbf{G_r M_r G_r x_r}\beta_0 + \mathbf{G_r M_r G_r^* x_r}\gamma_0 + \eta_r \mathbf{G_r M_r l}_{\mathbf{n}_r}. \tag{19}$$

First, we consider the case that $\mathbf{G_r}$ is row-normalized such that $\mathbf{G_r} = \mathbf{G_r^*}$. In this case, (19) can be simplified as

$$\mathrm{E}(\mathbf{G_r y_r}) = \mathbf{G_r^* x_r}\beta_0 + \mathbf{G_r^* M_r G_r^* x_r}(\phi_0\beta_0 + \gamma_0) + \frac{\eta_r}{1-\phi_0}\mathbf{l}_{\mathbf{n}_r}.$$

To illustrate the challenges in identification, we consider three cases. (1) $\beta_0 = \gamma_0 = 0$. This is the case when there is no relevant exogenous variables in the model. In this case, $\mathrm{E}(\mathbf{G_r y_r}) = \frac{\eta_r}{1-\phi_0}\mathbf{l}_{\mathbf{n}_r}$. Hence, the model is not identified because $[\frac{\eta_r}{1-\phi_0}\mathbf{l}_{\mathbf{n}_r}, \mathbf{x_r}, \mathbf{G_r^* x_r}, \mathbf{l}_{\mathbf{n}_r}]$ does not have full column rank. (2) $\beta_0\phi_0 + \gamma_0 = 0$. In this case, $\mathrm{E}(\mathbf{G_r y_r}) = \mathbf{G_r^* x_r}\beta_0 + \frac{\eta_r}{1-\phi_0}\mathbf{l}_{\mathbf{n}_r}$. The model can not be identified due to perfect collinearity. This corresponds to the case where the endogenous effect and exogenous effect exactly cancel out. Lee et al. (2010) have shown, in this case, the reduced form of (18) becomes a simple regression model with (spatially) correlated disturbances. In the reduced form, there are neither endogenous nor contextual effects. Interactions go through unobservables (disturbances) instead of observables. (3) $\beta_0\phi_0 + \gamma_0 \neq 0$. For this case, Bramoullé et al. (2009) and Lee et al. (2010) have derived some sufficient conditions for model identification, which are simpler to interpret. The identification can still be hard to achieve when the network is dense. For example, the "reflection problem", where the endogenous effects can not be identified from the contextual effects due to perfect collinearity, exists in the model of Manski (1993), which has the linear-in-mean specification such that $\mathbf{G_r} = \frac{1}{n_r}\mathbf{l}_{\mathbf{n}_r}\mathbf{l}_{\mathbf{n}_r}'$. When $\mathbf{G_r} = \frac{1}{(n_r-1)}(\mathbf{l}_{\mathbf{n}_r}\mathbf{l}_{\mathbf{n}_r}' - \mathbf{I}_{\mathbf{n}_r})$ and networks are of the same size such that $n_r = n/\bar{r}$, the model still can not be identified (see Moffitt, 2001). On the other hand, when $\mathbf{G_r} = \frac{1}{(n_r-1)}(\mathbf{l}_{\mathbf{n}_r}\mathbf{l}_{\mathbf{n}_r}' - \mathbf{I}_{\mathbf{n}_r})$ and there are variations in network sizes, Lee (2007) has shown that the model can be identified because the endogenous effect is stronger in small networks than in large networks. However, the identification can be weak when the all networks are large.

Row-normalization of $\mathbf{G_r}$ has some limitations. First, as in the structural model in this paper, one may be interested in the aggregate influence rather than average influence of the peers. Second, for some network structures, it is impossible to row normalize the adjacency matrix $\mathbf{G_r}$. For example, for an asymmetric $\mathbf{G_r}$, where agent i's outcome affects peers' outcomes but he/she is not affected by peers, the ith row of $\mathbf{G_r}$ would be all zeros. It would be impossible to normalize the ith row of $\mathbf{G_r}$ to sum to one. Finally, normalization may eliminate some useful information of the network structure. For the undirected friendship network, $\mathbf{G_r}$ will be a symmetric matrix. It should not be row-normalized because row-normalization would destroy the symmetry property.

Indeed, $\mathbf{G_r M_r l_{n_r}}$ is the measure of centrality in Bonacich (1987). The ith entry of $\mathbf{G_r M_r l_{n_r}}$ is the (weighted) sum of direct and indirect connections of agent i with others in the network. When $\mathbf{G_r}$ is not row-normalized, the entries of $\mathbf{G_r M_r l_{n_r}}$ in general is not all the same. The variation of this centrality measure in a network provides useful information for model identification. Even for the case that $\beta_0 = \gamma_0 = 0$, with non-row-normalized $\mathbf{G_r}$, $[\mathrm{E}(\mathbf{G_r y_r}), \mathbf{x_r}, \mathbf{G_r^* x_r}, \mathbf{l_{n_r}}] = [\eta_r \mathbf{G_r M_r l_{n_r}}, \mathbf{x_r}, \mathbf{G_r^* x_r}, \mathbf{l_{n_r}}]$ can still have full column rank. Hence, the model can be identified.

Under a certain regularity condition, $\mathbf{M_r} = \sum_{j=0}^{\infty}(\phi_0 \mathbf{G_r})^j$. It follows that $\mathbf{G_r M_r G_r x_r} = \sum_{j=0}^{\infty}(\phi_0 \mathbf{G_r})^j \mathbf{G_r^2 x_r}$, $\mathbf{G_r M_r G_r^* x_r} = \sum_{j=0}^{\infty}(\phi_0 \mathbf{G_r})^j \mathbf{G_r G_r^* x_r}$ and $\mathbf{G_r M_r l_{n_r}} = \sum_{j=0}^{\infty}(\phi_0 \mathbf{G_r})^j \mathbf{G_r l_{n_r}}$. Hence, from (19) we can use terms like $\mathbf{G_r l_{n_r}}$ as IVs for the endogenous effect in addition to the "traditional" IVs like $\mathbf{G_r x_n}$, $\mathbf{G_r^2 x_r}$ and/or $\mathbf{G_r G_r^* x_r}$ to help model identification and improve estimation efficiency (Liu and Lee, 2010).

Appendix 3: 2SLS and GMM estimators

2SLS Estimation From the reduced form equation (14), $\overline{\mathbf{Z}}$ =E($\mathbf{Z}$) = $[\mathbf{GM}(\mathbf{x}^*\delta_0 + \boldsymbol{\iota}\cdot\eta^*), \mathbf{x}^*]$. The best (in terms of efficiency) instrumental matrix for $\mathbf{JZ}$ in (13) is given by

$$\mathbf{F} = \mathbf{J}\overline{\mathbf{Z}} = \mathbf{J}[\mathbf{GMx}^*\delta_0 + \mathbf{GM}\boldsymbol{\iota}\eta^*, \mathbf{x}^*], \tag{20}$$

which is an $n \times (2m+1)$ matrix, where m is the dimension of $\mathbf{x}$. However, this instrumental matrix is infeasible as it involves unknown parameters δ_0 and η^*. Note that $\mathbf{F}$ can be considered as a linear combination of the IVs in $\mathbf{Q_0} = \mathbf{J}(\mathbf{GMx}^*, \mathbf{GM}\boldsymbol{\iota}, \mathbf{x}^*)$. Furthermore, as $\mathbf{M} = (\mathbf{I} - \phi_0\mathbf{G})^{-1} = \sum_{j=0}^{\infty}\phi_0^j\mathbf{G}^{j+1}$ when $|\phi_0\mu_1(\mathbf{G})| < 1$, $\mathbf{GMx}^*$ and $\mathbf{GM}\boldsymbol{\iota}$ can be approximated by linear combinations of $(\mathbf{Gx}^*, \mathbf{G}^2\mathbf{x}^*, \cdots)$ and $(\mathbf{G}\boldsymbol{\iota}, \mathbf{G}^2\boldsymbol{\iota}, \cdots)$ respectively, and, hence, $\mathbf{Q_0}$ can be approximated by a linear combination of $\mathbf{Q}_\infty = \mathbf{J}(\mathbf{Gx}^*, \mathbf{G}^2\mathbf{x}^*, \cdots, \mathbf{G}\boldsymbol{\iota}, \mathbf{G}^2\boldsymbol{\iota}, \cdots, \mathbf{x}^*)$.

For the estimation of (13), let $\mathbf{Q_K} = \mathbf{J}(\mathbf{G^{(p)}x}^*, \mathbf{G^{(p)}}\boldsymbol{\iota}, \mathbf{x}^*)$ be an $n \times K$ submatrix of $\mathbf{Q}_\infty$, where $\mathbf{G^{(p)}} = (\mathbf{G}, ..., \mathbf{G^p})$ for some p that increases as n increases. As $\boldsymbol{\iota}$ has $\bar{r}$ columns, the number of IVs in $\mathbf{Q_K}$ is large if the number of groups $\bar{r}$ is large. In general, more valid IVs would improve the efficiency of the estimator. However, the IV-based estimator could be asymptotically biased in the presence of many IVs.

Let $\mathbf{P_K} = \mathbf{Q_K}(\mathbf{Q_K'Q_K})^{-1}\mathbf{Q_K'}$. The many-IV 2SLS estimator is $\hat{\boldsymbol{\theta}}_{\mathbf{2sls}} = (\mathbf{Z'P_KZ})^{-1}\mathbf{Z'P_Ky}$. Let $\mathbf{e_1}$ denote the first column of an identity matrix. Liu and Lee (2010) have shown that, under some regularity assumptions, if $K/n \to 0$ then $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathbf{2sls}} - \boldsymbol{\theta}_{\mathbf{0}} - \mathbf{b_{2sls}}) \xrightarrow{d} N(0, \sigma_0^2\overline{\mathbf{h}}^{-1})$, where $\mathbf{b_{2sls}} = \sigma_0^2\mathrm{tr}(\mathbf{P_KGM})(\mathbf{Z'P_KZ})^{-1}\mathbf{e_1} = O_p(K/n)$. The term $\mathbf{b_{2sls}}$ is a bias due to the presence of many IVs. When $K^2/n \to 0$, the bias term $\sqrt{n}\mathbf{b_{2sls}}$ converges to zero so that $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathbf{2sls}} - \boldsymbol{\theta}_{\mathbf{0}}) \xrightarrow{d} N(0, \sigma_0^2(\lim_{n\to\infty}\frac{1}{n}\mathbf{F'F})^{-1})$. Hence, the sequence of IV matrices $\{\mathbf{Q_K}\}$ gives the asymptotically best IV estimator as the variance matrix attains the efficiency lower bound for the class of IV estimators.

To correct for the many-instrument bias in $\hat{\boldsymbol{\theta}}_{\mathbf{2sls}}$, we can adjust the many-IV 2SLS estimator by the estimated leading order bias. The bias-corrected many-IV 2SLS is given by $\hat{\boldsymbol{\theta}}_{\mathbf{c2sls}} = (\mathbf{Z'P_KZ})^{-1}\mathbf{Z'P_KY} - \hat{\mathbf{b}}_{\mathbf{2sls}}$, where $\hat{\mathbf{b}}_{\mathbf{2sls}}$ is a consistent estimator of b_{2sls}.[40] Liu and Lee (2010) have shown that, if $K/n \to 0$, then $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathbf{c2sls}} - \boldsymbol{\theta}_{\mathbf{0}}) \xrightarrow{d} N(0, \sigma_0^2(\lim_{n\to\infty}\frac{1}{n}\mathbf{F'F})^{-1})$.

Note that the number of IVs K is proportional to the number of groups $\bar{r}$. Hence, $K^2/n \to 0$ implies $\bar{r}^2/n = \bar{r}/\bar{m} \to 0$, where $\bar{m}$ is the average group size. So for asymptotic efficiency of the many-IV 2SLS estimator, the average group size needs to be large relative to the number of groups. On the other hand, $K/n \to 0$ implies $\bar{r}/n = 1/\bar{m} \to 0$. So for the

[40]For the explicit form of $\hat{\mathbf{b}}_{\mathbf{2sls}}$ (and that of $\hat{\mathbf{b}}_{\mathbf{gmm}}$ in the next Section), see Liu and Lee (2010).

bias-corrected many-IV 2SLS to be properly centered and asymptotically efficient, we only need the average group size to be large.

GMM Estimation The 2SLS approach can be generalized to the GMM with additional quadratic moment equations. While the IV moments use the information of the main regression function of the reduced form equation for estimation, the quadratic moments can explore the correlation structure of the reduced form disturbances. Let $\boldsymbol{\epsilon}(\theta) = \mathbf{J}(\mathbf{y} - \mathbf{Z}\theta)$ with $\theta = (\phi, \delta')'$. The IV moments $\mathbf{g_1}(\theta) = \mathbf{Q_K'}\boldsymbol{\epsilon}(\theta)$ are linear in $\boldsymbol{\epsilon}$ at $\boldsymbol{\theta_0}$. The quadratic moment is given by $\mathbf{g_2}(\theta) = \boldsymbol{\epsilon}'(\theta)\mathbf{U}\boldsymbol{\epsilon}(\theta)$ where $\mathbf{U} = \mathbf{JGMJ} - \text{tr}(\mathbf{JGM})\mathbf{J}/\text{tr}(\mathbf{J})$. At θ_0, $\text{E}[\mathbf{g_2}(\theta_0)] = 0$, because $\text{E}(\boldsymbol{\epsilon}'\mathbf{JUJ}\boldsymbol{\epsilon}) = \sigma_0^2\text{tr}(\mathbf{JU}) = 0$.[41] The vector of combined linear and quadratic empirical moments for the GMM estimation is given by $\mathbf{g}(\theta) = [\mathbf{g_1'}(\theta), \mathbf{g_2'}(\theta)]'$.

In order for asymptotic inference to be robust, we do not impose the normality assumption for the following results. For any $n \times n$ matrix $\mathbf{A} = [a_{ij}]$, let $\mathbf{A^s} = \mathbf{A} + \mathbf{A'}$ and $\text{vec}_D(\mathbf{A}) = (a_{11}, \cdots, a_{nn})'$. In general, μ_3 and μ_4 denote, respectively, the third and fourth moments of the error term. The variance matrix of $\mathbf{g}(\theta_0)$ is given by

$$\boldsymbol{\Omega} = \text{Var}[\mathbf{g}(\theta_0)] = \begin{pmatrix} \sigma_0^2\mathbf{Q_K'Q_K} & \mu_3\mathbf{Q_K'}\boldsymbol{\omega} \\ \mu_3\boldsymbol{\omega}'\mathbf{Q_K} & (\mu_4 - 3\sigma_0^4)\boldsymbol{\omega}'\boldsymbol{\omega} + \sigma_0^4\boldsymbol{\Delta} \end{pmatrix},$$

where $\boldsymbol{\omega} = \text{vec}_D(\mathbf{U})$ and $\boldsymbol{\Delta} = \frac{1}{2}\text{vec}(\mathbf{U^s})'\text{vec}(\mathbf{U^s})$. The optimal many-IV GMM estimator is given by $\hat{\boldsymbol{\theta}}_{\mathbf{gmm}} = \arg\min \mathbf{g}'(\theta)\boldsymbol{\Omega}^{-1}\mathbf{g}(\theta)$.

The optimal weighting matrix $\boldsymbol{\Omega}^{-1}$ involves unknown parameters σ_0^2, μ_3 and μ_4. In practice, with consistent initial estimators $\tilde{\sigma}^2$, $\tilde{\mu}_3$ and $\tilde{\mu}_4$, $\boldsymbol{\Omega}$ can be estimated as $\tilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega}(\tilde{\sigma}^2, \tilde{\mu}_3, \tilde{\mu}_4)$. Let $\mathbf{D_2} = \text{E}[\frac{\partial}{\partial\theta'}\mathbf{g}_2(\theta_0)] = -\sigma_0^2\text{tr}(\mathbf{U^sGM})\mathbf{e_1'}$, , where $\mathbf{e_1}$ is the first unit vector, and $\mathbf{B_{22}^{-1}} = (\mu_4 - 3\sigma_0^4)\boldsymbol{\omega}'\boldsymbol{\omega} + \sigma_0^4\boldsymbol{\Delta} - \frac{\mu_3^2}{\sigma_0^2}\boldsymbol{\omega}'\mathbf{P_K}\boldsymbol{\omega}$. Liu and Lee (2010) have shown that, if $K^{3/2}/n \to 0$, the feasible optimal many-IV GMM estimator $\hat{\boldsymbol{\theta}}_{\mathbf{gmm}} = \arg\min_{\theta\in\Theta} \mathbf{g}'(\theta)\tilde{\boldsymbol{\Omega}}^{-1}\mathbf{g}(\theta)$ has the asymptotic distribution

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathbf{gmm}} - \boldsymbol{\theta_0} - \mathbf{b_{gmm}}) \xrightarrow{d} N(0, [\sigma_0^{-2}(\lim_{n\to\infty}\frac{1}{n}\mathbf{F'F})^{-1} + \lim_{n\to\infty}\frac{1}{n}\bar{\mathbf{D}}_\mathbf{2}'\mathbf{B_{22}}\bar{\mathbf{D}}_\mathbf{2}]^{-1}), \quad (21)$$

where $\mathbf{b_{gmm}} = (\sigma^{-2}\mathbf{Z'P_KZ} + \check{\mathbf{D}}_\mathbf{2}'\mathbf{B_{22}}\check{\mathbf{D}}_\mathbf{2})^{-1}\text{tr}(\boldsymbol{\Psi}_\mathbf{K})\mathbf{e_1} = O(K/n)$, $\check{\mathbf{D}}_\mathbf{2} = \mathbf{D_2} - \frac{\mu_3}{\sigma_0^2}\boldsymbol{\omega}'\mathbf{P_KZ}$, and $\bar{\mathbf{D}}_\mathbf{2} = \mathbf{D_2} - \frac{\mu_3}{\sigma_0^2}\boldsymbol{\omega}'\mathbf{F}$.

As the asymptotic bias $\mathbf{b_{gmm}}$ is $O(K/n)$, the asymptotic distribution of the GMM estimator will be centered at $\boldsymbol{\theta_0}$ only when $K^2/n \to 0$. With the consistently estimated leading

[41] Liu and Lee (2010) have shown that the quadratic moment $\mathbf{g_2}(\theta) = \boldsymbol{\epsilon}'(\theta)\mathbf{U}\boldsymbol{\epsilon}(\theta)$ is the best (in terms of efficiency of the GMM estimator) under normality.

order bias $\hat{\mathbf{b}}_{\mathbf{gmm}}$, Liu and Lee (2010) have shown that, if $K^{3/2}/n \to 0$, the feasible bias-corrected many-IV GMM estimator $\hat{\boldsymbol{\theta}}_{\mathbf{cgmm}} = \hat{\boldsymbol{\theta}}_{\mathbf{gmm}} - \hat{\mathbf{b}}_{\mathbf{gmm}}$ is properly centered and has the asymptotic normal distribution as given in (21).

The asymptotic variance matrix of the many-IV GMM estimator can be compared with that of the many-IV 2SLS estimator. As $\bar{\mathbf{D}}_{\mathbf{2}}'\mathbf{b}_{\mathbf{22}}\bar{\mathbf{D}}_{\mathbf{2}}$ is nonnegative definite, the asymptotic variance of the many-IV GMM estimator is relatively smaller than that of the 2SLS estimator. The many-IV GMM estimator with additional quadratic moments improves efficiency upon the 2SLS estimator.

Appendix 4: Individual centrality measures and network characteristics

The simplest index of connectivity of individual i in network r is the number of direct friends divided by the maximum possible number of friends individual i can have (i.e. $n_r - 1$ individuals if everyone is directly connected to individual i), i.e. *degree centrality*:

$$\delta_{i,r}(g_r) = \frac{g_{i,r}}{n_r - 1} = \frac{\sum_{j=1}^{n_r} g_{ij,r}}{n_r - 1} \tag{22}$$

The standard measure of *closeness centrality* of individual i in network r is given by:

$$c_{i,r}(g_r) = \frac{n_r - 1}{\sum_j d_r(i,j)} \tag{23}$$

where $d_r(i,j)$ is the geodesic distance (length of the shortest path)[42] in network r between individuals i and j. As a result, the closeness centrality of individual i is the inverse of the sum of geodesic distances from i to the $n_r - 1$ other individuals (i.e. the reciprocal of its "farness") divided by $n_r - 1$, which is the maximum possible distance between two individuals in the network . Compared to degree centrality, the closeness measure takes into account not only direct connections among individuals but also indirect connections. However, compared to the Bonacich centrality, the closeness measure assumes a weight of one to each indirect connection, whereas the Bonacich centrality uses weights that depend on the strength of social interaction within the network.

The *betweenness centrality* measure of agent i in a network g_r can be defined as:

$$f_{i,r}(g_r) = \frac{1}{(n_r - 1)(n_r - 2)/2} \sum_{j,l}^{n_r} \frac{a_{jl,r}(i)}{a_{jl,r}} \tag{24}$$

where j and l denote two given agents in g_r, $a_{jl,r}(i)$ is the number of shortest paths between j and l through i in g_r, $a_{jl,r}$ is the number of shortest paths between j and l in g_r and $(n_r - 1)(n_r - 2)/2$ is the total number of links in a complete network.[43] Note that betweenness centrality, as the degree and closeness centrality measures, is a parameter-free index while the Bonacich centrality is not since it depends on the decay factor ϕ.

[42] The length of a shortest path is the smallest k such that there is at least one path of length k from i to j. Therefore we can find the length by computing $\mathbf{G}_r$, $\mathbf{G}_r^2$, $\mathbf{G}_r^3$, ..., until we find the first k such that the (i,j)th entry of $\mathbf{G}_r^k$ is not zero.

[43] Formula (24) is only true for undirected networks. For directed networks, a similar formula can be used but it has to be divided by $(n_r - 1)(n_r - 2)$ instead of $(n_r - 1)(n_r - 2)/2$.

The *clustering coefficient* of individual i in network r is given by:

$$\psi_{i,r}(g_r) = \frac{\sum_{l \in N_{i,r}(g_r)} \sum_{k \in N_{i,r}(g_r)} g_{lk,r}}{n_{i,r}(g_r)\left[n_{i,r}(g_r) - 1\right]} \qquad \text{for all } i \in \{i \in N_r \mid n_{i,r}(g_r) \geq 2\} \tag{25}$$

where N_r is the set of nodes in network g_r, $N_{i,r}(g_r) = \{j \neq i \mid g_{ij,r} = 1\}$ is the set of i's direct contacts and $n_{i,r}(g_r)$, it size (or cardinality of this set). $\psi_{i,r}(g_r)$ gives us the percentage of an individual's links who are linked to each other. This is an indication of the percentage of transitive triads[44] around individual i. It thus measures the probability with which two of i's friends are also friends.

Unit centralities in a network can have large or small variance. Network, where one unit (or a low number of units) has (have) much higher centrality than other units is highly centralized. On the other hand, if unit centrality measures do not differ significantly, the centrality of a network is low.

From these individual measures we can compute the corresponding measures at the network level using the definition provided by Freeman (1979). In our notation, the Freeman (1979)'s general network index for a given network g_r is

$$C_r^A(g_r) = \frac{\sum_{i=1}^{n_r}(C_{i^*,r}^A - C_{i,r}^A)}{\max \sum_{i=1}^{n_r}(C_{i^*,r}^A - C_{i,r}^A)}$$

where $C_{i^*,r}^A$ is the largest value of $C_{i,r}^A$ for any individual in the network and $\max \sum_{i=1}^{n_r}(C_{i^*,r}^A - C_{i,r}^A)$ is the maximum possible sum of differences in unit centrality for a network of n individuals. The network index is thus a number between 0 and 1, being 0 if all units have equal value, and 1, when one unit completely dominates all other units. Our four individual measures, then lead to four network properties, namely degree, closeness and betweenness network centrality and network clustering. Let us finally revise other widely used network characteristics.

The *average distance* of a network (also known as the *average path length*) is defined as the average number of steps along the shortest paths for all possible pairs of network nodes (i.e. $\sum_i \sum_j d_r(i,j) / [n_r(n_r - 1)]$).

The *average degree* is the total number of links divided by n_r (i.e. $\sum_i g_{i,r}/n_r$).[45]

The *diameter* of a network is the largest (shortest) distance between any two nodes in the network. It thus provides an upper-bound measure of the size of the network.

[44] A triad is the subgraph on three individuals, so that when studying triads, one has to consider the threesome of individuals and all the links between them. A triad involving individuals i, j, k is transitive if whenever $i \to j$ and $j \to k$, then $i \to k$.

[45] Remember that $g_{i,r} = \sum_{j=1}^{n_r} g_{ij,r}$ is the degree (i.e. the number of direct friends) of criminal i.

Network *density* is simply the fraction of ties present in a network over all possible ones (it is the average degree divided by $n_r - 1$). It ranges from 0 to 1 as networks get denser.

Network *asymmetry* is measured using the variance of connectivities (i.e. $(\frac{\max_i g_{i,r}}{\min_i g_{i,r}})/(n_r - 1)$). We normalize it, so that it reaches 1 for the most asymmetric network in the sample.

Network redundancy or *clustering* is the fraction of all transitive triads over the total number of triads. It measures the probability with which two of i's friends know each other.

Finally, network *assortativity* measures the correlation patterns among high-degree nodes. If high-degree nodes tend to be connected to other high-degree nodes, then the network is said to be positive assortative. The degree of assortativity of the network g_r is computed as: $\sum_i \sum_j (g_{i,r} - m_r)(g_{j,r} - m_r) / \sum_i (g_{i,r} - m_r)^2$, where m_r is the average degree in network r (i.e. $\sum_i g_{i,r}/n_r$).

Appendix 5: Key Player: complete list of results

Table A.2: Features of the Key Player – All crimes –

		Individual Characteristics		Friend Group Characteristics	
	Range	Mean	St.dev	Mean	St.dev
Socio-demographic variables					
Female	0-1	0.32	0.47	0.50	0.41
Religion practice	1-5	2.35	1.48	2.46	1.25
Student grade	7-12	9.15	1.59	9.18	1.49
Black or African American	0-1	0.19	0.39	0.19	0.38
Other races	0-1	0.07	0.25	0.06	0.22
Mathematics score	1-4	2.29	1.10	2.24	0.90
Self esteem	1-6	3.97	1.16	3.95	0.82
Physical development	1-5	3.45	1.23	3.32	0.80
Family background variables					
Household size	2-11	4.46	1.26	4.55	1.20
Two married parent family	0-1	0.66	0.48	0.70	0.41
Single parent family	0-1	0.27	0.44	0.24	0.38
Parent education	0-5	3.08	1.17	3.16	0.99
Parent occupation manager	0-1	0.15	0.36	0.10	0.25
Parent occupation professional/technical	0-1	0.17	0.38	0.19	0.35
Parent occupation office or sales worker	0-1	0.11	0.31	0.08	0.20
Parent occupation manual	0-1	0.29	0.45	0.32	0.42
Parent occupation military or security	0-1	0.01	0.08	0.03	0.15
Parent occupation farm or fishery	0-1	0.01	0.12	0.03	0.14
Parent occupation other	0-1	0.17	0.37	0.18	0.34
Protective factors					
School attachment	1-5	2.10	0.94	2.03	0.91
Relationship with teachers	0-4	1.38	1.03	1.01	0.84
Social inclusion	1-5	4.28	0.82	4.54	0.63
Parental care	0-1	0.83	0.37	0.93	0.22
Residential neighborhood variables					
Residential building quality	1-4	1.60	0.87	1.58	0.74
Residential area suburban	0-1	0.38	0.49	0.40	0.45
Residential area urban - residential only	0-1	0.23	0.42	0.25	0.41
Residential area industrial properties - mostly wholesale	0-1	0.00	0.00	0.00	0.00
Residential area other type	0-1	0.05	0.21	0.00	0.04

Table A.3: Features of the Key Player – Petty crimes –

		Individual Characteristics		Friend Group Characteristics	
	Range	Mean	St.dev	Mean	St.dev
Socio-demographic variables					
Female	0-1	0.42	0.50	0.51	0.44
Religion practice	1-5	2.25	1.34	2.36	1.29
Student grade	7-12	9.23	1.55	9.13	1.43
Black or African American	0-1	0.16	0.37	0.17	0.37
Other races	0-1	0.05	0.23	0.08	0.24
Mathematics score	1-4	2.40	1.07	2.10	0.88
Self esteem	1-6	4.01	1.02	3.94	0.93
Physical development	1-5	3.46	1.18	3.37	0.91
Family background variables					
Household size	2-11	4.53	1.38	4.44	1.24
Two married parent family	0-1	0.23	0.42	0.69	0.43
Single parent family	0-1	0.71	0.45	0.25	0.40
Parent education	0-5	3.14	1.09	3.17	1.05
Parent occupation manager	0-1	0.16	0.37	0.11	0.27
Parent occupation professional/technical	0-1	0.18	0.39	0.20	0.37
Parent occupation office or sales worker	0-1	0.14	0.34	0.05	0.18
Parent occupation manual	0-1	0.30	0.46	0.30	0.43
Parent occupation military or security	0-1	0.01	0.09	0.02	0.13
Parent occupation farm or fishery	0-1	0.03	0.17	0.01	0.09
Parent occupation other	0-1	0.11	0.32	0.19	0.37
Protective factors					
School attachment	1-5	2.05	0.97	1.99	0.90
Relationship with teachers	0-4	1.28	1.05	0.94	0.81
Social inclusion	1-5	4.21	0.85	4.52	0.61
Parental care	0-1	0.80	0.40	0.91	0.26
Residential neighborhood variables					
Residential building quality	1-4	1.56	0.87	1.52	0.74
Residential area suburban	0-1	0.38	0.49	0.43	0.47
Residential area urban - residential only	0-1	0.26	0.44	0.22	0.39
Residential area industrial properties - mostly wholesale	0-1	0.00	0.00	0.00	0.00
Residential area other type	0-1	0.02	0.12	0.02	0.15

Table A.4: Features of the Key Player – More serious crimes –

		Individual Characteristics		Friend Group Characteristics	
	Range	Mean	St.dev	Mean	St.dev
Socio-demographic variables					
Female	0-1	0.32	0.47	0.47	0.41
Religion practice	1-5	2.47	1.48	2.38	1.15
Student grade	7-12	9.05	1.58	9.16	1.47
Black or African American	0-1	0.29	0.46	0.33	0.46
Other races	0-1	0.07	0.25	0.06	0.23
Mathematics score	1-4	2.48	1.08	2.44	0.93
Self esteem	1-6	3.87	1.27	3.88	0.95
Physical development	1-5	3.60	1.26	3.35	0.83
Family background variables					
Household size	2-11	4.61	1.40	4.60	1.17
Two married parent family	0-1	0.27	0.45	0.64	0.42
Single parent family	0-1	0.64	0.48	0.29	0.39
Parent education	0-5	3.24	1.21	2.96	1.06
Parent occupation manager	0-1	0.12	0.33	0.09	0.25
Parent occupation professional/technical	0-1	0.16	0.37	0.19	0.36
Parent occupation office or sales worker	0-1	0.16	0.37	0.12	0.24
Parent occupation manual	0-1	0.27	0.45	0.31	0.42
Parent occupation military or security	0-1	0.00	0.00	0.00	0.00
Parent occupation farm or fishery	0-1	0.01	0.12	0.01	0.04
Parent occupation other	0-1	0.13	0.34	0.16	0.32
Protective factors					
School attachment	1-5	2.25	1.07	2.03	0.74
Relationship with teachers	0-4	1.52	1.39	1.16	0.90
Social inclusion	1-5	4.44	0.79	4.33	0.77
Parental care	0-1	0.88	0.33	0.90	0.25
Residential neighborhood variables					
Residential building quality	1-4	1.71	0.93	1.65	0.72
Residential area suburban	0-1	0.45	0.50	0.34	0.43
Residential area urban - residential only	0-1	0.19	0.39	0.31	0.43
Residential area industrial properties - mostly wholesale	0-1	0.00	0.00	0.00	0.00
Residential area other type	0-1	0.05	0.23	0.01	0.06

Table 2a: Model (11) Estimation Results for Undirected Networks

	Total crimes	Type 1 Crimes	Type 2 Crimes
2SLS finite IVs	0.067 (3.233)	0.06 (3.043)	0.097 (2.534)
2SLS large IVs	0.047 (2.549)	0.031 (1.733)	0.068 (1.997)
bias-corrected 2SLS	0.072 (3.945)	0.053 (2.901)	0.128 (3.677)
GMM finite IVs	0.056 (4.12)	0.042 (3.136)	0.097 (3.773)
GMM large IVs	0.045 (3.518)	0.03 (2.27)	0.072 (2.899)
bias-corrected GMM	0.052 (4.043)	0.036 (2.783)	0.08 (3.239)

Notes: Estimation has been performed using Matlab. T-tests are reported in parentheses.

Table 2b: Model (11) Estimation Results for Directed Networks

	Total crimes	Type 1 Crimes	Type 2 Crimes
2SLS finite IVs	0.097 (3.044)	0.089 (3.047)	0.189 (2.992)
2SLS many IVs	0.059 (2.521)	0.055 (2.381)	0.098 (2.191)
bias-corrected 2SLS	0.090 (3.854)	0.080 (3.470)	0.172 (3.833)
GMM finite IVs	0.089 (4.252)	0.074 (3.672)	0.188 (4.716)
GMM many IVs	0.072 (3.944)	0.059 (3.281)	0.114 (3.255)
bias-corrected GMM	0.088 (4.862)	0.072 (4.032)	0.144 (4.131)

Notes: Estimation has been performed using Matlab. T-tests are reported in parentheses.

Table 3a: Model (16) Estimation results for undirected networks

Dependent variable=1 if students *i* and *j* are friends and =0 otherwise	β	γ_1	γ_2
Female	-0.0195***	-0.0068	0.1518***
	(0.0048)	(0.0539)	(0.0512)
Religion practice	-0.0058***	-0.0013	0.0107
	(0.0020)	(0.0167)	(0.0168)
Student grade	-0.0386***	0.0435*	-0.0084
	(0.0020)	(0.0242)	(0.0180)
Black or African American	-0.0744***	0.0328	0.0340
	(0.0093)	(0.0756)	(0.0262)
Other races	-0.0201	-0.0133	-0.0242
	(0.0127)	(0.0335)	(0.0442)
Mathematics score	-0.0067**	-0.0177	0.0194
	(0.0027)	(0.0246)	(0.0293)
Self esteem	-0.0026	-0.0022	0.0082
	(0.0025)	(0.0167)	(0.0120)
Physical development	0.0003	-0.0167	0.0295
	(0.0018)	(0.0201)	(0.0198)
Household size	-0.0019	0.0001	0.0049
	(0.0019)	(0.0117)	(0.0144)
Two married parent family	-0.0113	-0.0890	0.0473
	(0.0074)	(0.0908)	(0.0908)
Parent education	-0.0038	0.0097	0.0131
	(0.0024)	(0.0111)	(0.0122)
Single parent family	0.0145**	0.1121	-0.1426*
	(0.0065)	(0.0776)	(0.0768)
Residential building quality	-0.0027	-0.0146	-0.0056
	(0.0023)	(0.0164)	(0.0211)
School attachment	-0.0031	-0.0336**	0.0226
	(0.0031)	(0.0163)	(0.0175)
Trouble relationship with teachers	-0.0035	0.0018	0.0015
	(0.0022)	(0.0204)	(0.0133)
Social inclusion	-0.0101***	-0.0044	0.0035
	(0.0025)	(0.0222)	(0.0183)
Parental care	0.0006	-0.0011	-0.0108
	(0.0048)	(0.0379)	(0.0409)
Constant		0.2130***	
		(0.0097)	
Observations		15093	
Number of networks		150	
R-squared		0.048	

Note. Obervations are all pairwise combinations of students across networks for total crime. A linear probability model is estimated via least squares with network fixed effects. Regressions also include parental occupation dummies and residential area dummies. Parameter estimates and bootstrapped standard errors (in parentheses) are reported. *** p<0.01, ** p<0.05, * p<0.1

Table3b: Model (17) Estimation results for Directed Networks

Dependent variable=1 if students *i* and *j* are friends and =0 otherwise	β	γ
Female	-0.0181***	0.0524*
	(0.0021)	(0.0275)
Religion practice	-0.0037***	-0.0121
	(0.0012)	(0.0084)
Student grade	-0.0235***	-0.0030
	(0.0009)	(0.0118)
Black or African American	-0.0446***	0.0013
	(0.0052)	(0.0275)
Other races	-0.0137*	-0.0480**
	(0.0082)	(0.0216)
Mathematics score	-0.0040**	0.0065
	(0.0018)	(0.0146)
Self esteem	-0.0026**	0.0067
	(0.0011)	(0.0078)
Physical development	-0.0001	0.0042
	(0.0010)	(0.0099)
Household size	-0.0020	0.0073
	(0.0012)	(0.0061)
Two married parent family	-0.0074	-0.0319
	(0.0045)	(0.0428)
Parent education	-0.0026**	0.0081
	(0.0012)	(0.0077)
Single parent family	0.0104**	0.0661
	(0.0045)	(0.0456)
Residential building quality	-0.0023	-0.0046
	(0.0015)	(0.0119)
School attachment	-0.0015	-0.0021
	(0.0016)	(0.0099)
Trouble relationship with teachers	-0.0035***	0.0159
	(0.0013)	(0.0138)
Social inclusion	-0.0059***	-0.0051
	(0.0016)	(0.0136)
Parental care	0.0012	-0.0025
	(0.0044)	(0.0257)
Constant	0.1338***	
	(0.0059)	
Observations	30186	
Number of networks	150	
R-squared	0.027	

Notes. Obervations are all pairwise combinations of students across networks for total crime. A linear probability model is estimated via least squares with network fixed effects. Regressions also include parental occupation dummies and residential area dummies. Parameter estimates and bootstrapped standard errors (in parentheses) are reported. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Who is the Key Player?
-Significant Differences-
All crimes

	All Criminals		Key Player Criminals		
	Mean	St. dev	Mean	St. dev	t-test
Individual characteristics					
Female	0.51	0.50	0.32	0.47	0.0000
Parent education	3.24	1.08	3.08	1.17	0.1025
Parent occupation military or security	0.02	0.15	0.007	0.08	0.0577
Residential area other type	0.01	0.11	0.05	0.21	0.0459
School attachment	1.92	0.92	2.10	0.94	0.0265
Trouble relationship with teachers	1.04	1.00	1.38	1.03	0.0002
Social inclusion	4.48	0.74	4.28	0.82	0.0102
Parental care	0.93	0.26	0.83	0.38	0.0350
Friends' characteristics					
Religious practice	2.25	1.21	2.46	1.25	0.0606
Student grade	8.97	1.47	9.18	1.49	0.1010
Residential area other type	0.02	0.12	0.004	0.04	0.0017
N.obs.	1147		150		

Notes: T-test for differences in means with unequal variances had been performed. P-values are reported

Table 5: Key Player versus Bonacich centrality
-Significant Differences-
All crimes

	Key Player Most Active Criminal		Key Player Not the Most Active Criminal		
	Mean	St. dev	Mean	St. dev	t-test
Individual characteristics					
Religion practice	2.47	1.53	2	1.26	0.0591
Mathematics Score	2.41	1.08	1.97	1.10	0.0349
Physical development	3.55	1.27	3.17	1.08	0.0704
Single parent family	0.30	0.46	0.17	0.38	0.0995
Residential area suburban	0.34	0.47	0.50	0.51	0.0798
Trouble relationship with teachers	1.51	1.02	1.02	0.97	0.0097
Social inclusion	4.20	0.81	4.50	0.82	0.0501
Friends' characteristics					
Religion practice	2.58	1.30	2.13	1.02	0.0294
Other races	0.08	0.24	0.02	0.10	0.0241
Parental education	3.07	1.03	3.39	0.83	0.0585
Parent occupation manual	0.35	0.43	0.23	0.37	0.0936
Residential building quality	1.51	0.69	1.77	0.84	0.0897
Residential area suburban	0.44	0.45	0.30	0.42	0.0777
Parental care	0.91	0.25	0.99	0.04	0.0009
N.obs.	110		40		

Notes: T-test for differences in means with unequal variances had been performed. P-values are reported

Table 6: Who is the Key Player?
-Significant Differences-
Petty crimes

	All Criminals		Key Player Criminals		
	Mean	St. dev	Mean	St. dev	t-test
Individual characteristics					
Female	0.52	0.50	0.42	0.49	0.0272
Blacks or African American	0.22	0.41	0.16	0.37	0.1009
Mathematics score	2.13	0.98	2.40	1.07	0.0057
Trouble relationship with teachers	1.04	0.98	1.28	1.05	0.0155
Social inclusion	4.48	0.74	4.21	0.85	0.0006
Parental care	0.94	0.25	0.80	0.40	0.0004
Friends' characteristics					
Parent occupation office or sales worker	0.10	0.22	0.05	0.18	0.0111
Parent occupation farm or fishery	0.02	0.12	0.008	0.09	0.0994
Residential area suburban	0.35	0.41	0.43	0.47	0.0559
Trouble relationship with teachers	1.08	0.81	0.94	0.80	0.0715
N.obs.	967		132		

Notes: T-test for differences in means with unequal variances had been performed. P-values are reported

Table 7: Who is the Key Player?
-Significant Differences-
More serious crimes

	All Criminals		Key Player Criminals		
	Mean	St. dev	Mean	St. dev	t-test
Individual characteristics					
Female	0.42	0.49	0.32	0.50	0.0818
Physical development	3.33	1.07	3.6	1.26	0.0876
Parent occupation military or security	0.01	0.11	0.00	0.00	0.0141
Residential area suburban	0.29	0.45	0.45	0.50	0.0100
Residential area urban-residential only-	0.31	0.46	0.19	0.39	0.0119
School attachment	2.03	0.98	2.25	1.07	0.0978
Trouble relationship with teachers	1.24	1.06	1.52	1.39	0.0939
Friends' characteristics					
Student grade	8.85	1.45	9.16	1.47	0.0890
Parent occupation military or security	0.006	0.05	0.00	0.00	0.0030
Parent occupation farm or fishery	0.02	0.10	0.006	0.04	0.1098
Residential area other type	0.03	0.14	0.007	0.06	0.0303
N.obs.	470		75		

Notes: T-test for differences in means with unequal variances had been performed. P-values are reported

Table 8: Key Player for Petty and Serious Crimes
-Significant Differences-

	Key Player Petty Crime		Key Player More Serious Crime		
	Mean	St. dev	Mean	St. dev	t-test
Individual characteristics					
Black or African American	0.16	0.37	0.30	0.46	0.0318
Social inclusion	4.21	0.85	4.44	0.79	0.0543
Friends' characteristics					
Black or African American	0.17	0.37	0.32	0.46	0.0146
Mathematics score	2.10	0.88	2.44	0.93	0.0130
Parent occupation office or sales worker	0.05	0.18	0.12	0.24	0.0560
Parent occupation military or security	0.02	0.14	0.00	0.00	0.0472
Residential area urban-residential only-	0.21	0.39	0.31	0.43	0.1039
Trouble relationship with teachers	0.94	0.80	1.16	0.90	0.0913
Social inclusion	4.51	0.61	4.33	0.77	0.0825
N.obs.	132		75		

Notes: T-test for differences in means with unequal variances had been performed. P-values are reported

Table 9: Key Player versus Bonacich centrality
-Significant Differences-
Petty crimes

	Key Player Most Active Criminal		Key Player Not the Most Active Criminal		
	Mean	St. dev	Mean	St. dev	t-test
Individual characteristics					
Religion practice	2.40	1.39	1.90	1.16	0.0368
Physical development	3.60	1.22	3.13	1.03	0.0251
Parent education	3.27	1.06	2.82	1.10	0.0342
Parent occupation manager	0.20	0.40	0.05	0.22	0.0065
Relationship with teachers	1.46	1.04	0.85	0.96	0.0016
Friends' characteristics					
Parent occupation office or sales worker	0.07	0.21	0.02	0.09	0.0826
Residential area suburban	0.47	0.48	0.33	0.45	0.1057
N.obs.	93		39		

Notes: T-test for differences in means with unequal variances had been performed. P-values are reported

Table 10: Key Player versus Bonacich centrality
-Significant Differences-
More Serious crimes

	Key Player Most Active Criminal		Key Player Not the Most Active Criminal		
	Mean	St. dev	Mean	St. dev	t-test
Individual characteristics					
Other races	0.09	0.29	0.00	0.00	0.0240
Household size	4.43	1.25	5.16	1.71	0.0993
Residential building quality	1.80	1.00	1.42	0.61	0.0526
Residential area suburban	0.39	0.49	0.63	0.49	0.0789
Residential area other type	0.07	0.26	0.00	0.00	0.0445
Friends' characteristics					
Two married parent family	0.69	0.42	0.49	0.40	0.0788
Single parent family	0.25	0.37	0.43	0.41	0.0988
Residential area urban-residential only-	0.37	0.46	0.15	0.28	0.0167
Parental care	0.88	0.28	0.97	0.09	0.0406
N.obs.	56		19		

Notes: T-test for differences in means with unequal variances had been performed. P-values are reported

Table 11: Key Players and network topology
All crimes

	Betweenness		Closeness		Clustering		Bonacich	
percentiles								
p50	0		0.50		0		2.16	
p75	0.50		0.67		0		3.32	
p90	0.67		0.75		0.27		4.70	
p95	0.73		0.83		0.50		5.58	
min	0		0.17		0		0.13	
max	1		1		1		9.63	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
>p90	4.5%	10%	11%	5%	10%	10%	14%	0%
>p95	4.5%	5%	4.5%	5%	4.5%	2.5%	6.4%	0%

(1) Key Players Most Active Criminals; (2) Key Players Not the Most Active Criminals

Table 12: Key Players and network topology
Petty crimes

	Betweenness		Closeness		Clustering		Bonacich	
percentiles								
p50	0		0.50		0		2.18	
p75	0.05		0.60		0		3.80	
p90	0.53		0.75		0.33		5.18	
p95	0.67		0.80		1		5.75	
min	0		0.13		0		0.20	
max	1		1		1		7.31	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
>p90	13%	2.5%	8.6%	7.6%	7.5%	0%	13%	2.6%
>p95	3.2%	0%	6.4%	5.1%	2.1%	0%	6.5%	0%

(1) Key Players Most Active Criminals; (2) Key Players Not the Most Active Criminals

Table 13: Key Players and network topology
More serious crimes

	Betweenness		Closeness		Clustering		Bonacich	
percentiles								
p50	0		0.50		0		2.45	
p75	0.67		0.75		0		4.53	
p90	0.67		0.75		0.33		5.61	
p95	0.69		1		0.33		6.48	
min	0		0.20		0		0.34	
max	1		1		1		12.55	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
>p90	1.8%	16%	5.3%	16%	3.6%	5.3%	12.5%	0%
>p95	1.8%	10%	3.6%	10%	3.6%	5.3%	5.4%	0%

(1) Key Players Most Active Criminals; (2) Key Players Not the Most Active Criminals

Table 14: Key Players and network topology
All crimes

	Key Player Most Active Criminal		Key Player Not the Most Active Criminal		
	Mean	St. dev	Mean	St. dev	t-test
Network characteristics					
Diameter	3.84	2.33	4.20	2.75	0.4701
Average distance	2.05	0.82	2.16	0.96	0.5183
Average degree	1.81	0.46	1.80	0.43	0.9043
Density	0.42	0.12	0.42	0.14	0.9074
Asymmetry	0.67	0.25	0.64	0.24	0.4750
Network clustering	0.10	0.20	0.10	0.18	0.9328
Network degree	0.13	0.10	0.11	0.09	0.3421
Network closeness	0.54	0.26	0.50	0.23	0.3511
Assortativity	9.30×10^{-18}	1.58×10^{-16}	6.03×10^{-17}	3.28×10^{-16}	0.2037
Network betweeness-	3.36	3.51	4.14	5.01	0.3723
N.obs.	110		40		

Table 15: Key Players and network topology
Petty crimes

	Key Player Most Active Criminal		Key Player Not the Most Active Criminal		
	Mean	St. dev	Mean	St. dev	t-test
Network characteristics					
Diameter	3.91	2.25	3.84	2.43	0.8817
Average distance	2.06	0.78	2.03	0.81	0.8484
Average degree	1.82	0.47	1.75	0.38	0.3719
Density	0.42	0.13	0.43	0.12	0.5053
Asymmetry	0.65	0.24	0.67	0.22	0.6335
Network clustering	0.10	0.19	0.09	0.19	0.7937
Network degree	0.12	0.10	0.12	0.08	0.9292
Network closeness	0.53	0.25	0.53	0.22	0.9894
Assortativity	-1.54×10^{-17}	3.24×10^{-17}	-4.39×10^{-17}	6.14×10^{-17}	0.6836
Network betweeness	3.69	4.39	3.27	3.30	0.5497
N.obs.	93		39		

Table 16: Key Players and network topology
More serious crimes

	Key Player Most Active Criminal		Key Player Not the Most Active Criminal		
	Mean	St. dev	Mean	St. dev	t-test
Network characteristics					
Diameter	3.91	2.01	3.42	1.46	0.2616
Average distance	2.07	0.73	1.92	0.61	0.3710
Average degree	1.79	0.41	1.66	0.21	**0.0849**
Density	0.42	0.13	0.46	0.10	0.1624
Asymmetry	0.65	0.23	0.69	0.21	0.5318
Network clustering	0.09	0.17	0.09	0.19	0.9753
Network degree	0.12	0.10	0.13	0.08	0.7559
Network closeness	0.52	0.24	0.53	0.22	0.8513
Assortativity	7.64×10^{-18}	1.37×10^{-16}	-1.24×10^{-17}	1.22×10^{-16}	0.5535
Network betweeness	3.63	3.62	2.72	2.12	0.1933
N.obs.	56		19		