Leaders in Juvenile Crime*

Carlos Díaz† Eleonora Patacchini‡ Thierry Verdier§ Yves Zenou¶

October 5, 2020 (Revised October 23, 2021)

Abstract

This paper presents a new theory of crime where leaders transmit a crime technology and act as a role model for other criminals. We show that, in equilibrium, an individual's crime effort and criminal decisions depend on the geodesic distance to the leader in his or her network of social contacts. By using data on friendship networks among U.S. high-school students, we structurally estimate the model and find evidence supporting its predictions. In particular, by using a definition of a criminal leader that is exogenous to the network formation of friendship links, we find that the longer is the distance to the leader, the lower is the criminal activity of the delinquents and the less likely they are to become criminals. We finally perform a counterfactual experiment that reveals that a policy that removes all criminal leaders from a school can, on average, reduce criminal activity by about 20% and the individual probability of becoming a criminal by 10%.

Key words: Crime leaders, social distance, criminal decisions, closeness centrality.

JEL classification: C31, D85, K42.

---

*We thank the editor, Sudipta Sarangi, two anonymous referees, and Michelle Rendall for very helpful comments. Thierry Verdier acknowledges financial support from the European Research Council grant number 324004 (TECTACOM project).

†Universidad Católica del Uruguay. E-mail: carlos.diaz@ucu.edu.uy.

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

§Paris School of Economics, ENPC ParisTech, PUC-Rio, and CEPR. E-mail: thierry.verdier@psemail.eu.

¶Monash University, and CEPR. E-mail: yves.zenou@monash.edu.

1 Introduction

Leaders play an important role in various activities. Religious leaders can influence the identity of their community (Hauk and Mueller, 2015; Prummer and Siedlarek, 2017; Verdier and Zenou, 2015, 2018; Prummer, 2019), business leaders affect the way in which companies are run and organized (Mendenhall et al., 2001), and gang crime leaders are responsible for more than two-thirds of their co-offenders' first-time convictions (Reiss and Farrington, 1991). However, economics research that aims to understand the mechanisms through which leaders influence their peers remains scarce. Of this limited literature, some theoretical works have recently studied the role of leaders in organizations. In these models, leaders act as coordination devices (Van den Steen, 2010; Myerson, 2011; Bolton et al., 2012; Hermalin, 2012) or as prominent and visible agents anchoring and changing social norms and beliefs on appropriate outcomes (Acemoglu and Jackson, 2014).

To build on the current body of knowledge on this topic, this study examines the impact of leaders on juvenile crime. We develop a simple theoretical model in which each delinquent decides how much crime effort to exert. The objective function of each individual related to the criminal leader has two key aspects. First, the proceeds from crime depend on a learning effect on how to commit crime that each individual obtains from the criminal leader in the network. Such a learning effect translates into an improvement of the technology of crime to individuals, but this technology decreases with the distance to the leader in the network. Second, because of a conformity effect, each delinquent wants to minimize the (social) distance between his or her criminal behavior and that of the leader, so that the closer he or she is to the leader in the network, the higher is the influence of the leader. In other words, each delinquent loses utility from failing to conform to the leader's criminal behavior, and this cost is higher the further away the delinquent is from the leader in the network.

We show that the criminal behavior of each individual depends on the distance to the leader in the social network. On top of the intensive margin of crime, we extend this model by introducing the extensive margin related to the crime decision of each individual. There are now two stages. In the first stage, each individual decides whether to become a criminal. If he or she chooses to do so, he or she decides how much crime to exert in the second stage. We show how, in this extended model, the distance to the criminal leader affects both the decision to become a criminal (extensive margin) and the number of crimes thereafter committed (intensive margin).

We structurally estimate this model by using data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) in the United States, which contains information on all students attending a random sample of U.S. high schools in 1995. This dataset provides unique information on friendship networks by asking students to nominate up to 10 friends from a school roster. It also contains detailed information on juvenile delinquency, including 12 types of crime. To identify criminal leaders in a way that is exogenous to the network formation process, we define a criminal leader as a teenager who has a level of crime more than three standard deviations above the median in the school. The distance to the leader is then calculated by using the (shortest) distance between any delinquent and a leader in the social network to which he or she belongs. Our baseline results are obtained when excluding individuals at distance 1 or 2 from the leader. This is because students choose their friends, and maybe the friends of their friends, but not beyond. As a result, the network distance from the leader for those further away may be reasonably assumed as exogenous.

Finally, we run some counterfactual exercises by simulating the effects of a policy aiming at removing all criminal leaders from a school. We show that this policy can, on average, reduce crime by 20% and decrease the individual probability of becoming a criminal by about 10%.

Our study contributes to the nascent economic literature on *leadership in social networks*. Management and psychology research increasingly recognizes the importance of social processes and relational linkages in shaping leadership. In addition to resources that stem from human capital, organizational capacities can be derived from social relationships, termed *social capital* (Putnam, 2000). In economics, while studies of the effects of social networks on a variety of outcomes are pervasive (Jackson, 2008; Jackson et al., 2017), the intersection between leadership and social networks has received limited attention. Of the recent studies that have investigated related issues, Tao and Lee (2014) identify the endogenous social interactions stemming from the best and the average of players' performance in college basketball. They find that peers' average performance measures are not significant, whereas the best performers' endogenous effect is significant and negative, signaling a highly competitive environment. By using data from German ninth-graders, Tatsi (2017) finds that *bad apples* are more important than classroom *stars* in affecting educational outcomes. Mastrobuoni and Patacchini (2012) show that network centrality, especially eigenvalue centrality, is an important predictor of leadership in the U.S. mafia organization. Finally, Islam et al. (2021) show that leaders in social networks (measured by the student who has the highest centrality in the network) have an important impact on the learning outcomes of primary-school students in Bangladesh.

To the best of our knowledge, this is the first study that measures the influence of a leader on other individuals by examining the *distance* between them in terms of the shortest paths in a social network. Indeed, most of the studies cited above have no common social space in which both leaders and the persons they influence are located. In addition, although researchers have investigated the role of network centrality on outcomes, showing that degree centrality (Christakis and Fowler, 2010), eigenvector and diffusion centrality (Banerjee et al., 2013; Islam et al., 2021), Katz–Bonacich centrality (Calvó-Armengol et al., 2009; Battaglini and Patacchini, 2018), betweenness centrality (Burt, 1992; Padgett and Ansell, 1993; Mehra et al. 2001), and key player centrality (Ballester et al., 2006; Lindquist and Zenou, 2014; Lee et al., 2021), we emphasize the importance of the *closeness centrality* of criminal leaders in explaining criminal behaviors.[1]

Our study is also related to the literature on *peer and network effects in crime*. A growing body of the empirical literature suggests that peer effects are strong in crime decisions. Ludwig et al. (2001) and Kling et al. (2005) study the relocation of families from high- to low-poverty neighborhoods by using data from the Moving to Opportunity experiment. They find that this policy reduces juvenile arrests for violent offences by 30–50% relative to a control group. This result also suggests strong social interactions in crime behaviors. Patacchini and Zenou (2008, 2012) find that peer effects in crime are strong, especially for petty crime. Bayer et al. (2009) consider the influence of juvenile offenders serving time in the same correctional facility on each other's subsequent criminal behavior. They also find strong evidence of learning effects in criminal activities since exposure to peers with a history of committing a particular crime increases the probability that an individual who has already committed the same type of crime recidivates that crime. Damm and Dustmann (2014) show that growing up in a neighborhood in which a relatively high share of young people have committed crime increases the individual's probability of committing crime later on.[2] In this study, we consider a different approach by examining the effect of a specific type of peer: the (criminal) leader.

The rest of the paper unfolds as follows. In Section 2, we expose our main theoretical framework. Section 3 describes the data and provides some preliminary evidence that puts forward the negative relationship between distance to leader and criminal behavior. Section 4 explains our empirical strategy and presents our empirical results. In Section 5, we extend our model to include the decision to become a criminal and empirically test the results of this model, while in Section 6 we employ an alternative definition of criminal leader and include non-leader peers as robustness checks. Section 7 is devoted to the policy experiments where we remove all leaders in the school and examine the impact on the decision to become a criminal and the number of crimes committed by each delinquent. Finally, Section 8 concludes. Appendix A provides two extensions of our theoretical model. Appendix B extends our model to incorporate the decision of becoming a criminal leader. Appendix C describes our data.

[1]For the definition of these centrality measures, see Wasserman and Faust (1994) and Jackson (2008).

[2]For theoretical foundations on peer and network effects in crime, see Glaeser et al. (1996), Ballester et al. (2006, 2010), Calvó-Armengol and Zenou (2004), Calvó-Armengol et al. (2007), Cortés et al. (2019), and Bezin et al. (2021). For an overview of the literature on crime and networks, see Lindquist and Zenou (2019).

2 Theoretical framework

We consider a simple model in which adolescents decide how much crime to commit.

The network $N = \{1, \ldots, n\}$ is a finite set of agents. The $n-$square adjacency matrix $\mathbf{G}$ of network g keeps track of the direct connections in this network. Here, two individuals, i and j, are connected (i.e., best friends) in g if and only if $g_{ij} = 1$, and $g_{ij} = 0$ otherwise. Given that friendship is a reciprocal relationship, we set $g_{ij} = g_{ji}$.[3] We also set $g_{ii} = 0$. The set of individual i's best friends (direct connections) is $N_i(g) = \{\text{all } j \neq i \mid g_{ij} = 1\}$, which is of size $\overline{g}_i$ (i.e., $\overline{g}_i = \sum_{j=1}^{n} g_{ij}$ is the number of direct links of individual i).

We have the following standard network-related definitions. A *walk* in network g refers to a sequence of nodes, $i_1, i_2, i_3, \ldots, i_{L-1}, i_L$ such that $g_{i_l i_{l+1}} = 1$ for each l from 1 to $L-1$. The *length* of the walk is the number $L-1$ of links in it. A *path* in network g is a walk in g, $i_1, i_2, i_3, \ldots, i_{K-1}, i_K$, such that all the nodes are distinct. The (geodesic) *distance* between nodes i and j in a network is the length of the *shortest path* between them.

Preferences Delinquents in network g belonging to school s decide how much crime effort to exert (i.e., how many crimes to commit). We denote by $y_{i,gs}$ the delinquency effort level of delinquent i in network g belonging to school s and by $\boldsymbol{y}_{gs} = (y_{1,gs}, ..., y_{n,gs})'$ the population delinquency profile in network g belonging to school s.[4] Each agent i selects an effort $y_{i,gs} \geq 0$ and obtains a payoff $u_{i,gs}(\boldsymbol{y}_{gs})$, which depends on the effort profile $\boldsymbol{y}_{gs}$ and underlying network g, in the following way:

$$u_{i,gs}(\boldsymbol{y}_{gs}) = \underbrace{(\alpha_{i,gs} + \xi_s + \epsilon_{i,gs})\, y_{i,gs}}_{\text{Proceeds}} - \underbrace{\frac{1}{4} y_{i,gs}^2}_{\text{Cost of crime effort}} - \underbrace{p\, f\, y_{i,gs}}_{\text{Cost of being caught}} - \underbrace{\frac{1}{4}\left(y_{i,gs} - \phi^{d_{iL,gs}} y_{L,gs}\right)^2}_{\text{Cost of distance to ``leader''}} \tag{1}$$

where $0 < \phi < 1$ and $d_{iL,gs}$ is the geodesic distance in network g between individual i and the leader, L (i.e., the length of the shortest path between i and L).

This utility has a standard cost/benefit structure (as in Becker, 1968). The cost of crime is captured by the probability of being caught $0 < p < 1$ times the fine f, which increases with an individual's effort $y_{i,gs}$, as the severity of the punishment increases with involvement in crime. In addition, individuals have a *direct* cost of committing crime equal to $\frac{1}{4} y_{i,gs}^2$, which is also increasing in own crime effort $y_{i,gs}$. The proceeds from crime are given by $(\alpha_{i,gs} + \xi_s + \epsilon_{i,gs})\, y_{i,gs}$ and are increasing in individuals' own effort $y_{i,gs}$, where $\alpha_{i,gs}$ denotes the (marginal) *productivity*

[3] All our theoretical results hold even when $g_{ij} \neq g_{ji}$.

[4] In our data, there can be more than one network in a school.

of committing crime, ξ_s denotes the unobservable characteristics of the environment such as the prosperity level of the neighborhood/school (i.e., more prosperous neighborhoods lead to higher proceeds from crime), and $\epsilon_{i,gs}$ is an error term that captures other uncertainty in the proceeds from crime. Both ξ_s and $\epsilon_{i,gs}$ are observed by the delinquents (when choosing their effort level) but not by the econometrician.

The individual productivity of crime is given by

$$\alpha_{i,gs} = \underbrace{\beta_{i,gs}}_{\text{crime technology from own characteristics}} + \underbrace{\max\left[\left(\omega_{gs} - \theta d_{iL,gs}\right), 0\right] \times 1_{L,gs}}_{\text{technology gain from the leader}} \tag{2}$$

where $\theta > 0$ and $1_{L,gs} = 1$ if there is a leader in the school and 0 otherwise. This means that the leader in the network provides some fixed crime technology to all criminals (equal to ω_{gs}) but this technology decreases with distance to the leader (the total technology gain from the leader is $\omega_{gs} - \theta d_{iL,gs}$): the closer someone is to the leader, the more he or she learns from this leader. The max function ensures that the total technology gain from the leader is always positive. We, indeed, assume that $\omega_{gs} > \theta d_{iL,gs}$ for all $d_{iL,gs}$. To facilitate the presentation of the model, we assume that there is always one leader in each network (as we see below, this is not always true in the data since the leader is defined at the school level and not the network level), meaning that $\alpha_{i,gs}$ can be written as

$$\alpha_{i,gs} = \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} > 0$$

Furthermore, in (2), $\beta_{i,gs}$ captures the *observable* characteristics of individual i. This can be written as

$$\beta_{i,gs} = \sum_{m=1}^{M} \beta_{1m} x_{i,gs}^{m} \tag{3}$$

where $x_{i,gs}^{m}$ denotes the set of M observable variables (e.g., sex, race, age, parental education). In other words, $\alpha_{i,gs}$, the crime productivity of individual i, is positively affected by his or her characteristics[5] and negatively affected by the distance to the leader, $d_{iL,gs}$. In other words, being closer to a leader increases the efficiency of the technology of committing crime. Indeed, a delinquent learns to commit crime more efficiently and has higher productivity the closer he or she is to the leader. Our formulation therefore implies the following. When individual i is at a distance of one node from the leader (direct friendship), then i's marginal productivity is $\alpha_{i,gs} = \beta_{i,gs} + \omega_{gs} - \theta$, which is greater than $\beta_{i,gs}$ (i's productivity when i is alone) since $\omega_{gs} > \theta$. Then, when the leader

[5] In fact, as we see in the empirical analysis, the sign of the effect of $\beta_{i,gs}$ on crime productivity depends on which characteristic we are considering. If we think of height or self-esteem, then being tall or having high self-esteem may increase an individual's crime productivity. On the contrary, if we consider parental education, students with more educated parents may have lower crime productivity.

is at a distance of two nodes from individual i, i's productivity is given by $\beta_{i,gs} + \omega_{gs} - 2\theta < \beta_{i,gs} + \omega_{gs} - \theta$, and so on.

Another interesting aspect of this utility function is the last term $\frac{1}{4}\left(y_{i,gs} - \phi^{d_{iL,gs}} y_{L,gs}\right)^2$. This is such that each individual i wants to minimize the distance between his or her crime effort and that of the leader. Indeed, the individual loses utility $\frac{1}{4}\left(y_{i,gs} - \phi^{d_{iL,gs}} y_{L,gs}\right)^2$ from failing to conform to the leader's behavior. This term captures the fact that the influence of a leader on the "reference behavior" of an individual is declining with the geodesic distance to the leader. When $\phi < 1$, the quadratic cost term of conformism is an increasing function of the geodesic distance $d_{iL,gs}$ as $\phi^{d_{iL,gs}}$ is decreasing in $d_{iL,gs}$ and the cost at distance 2 is therefore larger than the cost at distance 1. Hence, it is easily verified that:

$$\frac{\partial u_{i,gs}}{\partial y_{i,gs} \partial d_{iL,gs}} = -\theta + \frac{1}{2} y_{L,gs} \phi'(d_{iL,gs}) < 0 \text{ and } \frac{\partial u_{i,gs}}{\partial y_{i,gs} \partial y_{L,gs}} = \frac{1}{2} \phi^{d_{iL,gs}} > 0$$

where $\phi(d_{iL,gs}) \equiv \phi^{d_{iL,gs}}$ with $\phi'(d_{iL,gs}) < 0$. This means that the higher is the distance between individual i and the leader, the lower is the marginal utility of exerting crime effort for i.[6] Moreover, the leader always exerts a positive influence on individual i in terms of crime since the higher is the leader's crime effort, the higher is the utility to the delinquent of providing his or her own crime effort.[7]

Note that we focus here on the impact of leaders on individual crime, without modeling the role of peers who are not leaders. As stated in the Introduction, network modeling studies exclusively focus on the impact of peers on outcomes (such as crime) by using a conformist model (see Akerlof, 1997; Patacchini and Zenou, 2012; Boucher, 2016; Ushchev and Zenou, 2020). Our aim here is to focus on the role of the leader only, abstracting from other peer effects.

To summarize, in our model, leaders plays two roles: they transmit the criminal technology, which increases the individual productivity of crime, and they act as a role model for criminals. The utility function of individual $i \neq L$ can thus be written as

$$u_{i,gs}(\boldsymbol{y}_{gs}) = \left[\beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \eta_s + \epsilon_{i,gs}\right] y_{i,gs} - \frac{1}{4} y_{i,gs}^2 - \frac{1}{4}\left(y_{i,gs} - \phi^{d_{iL,gs}} y_{L,gs}\right)^2 \tag{4}$$

where $\eta_s = \xi_s - pf$. Here, utility (4) is concave in an individual's own crime decisions and displays

[6] This theoretical framework allows for the existence of several criminal leaders in the network by assuming that leaders do not affect each other and that only the distance to the closest leader (i.e., the geodesic distance $d_{iL,gs}$) matters for each delinquent i in network g belonging to school s when deciding how much crime effort to exert. In such a case, the distance to any other leader has no impact on the criminal activity on each individual.

[7] In Appendix A.1, we consider an alternative to the utility function (1), where the only difference is the conformist component, which is given by $\frac{\phi^{d_{iL,gs}}}{4}\left(y_{i,gs} - y_{L,gs}\right)^2$ instead of $\frac{1}{4}\left(y_{i,gs} - \phi^{d_{iL,gs}} y_{L,gs}\right)^2$. We show that the theoretical predictions of this model are similar to those presented in this section.

decreasing marginal returns in his or her own effort levels.

For the leader, the utility function can be written as

$$u_{L,gs}(y_{L,g}) = \beta_{L,gs} y_{L,gs} - \frac{1}{2} y_{L,gs}^2 \tag{5}$$

where

$$\beta_{L,gs} \equiv \sum_{m=1}^{M} \beta_{1m} x_{L,gs}^m \tag{6}$$

as the leader does not compare his or her criminal activity with that of anybody else and his or her crime productivity is only influenced by his or her own characteristics.[8]

Nash equilibrium Let us first determine the criminal behavior of the leader. Maximizing (5) leads to

$$y_{L,gs}^* = \sum_{m=1}^{M} \beta_{1m} x_{L,gs}^m \tag{7}$$

We can now characterize the Nash equilibrium of the game for all individuals but the leader, where agents choose their effort level $y_{i,gs} \geq 0$ simultaneously. In equilibrium, each agent maximizes his or her utility (4). For each $i \neq L$, we easily obtain

$$y_{i,gs}^* = \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \eta_s + \epsilon_{i,gs} + \frac{\phi^{d_{iL,gs}}}{2} y_{L,gs}^* \tag{8}$$

The equilibrium crime effort of each individual i is a combination of his or her crime productivity and the crime effort of the leader. The higher is the crime effort of the leader $y_{L,gs}^*$, the higher is the crime effort of the delinquent $y_{i,gs}$. Naturally, the effect of the leader on the delinquent's crime is reduced the further away individual i is from the leader in the network. Indeed,

$$\frac{\partial y_{i,gs}}{\partial d_{iL,gs}} = -\theta + \left(\frac{\log \phi}{2} \right) \phi^{d_{iL,gs}} < 0$$

since $0 < \phi < 1$. We can easily calculate the equilibrium utility of each individual $i \neq L$ by plugging the value of $y_{i,gs}^*$ from (8) into (4). We obtain

$$u_{i,gs}^*(\boldsymbol{y}_{gs}) = \frac{1}{2} \left(\pi_{i,gs} - \theta d_{iL,gs} \right)^2 - \frac{1}{8} \left(\phi^{d_{iL,gs}} y_{L,gs}^* \right)^2 + \frac{1}{2} \left(\pi_{i,gs} - \theta d_{iL,gs} \right) \phi^{d_{iL,gs}} y_{L,gs}^* \tag{9}$$

[8]In Appendix A.2, we relax this assumption and assume instead that the crime decision of the leader is a function of his or her characteristics as well as the total criminal activity in the network to which the leader belongs. We show that the empirical predictions are similar to those obtained in this model.

where $\pi_{i,gs} \equiv \beta_{i,gs} + \omega_{gs} + \eta_{gs} + \epsilon_{i,gs}$. For the leader, we have

$$u_{Ls}^*(y_{L,gs}^*) = \frac{1}{2}\beta_{L,gs}^2 \tag{10}$$

We can write equation (8) as follows:

$$y_{i,gs} = \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \phi^{d_{iL,gs}*}\beta_{L,gs} + \eta_s + \epsilon_{i,gs} \tag{11}$$

where $\beta_{i,gs}$ and $\beta_{L,gs} \equiv \sum_{m=1}^{M} \beta_{1m} x_{L,gs}^m$ are given by (3) and (6), $\phi^{d_{iLs}*} \equiv \phi^{d_{iLs}}/2$. In the empirical analysis, we test equation (11) to examine how the delinquent behavior of individual i is affected by his or her observable characteristics, that of the leader and his or her distance to the leader in the social network.

In Appendix B, we extend the model to allow for endogenous leadership, that is, we explain how to become a criminal leader in a network. Define the criminal leader as the criminal $i = L$ with the highest value of $\pi_{i,gs}^0 = \beta_{i,gs} + \eta_{gs} + \epsilon_{i,gs}$. We show that, if $\pi_{L,gs}^0 \geq \widetilde{\pi}_{L,gs}^0$, where $\widetilde{\pi}_{L,gs}^0$ is defined in (B.4), then there exists one criminal leader in the network and, thus, our model and all our results remain the same.

3 Data description

Our analysis is based on information from the National Longitudinal Study of Adolescent to Adult Health (Add Health).[9] The Add Health survey 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 the years 1994-1995 (Wave I). Every pupil attending the sampled schools on the interview day was asked to compile a questionnaire (in-school data) containing questions on respondents' demographic and behavioral characteristics, education, family background and friendship. A subset of adolescents selected from the rosters of the sampled schools, about 20,000 individuals, was then asked to complete a long in-home questionnaire that contains an extensive set of questions on juvenile delinquency encompassing 12 items.[10] The survey asks

[9] Add Health is a program directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain Add Health data files is available from the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from Grant P01-HD31921 for this analysis.

[10] These 12 delinquency items are: painting graffiti or signs on someone else's property or in a public place; deliberately damaging property that didn't belong to you; getting into a serious physical fight; hurting someone

students how often they participated in each criminal activity during the past year.[11] Each response is coded by using a four-point ordinal scale: 0 (never participate), 1 (participate once or twice), 2 (participate three or four times), and 3 (participate five or more times).[12] Table 1 describes the criminal activities as well as the percentage of the students involved in each crime type. Petty theft and physical fights are the most recurrent types of juvenile delinquency. We measure individual criminal activity by using the total number of committed crimes (i.e., the sum of the responses to each question listed in Table 1).[13] Non-criminal individuals are defined as those who report to have never participated in any delinquent activity.

[*Insert Table* 1 *here*]

From a network perspective, the most interesting aspect of the Add Health data is the friendship information, which is based upon actual friends' nominations. Indeed, pupils were asked to identify their best friends from a school roster (up to five males and five females). We set $g_{ij} = 1$ when individual i has nominated individual j as his or her best friend or vice versa. As a result, one can reconstruct the whole geometric structure of the friendship networks.[14]

To obtain a definition of a crime leader that is *exogenous* to the network formation, we define leaders as students who are outliers in terms of crimes committed by all the students *in the school*. More specifically, we identify the leaders as those individuals whose criminal activity is more than

badly enough to need bandages or care from a doctor or nurse; stealing something worth more than $50; going into a house or building to steal something; using or threatening to use a weapon to get something from someone; selling marijuana or other drugs; stealing something worth less than $50; taking part in a fight where a group of your friends was against another group; pulling a knife or gun on someone; and shooting or stabbing someone. The Add Health dataset contains information on other activities not considered here since they are undesirable behaviors rather than criminal acts (e.g., lying to your parents or guardians about where you have been or whom you were with), or items that are already included in those listed above (e.g., taking something from a store without paying for it).

[11] Respondents listened to pre-recorded questions through earphones and then 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.

[12] The questions related to pulling a knife or gun on someone and shooting or stabbing someone are coded slightly differently: 0 (never), 1 (once), and 2 (more than once).

[13] This is a standard approach in sociology (see Guo et al., 2008). We also considered an alternative measure of crime activity using factor analysis. The qualitative results remained unchanged, although they are more difficult to interpret.

[14] The limit in the number of nominations is not binding (even by sex). Fewer than 1% of the students in our sample show a list of 10 best friends.

three median absolute deviations from the school's median.[15,16,17] This definition is supported by evidence since leaders tend to commit much more crime than other criminals, especially for juvenile crime (Reiss, 1988; Reiss and Farrington, 1991; Amemiya et al., 2016). For each individual, we then use the friendship links to calculate the distance to a leader.[18] A given individual may not even know (or ever meet) the leader, who can be someone in a different grade/class or of a different race/sex. In our sample, friendship networks are dense: roughly 40% of the students are directly or indirectly connected to the leaders through a friendship chain (e.g., friends of friends of friends). Distances range between zero nodes (for the leader) and seven nodes.[19] We label those students as "inside a leader network" and focus our regression analysis on those individuals. We label the remaining sample of students as "outside a leader network". We use this sample of students who are outside a leader's reach as a counterfactual in our simulation experiment presented in Section 7 to assess the effect of a policy that removes all leaders.

Figure 1 displays the distribution of the students in our sample according to their distance to a leader. The distribution is positively skewed, which is expected given that adolescent friendship networks are dense.

[*Insert Figure* 1 *here*]

By matching the identification numbers of the friendship nominations to the respondents' identification numbers, one can obtain information on the characteristics of nominated friends. This unique feature of Add Health enables our empirical exercise. Indeed, for each individual, we know a vast array of characteristics, including crime activity and parental education.[20] Figure 2 reports the correlation between the distance to a leader and crime activities in the raw data. The figure shows the average crime level for individuals at different path lengths from the leader, highlighting

[15] An alternative definition could be based on standard deviations rather than absolute deviations. However, since the distance from the median is squared in the standard deviation case, large deviations have more weight. As a result, outliers may be too influential in our leader definition. This is why we prefer to use the median absolute deviations from the data's median.

[16] We do not define the leader as the delinquent with the highest crime rate in the school because, in that case, we would have just one leader in the entire school. Our purpose is to single out the most notable criminals, and the fact that there may be more than one leader in a network makes the definition of the "distance to the leader" more meaningful.

[17] In Section 6.1 below, we provide a robustness check using a different definition of a criminal leader. We use a more general measurement by considering the top 10 percentile of individual crime index distribution in each school.

[18] While we consider all networks (including those having more than one individual with an extreme level of crime), each student is uniquely matched to one leader, the one who is his or her closest (in terms of geodesic distance) leader in the network. We do so to exactly match our theoretical model of Section 2 where only the distance to the closest leader matters.

[19] Fewer than 2% of students are more than seven nodes away from a leader. They are therefore considered to be unconnected to the leader.

[20] The other existing surveys that report friends' nominations are ego-networks, namely the respondent lists his or her contacts and their basic characteristics such as sex, education, and employment status. Hence, detailed information about nominated contacts (e.g., crime activity) is not typically available.

the unambiguous *negative correlation* between crime and the distance to a leader.

[*Insert Figure* 2 *here*]

Our final sample consists of 6,557 observations distributed over 353 networks and 135 schools. The large reduction in the sample size compared with the original sample is mainly due to the network construction procedure and to missing data for the variables.[21]

Tables C.1−C.5 in Appendix C describe the variables used in our study. Table C.1 reports the summary statistics for the entire sample and for the sample of criminals. Table C.2 compares the samples of criminals and non-criminals and Table C.3 examines leaders and non-leaders. Table C.4 compares the characteristics of delinquents at different distances to the criminal leaders. Table C.5 examines the characteristics of adolescents inside and outside a leader network.

Table C.1 shows that among the adolescents selected in our sample of students, 46% are female and 40% are non-white. Average parental education is high-school graduate. More than half of our students declare to have committed at least one of the delinquent activities reported in the questionnaire (Table C.2). This is not surprising given that most of the activities are petty crimes. Criminals, on average, are less likely to be females and have parents going less often to religious services.[22] The remaining characteristics do not exhibit particularly marked differences between criminals and non-criminals. Table C.3 shows that the observable characteristics of leaders and non-leaders are relatively similar, apart from criminal activities. Table C.4 shows that the observable characteristics of delinquents at different distances to criminal leaders are similar. Finally, Table C.5 also suggests that the samples of students inside and outside a leader network are comparable.

4 Empirical analysis

4.1 Preliminary evidence

Let $\overline{g}$ be the total number of networks in the sample, n_{gs} be the number of individuals in network g in school s, and $n_s = \sum_{g=1}^{\overline{g}} n_{gs}$ be the total number of sample observations. We begin our empirical investigation with a non structural test of the main prediction of our model. We study the existence of a negative correlation between individual crime activity and distance to the leader by using the following linear regression model:

[21] This is common when working with AddHealth data. A comparison of the summary statistics between the original and selected samples shows that representativeness is preserved. This table is available upon request.

[22] The difference in religion practice intensity between criminals and non-criminals is 0.156, statistically significant (at the 1% level) in favor of the non-criminal group.

$$y_{i,gs} = \omega - \sum_{\tau=1}^{7} \theta_\tau D_{\tau,i,gs} + \sum_{m=1}^{M} \beta_{1m} x_{i,gs}^m + \sum_{m=1}^{M} \beta_{2m} x_{L,gs}^m + \eta_{gs} + \epsilon_{i,gs} \tag{12}$$

where $D_{1,i,gs} = 1$ if individual i in network g in school s is at a distance of one node from the leader and 0 otherwise, $D_{2,i,gs} = 1$ if individual i in network g in school s is at a distance of two nodes from the leader and 0 otherwise, and so on. In other words, to capture the distance to the leader, we create seven dummy variables $D_{1,i,gs}, \cdots, D_{7,i,gs}$, where $D_{0,i,gs}$ corresponds to the criminal leaders, our reference group, whose distance to the criminal leader is obviously 0. Denote by $\widehat{\theta}_\tau$ and $\widehat{\omega}$ the estimated values of θ_τ and ω, respectively. Since the technology gain from the leader should always be positive, we need to impose: $\widehat{\omega} - \widehat{\theta}_\tau \geq 0$, $\forall \tau = 1, \cdots, 7$. To be sure that the estimation will satisfy that condition, we impose that $\widehat{\omega} = \widehat{\theta}_7$, since the maximum distance to the leader in the data is seven (see Section 3).[23]

The first two columns of Table 2a show the OLS estimation results without and with *school fixed effects*, respectively. In line with the predictions of the theoretical model, it appears that the effect of the distance to the leader (i.e., θ_τ) is positive and statistically significant, which means that magnitude of the coefficients increases in absolute value with the distance to the leader, indicating that the further away is a leader, the lower is his or her influence on the individual in terms of criminal activities.

[*Insert Table* 2a *here*]

Figure 3 plots the marginal effects. This figure confirms the results of Table 2a by showing that the estimated coefficients of the distance dummies are increasing in absolute value with distance: the larger is the distance to the leader, the lower is the influence of the leader on other delinquents. We also test if coefficients of distance-to-leader dummies are actually statistically different from each other. Table 2b reports statistically significant differences between the impact of the different θ coefficients.

[*Insert Table* 2b *here*]

The effects of the other control variables are in line with our expectations. Being female, having a better high-school performance, living in a well-kept neighborhood, and having high-educated parents are negatively correlated with criminal activity, whereas non-white students seem to be more likely to commit crime.

[*Insert Figure* 3 *here*]

[23] Observe that equation (12) is slightly different from equation (11), as we do not include explicitly $d_{iL,gs}$, the distance to the leader, in this equation but instead add the subscript τ for each θ to indicate the distance to the leader. As a result, the condition that the technology gain from the leader is positive is written as $\omega \geq \theta d_{iL,gs}$, $\forall d_{iL,gs}$, in (11), while, for (12), it is given by $\omega \geq \theta_\tau$, $\forall \tau$.

The remaining columns of Table 2a display the results of two robustness checks. Our identification strategy raises one main concern, which is the fact that the decreasing criminal activity of individuals at increasing distances from the leader may simply occur because of a particularly high crime level of the individuals directly connected to the leader. If unobserved characteristics are driving both friends' nominations and criminal activity, then our results capture a sorting effect given that individuals at a distance of one node from the leader can indeed choose him or her as a friend. To address this concern, we repeat our analysis after dropping individuals located one link away from the leader (columns (3) and (4)) and two links away from the leader (columns (5) and (6)). If students choose their direct friends and may choose their friends of friends, it is, indeed, unlikely that they choose individuals at a distance of three or more nodes from them. Table 2a shows that the evidence from these robustness checks remains roughly unchanged.

4.2 Structural estimation

Let us now add the data to the specific structure of the theoretical model. According to our theory, preferences for conformity, as captured by the parameter ϕ and the productivity gains $\omega-\theta$, explain juvenile delinquency according to expression (11), which is non-linear in parameters. As a result, we estimate these parameters by non-linear least squares (NLS), where the Gauss-Newton algorithm is used to solve the associated minimization problem. In other words, we need to solve the following unconstrained minimization problem:[24]

$$\min_{\boldsymbol{\nu}} \sum_i \left(y_{i,gs} - f_{i,gs}(\mathbf{z},\boldsymbol{\nu})\right)^2$$

where $y_{i,gs}$ is the crime effort of individual i belonging to network g and studying in school s, $\mathbf{z}$ is the vector of all observable characteristics in our empirical model, whereas $\boldsymbol{\nu} = \left(\boldsymbol{\beta}, \omega, \theta, \phi^D\right)$.[25] In our case,

$$f_{i,gs}(\mathbf{z},\boldsymbol{\nu}) = \begin{cases} \sum_{m=1}^{M} \beta_{1m} x_{i,gs}^m + \omega - \theta d_{iL,gs} + \frac{1}{2}\phi_D^{d_{iL,gs}^D} \sum_{m=1}^{M} \beta_{2m} x_{L,gs}^m + \eta_s & \text{if} \quad i \neq L \\ \sum_{m=1}^{M} \beta_{0m} x_{L,gs}^m + \omega + \eta_s + \epsilon_{L,gs} & \text{if} \quad i = L \end{cases}$$

A simple way to solve this optimization problem, especially when $f_{i,gs}(\mathbf{z},\boldsymbol{\nu})$ is *nonlinear* in the parameters, is by using the Gauss-Newton method. This algorithm iteratively finds the value of $\boldsymbol{\nu}$

[24]In all our estimations, we assume that the network is exogenous, which means that the distance to the leader is fixed and exogenous. Tackling the issue of network endogeneity is a very complicated issue and is beyond the scope of this paper.

[25]For each individual (i, gs), $y_{i,gs}$ is measured using the $CAI_{i,gs,t}$, i.e., the Criminal Activity Index of individual i belonging to network g in school s in Wave I (see equation (11)).

that minimizes $\sum_i r_{i,gs}^2(\boldsymbol{\nu})$, where $r_{i,gs}(\boldsymbol{\nu})$ denotes the residuals, i.e., $r_{i,gs} \equiv y_{i,gs} - f_{i,gs}(\mathbf{z}, \boldsymbol{\nu})$. We start with an initial guess for $\boldsymbol{\nu}^{(0)}$ and, at each iteration, our model is linearized by approximation to a first-order Taylor series expansion about $\boldsymbol{\nu}^{(k)}$. As a result, the problem is now a *linear* least-square problem given by:

$$\min_{\boldsymbol{\nu}} \sum_i \left[r_{i,gs}\left(\boldsymbol{\nu}^{(k)}\right) + \nabla r_{i,gs}\left(\boldsymbol{\nu}^{(k)}\right)^T \left(\boldsymbol{\nu} - \boldsymbol{\nu}^{(k)}\right) \right]^2,$$

where $\nabla r_{i,gs}\left(\nu^{(k)}\right)$ is the vector of the first partial derivatives of $r_{i,gs}(\boldsymbol{\nu})$ evaluated at $\boldsymbol{\nu}^{(k)}$.

Panel I (NLS) in Table 3 reports the estimation results.[26] While column (1) presents the results for the entire sample, columns (2) and (3) report the results when dropping individuals located one link away from the leader (column (2)) and two links away from the leader (column (3)). The coefficient estimates of the key parameters are both statistically significant and consistent with our theoretical framework: $0 < \widehat{\theta} < \widehat{\omega}$ and $0 < \widehat{\phi} < 1$, e.g., $\widehat{\omega} = 7.019$, $\widehat{\theta} = 1.003$ and $\widehat{\phi} = 0.868$, in column (3), when we drop individuals at distances of one and two nodes from the leader.[27] Regarding the marginal effects, being located one link away from the leader reduces criminal activity by 0.2 units in the reported criminal activity of a delinquent. Since the average number of crimes committed by non-leaders is 1.3, the marginal effects suggest a reduction of 15% in criminal activity for the average agent in the network compared with someone located one link closer to a leader.

[*Insert Table* 3 *here*]

Because leaders may have different impact on delinquents depending on the types of crimes they commit, we differentiate between violent and non-violent crimes (see Table 1 for a definition of each type of crime).[28] The results for the non-linear least square estimations are reported in panel I of Table 4a (violent crimes) and Table 4b (non-violent crimes). We see that effects are stronger for violent crimes than non-violent crimes.

[*Insert Tables* 4a *and* 4b *here*]

[26] As stated in footnotes 6 and 18, while we allow for several leaders in a network, each student is uniquely matched with one leader, who is his or her closest leader in the network. For robustness, our analysis has also been performed when removing multi-leader networks. The results remain qualitatively unchanged.

[27] Observe that, following our theoretical model in (11), in order for the technology gain from the leader to be positive, we need that $\omega \geq \theta d_{iL,gs}$, $\forall d_{iL,gs} = 1, \cdots, 7$.

[28] Instead of violent versus non-violent crimes, one could have divided crimes differently. Ideally, the analysis should be run for each crime type separately since the crime technology is certainly crime-specific. The small size of our sample prevents us from doing so.

5 The decision to become a criminal

Let us now extend our analysis to incorporate the decision to become a criminal.

5.1 Theory

Consider now a two-stage model in which the students decide whether they want to be a delinquent in the first stage and then, if they decide to be so, decide how many crimes to commit in the second stage. As usual, let us solve the model by backward induction.

Second stage: ***Crime effort.*** We have already solved this stage. If person i is a criminal, then his or her equilibrium effort is given by equation (8), or

$$y_{i,gs} = \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \eta_s + \epsilon_{i,gs} + \frac{\phi^{d_{iL,gs}}}{2} y^*_{L,gs} \tag{13}$$

and the equilibrium utility is given by (see (9))

$$u^*_{i,gs}(\boldsymbol{y}_{gs}) = \frac{1}{2}\left(\pi_{i,gs} - \theta d_{iL,gs}\right)^2 - \frac{1}{8}\left(\phi^{d_{iL,gs}} y^*_{L,gs}\right)^2 + \frac{1}{2}\left(\pi_{i,gs} - \theta d_{iL,gs}\right)\phi^{d_{iL,gs}} y^*_{L,gs} \tag{14}$$

where $\pi_{i,gs} \equiv \beta_{i,gs} + \omega_{gs} + \eta_s + \epsilon_{i,gs}$.

First stage: ***Crime decision.*** Each individual must decide whether he or she wants to become a criminal. Assume that the outside option of being a non-criminal for individual i is given by $\overline{u}_{i,gs}$. Then, each individual i will trade off the utility of being a criminal (given by (14)) and that of being a non-criminal (equal to $\overline{u}_{i,gs}$). As a result, he or she decides to become a criminal if and only if $u^*_{i,gs}(\boldsymbol{y}_{gs}) \geq \overline{u}_{i,gs}$ and a non-criminal otherwise. The value of $\pi_{i,gs}$, denoted by $\widetilde{\pi}_{i,gs}$, for which individual i is indifferent between being a criminal and not being so is given by $u^*_{i,gs}(\boldsymbol{y}_{gs}) = \overline{u}_{i,gs}$, or equivalently

$$4\widetilde{\pi}^2_{i,gs} + 4\left(\phi^{d_{iL,gs}} y^*_{L,gs} - 2\theta d_{iL,gs}\right)\widetilde{\pi}_{i,gs} - \left[\left(\phi^{d_{iL,gs}} y^*_{L,gs}\right)^2 + 4\theta d_{iL,gs}\left(\phi^{d_{iL,gs}} y^*_{L,gs} - \theta d_{iL,gs}\right) + 8\overline{u}_{i,gs}\right] = 0$$

Assume that $\phi^{d_{iL,gs}} y^*_{L,gs} > 2\theta d_{iL,gs}$, $\forall d_{iL,gs}$. Then, it is easily verified that there is a unique strictly positive solution in $\widetilde{\pi}_{i,gs}$ to this equation, which we denote by $\widetilde{\pi}^1_{i,gs} \equiv \widetilde{\pi}_{i,gs}\left(\pi_{i,gs}, \theta, \phi, d_{iL,gs}\right)$. This means that all individuals for which $\pi_{i,gs} \geq \widetilde{\pi}^1_{i,gs}$ will commit crime, while those for which $\pi_{i,gs} < \widetilde{\pi}^1_{i,gs}$ will not commit crime. Clearly, $\widetilde{\pi}^1_{i,gs}$ is increasing with $\overline{u}_{i,gs}$ since a better outside option leads to less crime.

As a result, the equilibrium effort and utility of individuals are given by

$$y^*_{i,gs} = \begin{cases} \beta_{i,gs} + \omega 1_{L,gs} - \theta d_{iL,gs} + \eta_s + \epsilon_{i,gs} + \frac{\phi^{d_{iL,gs}}}{2} y^*_{L,gs} & \text{if} \quad \pi_{i,gs} \geq \widetilde{\pi}^1_{i,gs} \\ 0 & \text{if} \quad \pi_{i,gs} < \widetilde{\pi}^1_{i,gs} \end{cases} \tag{15}$$

and

$$u^*_{i,gs}(\boldsymbol{y}_{gs}) = \begin{cases} \frac{1}{2} \left(\pi_{i,gs} - \theta d_{iL,gs}\right)^2 - \frac{1}{8} \left(\phi^{d_{iL,gs}} y^*_{L,gs}\right)^2 + \frac{1}{2} \left(\pi_{i,gs} - \theta d_{iL,gs}\right) \phi^{d_{iL,gs}} y^*_{L,gs} & \text{if} \quad \pi_{i,gs} \geq \widetilde{\pi}^1_{i,gs} \\ \overline{u}_{i,gs} & \text{if} \quad \pi_{i,gs} < \widetilde{\pi}^1_{i,gs} \end{cases}$$

where $\pi_{i,gs} \equiv \beta_{i,gs} + \omega_{gs} + \eta_s + \epsilon_{i,gs}$.

Denote by $c_{i,gs}$ the decision to commit crime. Hence, $c_{i,gs} = 1$ means that individual i, gs becomes a criminal, while $c_{i,gs} = 0$ means that individual i, gs does not. Then, equation (15) implies that the probability of becoming a criminal can be written as $P(c_{i,gs} = 1 | \pi_{i,gs}, \theta, \phi, d_{iL,gs}, \overline{u}_{i,gs})$.

5.2 Empirical analysis

We now test model (15) in which $\overline{u}_{i,gs}$ acts as an exclusion restriction since it affects the decision to become a criminal but not the crime effort. In the data, $\overline{u}_{i,gs}$ can be approximated by the individual moral cost, which we measure by the intensity of the religion of the parents. Indeed, the higher is the moral cost of committing crime, the less likely an individual is to become a criminal. This cost is clearly individual-specific and we could understand it as the cost associated to disappoint parents. We assume this is a cost that responds to the *extensive margin* only: once the parents already know that their kid is a criminal, they are not disappointed anymore and are less likely to have an impact on how many crimes he or she will commit.

In the Add Health survey, a parent (preferably the resident mother) of each adolescent respondent interviewed in Wave I is asked to complete an interviewer-administered questionnaire. This questionnaire includes the following question: "How often have you gone to religious services in the past year?" The response is coded by using a five-point ordinal scale: 0 (no religion), 1 (never), 2 (less than once a month), 3 (less than once a week, but at least once a month), and 4 (once a week or more). The higher is this value, the higher is the moral cost. We believe that the higher is the religiosity of the parents, the higher is the moral cost of committing crime of their offspring. Indeed, as shown in Table C.2 in Appendix C, the moral cost (as measured by the parents' frequency of going to religious services) is higher for non-criminals (and the difference is statistically significant at the 1% level). In the empirical analysis, according to our model, the first-stage equation is

$$P(c_{i,gs} = 1 | d_{iL,gs}, x_{L,gs}, x_{i,gs}, \overline{u}_{i,gs}, \eta_s, \epsilon_{i,gs}) = \Lambda\left(\gamma_0 d_{iL,gs} + \sum_{m=1}^{M} \gamma_{1m} d_{iL,gs} x_{L,gs}^m + \sum_{m=1}^{M} \gamma_{2m} x_{i,gs}^m + \gamma_3 \overline{u}_{i,gs} + \eta_s + \epsilon_{i,gs}\right) \tag{16}$$

where $P(c_{i,gs} = 1 | d_{iL,gs}, x_{L,gs}, x_{i,gs}, \overline{u}_{i,gs}, \eta_{gs}, \epsilon_{i,gs})$ is the probability of becoming a criminal ($c_{i,gs} = 1$), $\Lambda(\cdot)$ is the logistic distribution, and $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ are the parameters governing the crime decision. As in the theoretical model, equation (16) explains the crime decision of individual i in network g in school s by $d_{iL,gs}$, the distance to the leader, the observable characteristics of the leader (i.e., $d_{iL,gs}$), his or her own observable characteristics, the outside option $\overline{u}_{i,gs}$ captured by the moral cost (parental religion practice), and his or her unobservable characteristics (i.e., $\epsilon_{i,gs}$).

Table 3 (panels II and III) reports the results that take the *selection into criminal activities* into account. We estimate model (15) by using Heckman's two-step approach. In the first stage, as in (16), the probability of being a criminal is modeled as a function of the observed characteristics. In the second stage, the predicted probability is used as a correction term (Mills ratio). As an exclusion restriction, we use the moral cost ($\overline{u}_{i,gs}$) as measured by the intensity of the religious activities of the parent.

According to the non-linear least square estimations, for each one-node increase in the distance to the leader, the total number of crimes committed by each student decreases, on average, by 0.3 units (panel I, column (1)). When we only look at criminals (Heckman's second stage, panel II, column (1)), the average effect increases to 0.4. When we drop individuals who are at a distance one or at a distance two from a network leader (column (3)), the average marginal effect increases from 0.2 to 0.5.

When all the controls are included, comparing the least squares estimation results (Table 3, panel I, column (1)) with Heckman's estimation results (Table 3, panel III, column (1)), it appears that the individual moral cost only affects the first stage of Heckman's model but not the second stage, and thus seems to be a valid exclusion restriction. This means that family values (as measured by parental religious practice) are important in shaping adolescents' decisions to commit crime rather than affecting their crime level. This finding suggests that once a child becomes a criminal (and thus disappoints his or her parents), other factors than family values affect his or her decision on how much crime to commit. On the contrary, the distance to the leader is relevant in shaping both the level of criminal activity and the decision to become a criminal. Indeed, the marginal effect is statistically significant both when we drop individuals who are at a distance one or two from the leader and when we do not drop them.

The last two panels in Tables 4a and 4b show the results of the Heckman's two-step estimation for violent (Table 4a) and non-violent (Table 4b) crime activities. We see that the results at the extensive and intensive margins are qualitatively similar to those in the corresponding panels in Table 3.

6 Robustness analysis

6.1 Criminal leaders

In this section, we conduct a robustness check by having a different definition of criminal leaders. Instead of defining them as individuals whose criminal activity is more than three median absolute deviations from the school's median, let us now use a more general measurement for the leader definition by considering the top 10 percentile of individual crime index distribution in each school. Table 5 shows the results of the estimations both for the model without (panel I) and with the Heckman's two-step estimation (panels II and III). We can see that there are very similar to those displayed in Table 3.

[*Insert Table* 5 *here*]

6.2 Non-leader peers

So far, the model does not consider the influence of peers different from the network leader. In this section we extend the model to the presence of heterogeneous peer effects and check if our results are robust to the inclusion of non-leader peer effects.

6.2.1 Theory

Let us modify the utility function of individual $i \neq L$, which was given by (4), to

$$u_{i,gs}(\boldsymbol{y}_{gs}) = [\beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \eta_s + \epsilon_{i,gs}]\, y_{i,gs} - \frac{1}{4} y_{i,gs}^2 - \frac{1}{4}\left(y_{i,gs} - \phi^{d_{iL,gs}} y_{L,gs}\right)^2 + \sum_j g_{ij}\rho_j y_{i,gs} y_{j,gs} \tag{17}$$

where $\eta_s = \xi_s - pf$ and the last term of the utility function captures the influence of peers on individual i. By having different ρ_j, we allow for the possibility that each of i's peer (direct connection) has a different influence on individual i. The first-order condition for individual $i \neq L$

is equal to:

$$\frac{\partial u_{i,gs}(\boldsymbol{y}_{gs})}{\partial y_{i,gs}} = \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \eta_s + \epsilon_{i,gs} - y_{i,gs} + \frac{1}{2}\phi^{d_{iL,gs}} y_{L,gs} + \sum_j g_{ij}\rho_j y_{j,gs} = 0$$

This leads to

$$y_{i,gs} = \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \frac{\phi^{d_{iL,gs}}}{2} \sum_{m=1}^{M} \beta_{1m} x_{L,gs}^{m} + \sum_j g_{ij}\rho_j y_{j,gs} + \eta_s + \epsilon_{i,gs} \tag{18}$$

where by (7), $y_{L,gs}^{*} = \sum_{m=1}^{M} \beta_{1m} x_{L,gs}^{m}$. Denote by $\Lambda_{i,gs} := \alpha_{i,gs} + \eta_s + \epsilon_{i,gs} + \frac{\phi^{d_{iL,gs}}}{2} \sum_{m=1}^{M} \beta_{1m} x_{L,gs}^{m}$, where as above, $\alpha_{i,gs} := \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs}$. Denote by $\boldsymbol{\Lambda}_{gs} = (\Lambda_{1,gs}, \cdots, \Lambda_{n,gs})'$ the corresponding vector. Define $\boldsymbol{y}_{gs} = (y_{1,gs}, ..., y_{n,gs})'$. We can then write this first-order conditions for all agents in vector-matrix form as follows:

$$\boldsymbol{y}_{gs} = \boldsymbol{\Lambda}_{gs} + (\mathbf{R} \circ \mathbf{G})\, \boldsymbol{y}_{gs}, \tag{19}$$

where $\circ$ is the Hadamard product (entrywise product) and $\mathbf{R}$ is the $(n \times n)$-matrix defined by

$$\mathbf{R} = \begin{pmatrix} \rho_1 & \rho_2 & \cdots & \cdots & \rho_n \\ \rho_1 & \rho_2 & \cdots & \cdots & \rho_n \\ \rho_1 & \rho_2 & \cdots & \cdots & \rho_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_1 & \rho_2 & \cdots & \cdots & \rho_n \end{pmatrix}$$

Denote by $\mathbf{I}$ the identity matrix. Then, (19) can be written as:

$$\boldsymbol{y}_{gs} = (\mathbf{I} - \mathbf{R} \circ \mathbf{G})^{-1} \boldsymbol{\Lambda}_{gs}. \tag{20}$$

Denote by $\lambda_{\max}(\mathbf{A})$ the spectral radius of matrix $\mathbf{A}$. Then, if $\lambda_{\max}(\mathbf{R} \circ \mathbf{G}) < 1$, there exists a unique Nash equilibrium given by

$$\boldsymbol{y}_{gs} = (\mathbf{I} - \mathbf{R} \circ \mathbf{G})^{-1} \boldsymbol{\Lambda}_{gs} = \sum_{k=0}^{\infty} (\mathbf{R} \circ \mathbf{G})^{k} \boldsymbol{\Lambda}_{gs}. \tag{21}$$

6.2.2 Empirical analysis

An empirical test of this model is quite challenging and beyond the scope of the present paper. Still, in order to test for the robustness of our results with respect to peer effects, we restrict ourselves to the case where $\rho_j = \rho$ for all j and estimate (18) when $\rho_j = \rho$ for all j. That is,

$$y_{i,gs} = \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \frac{\phi^{d_{iL,gs}}}{2} \sum_{m=1}^{M} \beta_{1m} x_{L,gs}^{m} + \rho \sum_{j} g_{ij} y_{j,gs} + \eta_s + \epsilon_{i,gs}, \tag{22}$$

which incorporates both the leader effect as well as non-leader peer effects, as measured at an aggregated level by the average level of crime in the network. Table 6 displays the results. When we compare the results with that of Table 3 (estimating the same equation without controlling for non-leader peer effects), we see that the effect of criminal leaders, both in terms of technology and social distance, remains qualitatively the same. This indicates that, beyond non-leader peer effects, the distance to the leader in the criminal network has a distinct impact on delinquents.

[*Insert Table* 6 *here*]

7 Policy experiments

We now present the results of counterfactual exercises carried out to evaluate the importance of leaders in crime decisions. For that, we simulate the impact on criminal activities of a policy that removes all leaders from a school. It is worth noting that, in the simulation exercises, we assume that no new leader will emerge after the removal of current leaders. Thus, we need to interpret the outcomes of these simulations as *short-term* policies, because the model does not include dynamics.

As detailed in the Data section (Wave I), we have samples of 2,892 students inside a leader network and 3,665 students outside a leader network. However, the observable characteristics of these individuals are similar (see Table C.5). Our counterfactual exercises simulate the crime decisions and activities of students inside a leader network by using the parameter estimates of those outside a leader network. We next describe the implementation of this exercise in each stage.

7.1 Removing the leaders: Intensive margin

For the entire sample of students, our empirical model of criminal activity is

$$y_{i,gs} = \begin{cases} \omega - \theta\, d_{iL,gs} + \sum_{m=1}^{M} \beta_{1m} x_{i,gs}^{m} + \sum_{m=1}^{M} \beta_{2m} x_{L,gs}^{m} \\ + \sum_{m=1}^{M} \beta_{3m} x_{L,gs}^{m} d_{iL,gs} + \eta_s + \epsilon_{i,gs}, & \text{if } 0 \leq d_{iL,gs} \leq 7 \\ \\ \sum_{m=1}^{M} \beta_{4m} x_{i,gs}^{m} + \eta_s + \epsilon_{i,gs}, & \text{otherwise} \end{cases} \tag{23}$$

In other words, students within a leader network are affected by both their observable characteristics and the distance to the leader, while those outside a leader network are only affected by their observable characteristics. In our simulation exercises, we predict the delinquency level of each individual i with $y_{i,gs} > 0$ and $0 < d_{iL,g} \leq 7$ *after* the removal of all leaders using the estimated coefficients obtained from the second equation of the above model (23):[29]

$$\widehat{y}_{i,gs}^{after} = \sum_{m=1}^{M} \widehat{\beta}_{4m} x_{i,gs}^{m} + \widehat{\eta}_s$$

For each individual i, this value estimates the effect of his or her own productivity on his or her own crime. Total crime is thus $TC^{after} = \sum_i \widehat{y}_{i,gs}^{after}$.

We then compare TC^{after} with the original prediction of the model (i.e., before the removal of the leaders):

$$\widehat{y}_{i,gs}^{before} = \widehat{\omega} - \widehat{\theta}\, d_{iL,gs} + \sum_{m=1}^{M} \widehat{\beta}_{1m} x_{i,gs}^{m} + \sum_{m=1}^{M} \widehat{\beta}_{2m} x_{L,gs}^{m} + \sum_{m=1}^{M} \widehat{\beta}_{3m} x_{L,g}^{m} d_{iL,gs} + \widehat{\eta}_s$$

Total crime is $TC^{before} = \sum_i \widehat{y}_{i,gs}^{before}$. To determine the crime reduction, we calculate

$$\frac{TC^{after} - TC^{before}}{TC^{before}} \tag{24}$$

Figure 4 displays the average crime reduction given by (24). The left panel (NLS) shows that when we remove all the leaders from our sample, total crime is reduced by 22% on average. This effect differs for delinquents at a distance of one node from the leader and those located further away in the network. As a result, in Figure 5, we compare the actual criminal activities before

[29]From the theoretical model, $\omega 1_{L,gs} = \omega$ if there is a leader in the school and $\omega 1_{L,gs} = 0$ otherwise. As a result, when we remove all the leaders from the school, this term disappears.

the removal of the leaders (blue curve) with the model estimates when we remove all the leaders (orange curve) at *each distance level* to the leader. We see that the crime reduction is much higher for delinquents at a distance of one node to the leader (up to a 33% reduction in criminal activities) than those at a distance of four nodes to the leader (up to a 9% reduction in criminal activities). This effect vanishes for distances of five to seven nodes, where there is no statistical difference between the actual and estimated crime without a leader's influence. This figure thus shows that in the absence of a leader (i.e., criminal activities are only based on the delinquent's own characteristics), the students commit, on average, one crime irrespective of the distance to the leader.[30]

[*Insert Figures* 4 *and* 5 *here*]

7.2 Removing the leaders: Intensive and extensive margins

Let us perform the same exercises when controlling for selection into crime, that is when using Heckman's selection model. The first stage is given by

$$P(c_{i,gs}=1|\psi_{iL,gs}) = \begin{cases} \Lambda\left(\gamma_0 d_{iL,gs} + \sum_{m=1}^{M}\gamma_{1m} d_{iL,gs} x_{L,gs}^m + \right. \\ \left. + \sum_{m=1}^{M}\gamma_{2m} x_{i,gs}^m + \gamma_3 \overline{u}_{i,gs} + \eta_s + \epsilon_{i,gs}\right), & \text{if } 0 \leq d_{iL,gs} \leq 7 \\ \\ \Lambda\left(\sum_{m=1}^{M}\gamma_{4m} x_{i,gs}^m + \gamma_5 \overline{u}_{i,gs} + \eta_s + \epsilon_{i,gs}\right), & \text{otherwise} \end{cases} \tag{25}$$

where $\psi_{iL,gs} = (d_{iL,gs}, x_{L,gs}, x_{i,gs}, \overline{u}_{i,gs}, \eta_s, \epsilon_{i,gs})$.

We predict the probability of being a criminal for each individual i with $y_{i,gs} > 0$ and $0 < d_{iL,g} \leq 7$ after the removal of the leaders, using the estimated coefficients obtained from the second equation of the above model (25):

$$\widehat{P}^{after}(c_{i,gs}=1|\psi_{iL,gs}) = \Lambda\left(\sum_{m=1}^{M}\widehat{\gamma}_{4m} x_{i,gs}^m + \widehat{\gamma}_5 \overline{u}_{i,gs} + \widehat{\eta}_s\right)$$

[30] We replicate these policy exercises by using propensity scores to match the characteristics of individuals inside and outside networks. The results are qualitatively similar.

For each individual i, this value estimates the effect of his or her own characteristics on his or her probability of becoming a criminal. The average probability of becoming a criminal is equal to $PC^{after} = \frac{\sum_i \widehat{p}_{i,gs}^{after}}{N}$, where $\widehat{p}_{i,gs}^{after} = \widehat{P}^{after}(c_{i,gs} = 1|\psi_{iL,gs})$ and N is the number of criminals that belong to a leader network.

We again compare PC^{after} with the original prediction of the model:

$$\widehat{P}^{before}(c_{i,gs} = 1|\psi_{iL,gs}) = \Lambda\left(\widehat{\gamma}_0 d_{iL,gs} + \sum_{m=1}^{M} \widehat{\gamma}_{1m} d_{iL,gs} x_{L,gs}^m + \sum_{m=1}^{M} \widehat{\gamma}_{2m} x_{i,gs}^m + \widehat{\gamma}_3 \overline{u}_{i,gs} + \widehat{\eta}_s\right)$$

The average probability of becoming a criminal is $PC^{before} = \frac{\sum_i \widehat{p}_{i,gs}^{before}}{N}$, where $\widehat{p}_{i,gs}^{before} = \widehat{P}^{before}(c_{i,gs} = 1|\psi_{iL,gs})$ and N is the number of criminals that belong to a leader network. To determine the reduction in the average probability of becoming a criminal, we calculate

$$PC^{after} - PC^{before} \tag{26}$$

In the second stage, we first re-estimate model (23) for all individuals with $y_{i,gs} > 0$, controlling for selection into crime (i.e., the Mills ratio obtained from the first stage). Then, with the new sets of coefficients, we compute total crime before and after the removal of the leaders, as explained in Section 7.1. Finally, we calculate (24) to obtain the estimated crime reduction.

Figure 4 (middle and right panels) displays the results. It shows a decrease in both the extensive (i.e., probability of becoming a criminal) and the intensive (i.e., total amount of crime) margins when we eliminate the influence of the leader. Indeed, the average probability of becoming a criminal decreases by 11% when all leaders are removed.

As before, by using the previous estimations, we can compute the actual and estimated levels of crime for each possible distance within the criminal network, $d_{iL,g}$. This is reported in Figure 6, where the blue and orange curves represent actual criminal activities and the model estimates without the leader's influence, respectively. We see that the influence of a leader decreases as delinquents are located further away in the network. The effect is, however, different to the case without crime decisions (Figure 5). Indeed, criminals at distances of one to four nodes experience a relatively similar reduction in their delinquent activities on average when we remove the effects of the network leader (up to 20%). Again, the effects vanish for those criminals at distances of five to seven nodes.

[*Insert Figure* 6 *here*]

8 Concluding remarks

In this study, we propose a theory explaining why targeting crime leaders could be an effective way for reducing crime and test it using data on juvenile delinquency in US high-schools. Other studies in the literature have suggested that targeting the *key players* in social networks could be another effective policy for reducing crime (Ballester et al., 2006; Zenou, 2016). The empirical implementation of the key player policy, however, presents a number of empirical challenges and, most importantly, requires that the planner knows the exact topology of the network (Lee et al., 2021). In our approach, instead, crime leaders may be empirically defined using only information on outcomes. It can thus be the only alternative if network information is not available and too costly to obtain. If the policy maker has only partial knowledge about the network topology, then our theory suggests that "key leaders" should be targeted, that is individuals with high criminal activities and high closeness centrality. In order words, the crime leader property (high crime rate) is complementary to the connectivity property (high closeness centrality) in the proposed crime reduction policy.

References

[1] Acemoglu, D. and Jackson, M. O. (2014), "History, expectations, and leadership in the evolution of social norms," *Review of Economic Studies* 82, 423–456.

[2] Akerlof, G. A. (1997), "Social distance and social decisions," *Econometrica* 65(5), 1005–1027.

[3] Amemiya, J., Monahan, K. C. and Cauffman, E. (2016), "Leaders and followers in juvenile offending. Distinguishing correlates and adjustment to incarceration," *Criminal Justice and Behavior* 43, 899–922.

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

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

[6] Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O. (2013),"The diffusion of microfinance," *Science* 341 (6144), 1236498.

[7] Battaglini, M. and Patacchini, E. (2018), "Influencing connected legislators," *Journal of Political Economy* 126, 2277–2322.

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

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

[10] Bezin, E., Verdier, T. and Zenou, Y. (2021), "Crime, broken families and punishment," *American Economic Journal: Microeconomics*, forthcoming.

[11] Bolton, P., Brunnermeier, B. K. and Veldkamp, L. (2012), "Leadership, coordination and corporate culture," *Review of Economic Studies* 80, 512–537.

[12] Boucher, V. (2016), "Conformism and self-selection in social networks," *Journal of Public Economics* 136, 30–44.

[13] Burt, R. S. (1992), *Structural Holes*, New York: Academic Press.

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

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

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

[17] Christakis, N. and Fowler, J. H. (2010), "Social network sensors for early detection of contagious outbreaks," *PLOS One*: e12948 5(9).

[18] Cortés, D., Friebel, G. and Maldonado, D. (2019), Crime and education in a model of information transmission," *Annals of Public and Cooperative Economics* 91(1), 71–93.

[19] Damm, A. P. and Dustmann, C. (2014), "Does growing up in a high crime neighborhood affect youth criminal behavior?" *American Economic Review* 104, 1806–1832.

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

[21] Guo, G., Roettger, M. and Cai, T. (2008), "The integration of genetic propensities into social-control models of delinquency and violence among male youths," *American Sociological Review* 73, 543–568.

[22] Hauk, E. and Mueller, H. (2015), "Cultural leaders and the clash of civilizations," *Journal of Conflict Resolution* 59, 367–400.

[23] Hermalin, B. (2012), "Leadership and corporate culture," In: R. Gibbons and J. Roberts (Eds.), *Handbook of Organizational Economics*, Chap. 11, Princeton: Princeton University Press.

[24] Islam, A., Vlassopoulos, M., Zenou, Y. and Zhang, X. (2021), "Centrality-based spillover effects," CEPR Discussion Paper No. 16321.

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

[26] Jackson, M. O., Rogers, B. and Zenou, Y. (2017), "The economic consequences of social network structure," *Journal of Economic Literature* 55, 49–95.

[27] Kling, J. R., Ludwig, J. and Katz, L. F. (2005), "Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment," *Quarterly Journal of Economics* 120, 87–130.

[28] Lee, L-F., Liu, X., Patacchini, E. and Zenou, Y. (2021), "Who is the key player? A network analysis of juvenile delinquency," *Journal of Business and Economic Statistics* 39(3), 849–857.

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

[30] Lindquist, M. J. and Zenou, Y. (2019), "Crime and networks: 10 policy lessons," *Oxford Review of Economic Policy* 35(4), 746–771.

[31] 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.

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

[33] Mendenhall, M. E., Kühlmann, T. M. and Stahl, G. K. (2001), *Developing Global Business Leaders: Policies, Processes, and Innovations* Westport, CT: Quorum Books.

[34] Mehra, A., Kildu, M. and Brass, D. J. (2001), "The social networks of high and low self-monitors: Implications for workplace performance," *Administrative Science Quarterly* 46, 121–146.

[35] Myerson R. (2011), "Towards a theory of leadership and state-building," *Proceedings of the National Academy of Sciences of the USA* 108, 21297–21301.

[36] Padgett, J. F. and Ansell, C. K. (1993), "Robust action and the rise of the Medici 1400-1434," *American Journal of Sociology* 98, 1259–1319.

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

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

[39] Prummer, A. (2019), "Religious and cultural leaders," In: J-P. Carvalho, S. Iyer and J. Rubin (Eds.), *Advances in the Economics of Religion*, Palgrave Macmillan.

[40] Prummer, A. and Siedlarek, J-P. (2017), "Community leaders and the preservation of cultural traits," *Journal of Economic Theory* 168, 143–176.

[41] Putnam, R. D. (2000), *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon & Schuster.

[42] Reiss, A. (1988), "Co-offending and criminal careers," In: N. Morris and M. Tonry (Eds.), *Crime and Justice*, Chicago, IL: University of Chicago Press, pp. 55–68.

[43] Reiss, A. and Farrington, D. P. (1991), "Advancing knowledge about co-offending: Results from a prospective longitudinal survey of London males," *Journal of Criminal Law and Criminology* 82, 360–395.

[44] Tatsi, E. (2017), "Endogenous social interactions: Which peers matter?" Unpublished manuscript, SOFI, Stockholm University.

[45] Tao, J. and Lee, L-F. (2014), "A social interaction model with an extreme order statistic," *Econometrics Journal* 17, 197–240.

[46] Ushchev, P. and Zenou, Y. (2020), "Social norms in networks," *Journal of Economic Theory* 185, 104969.

[47] Van den Steen, E. (2010), "Interpersonal authority in a theory of the firm," *American Economic Review* 100, 466–490.

[48] Verdier, T. and Zenou, Y. (2015), "The role of cultural leaders in the transmission of preferences," *Economics Letters* 136, 158–161.

[49] Verdier, T. and Zenou, Y. (2018), "Cultural leaders and the dynamics of assimilation," *Journal of Economic Theory* 175, 374–414.

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

[51] Zenou, Y. (2016), "Key players," In: Y. Bramoullé, B. W. Rogers and A. Galeotti (Eds.), *Oxford Handbook on the Economics of Networks*, Oxford: Oxford University Press, pp. 244–274.

Table 1: Criminal Activities

Crime	Question	Category	Percentage
Fight	*In the past 12 months, how often did you get into a serious physical fight?*	Violent	**19%**
Steal < $50	*In the past 12 months, how often did you steal something worth less than $50?*	Non-Violent	**17%**
Damage	*In the past 12 months, how often did you deliberately damage property that didn't belong to you?*	Violent	**12%**
Group Fight	*In the past 12 months, how often did you take part in a physical fight where a group of your friends was against another group?*	Violent	**12%**
Hurt	*In the past 12 months, how often did you hurt someone badly enough in a physical fight that he or she needed care from a doctor or nurse?*	Violent	**10%**
Graffiti	*In the past 12 months, how often did you paint graffiti or signs on someone else's property or in a public place?*	Non-Violent	**7%**
Drugs	*In the past 12 months, how often did you sell marijuana or other drugs?*	Non-Violent	**7%**
Burglary	*In the past 12 months, how often did you go into a house or building to steal something?*	Non-Violent	**4%**
Steal > $50	*In the past 12 months, how often did you steal something worth more than $50?*	Non-Violent	**5%**
Knife/Gun	*During the past 12 months, how often you pulled a knife or gun on someone?*	Violent	**3%**
Threat	*In the past 12 months, how often did you use or threaten to use a weapon to get something from someone?*	Violent	**3%**
Shot/Stab	*During the past 12 months, how often you shot or stabbed someone?*	Violent	**1%**

Notes: Criminal activities are based on answers to Add Health questions from Wave I, Section 29: Delinquency Scale, where respondents are asked to report their recent delinquent or undesirable behaviors. Percentages are relative to the total amount of crime in the sample.

Table 2a: Criminal Activity, Distance to Leader and Moral Cost

	Dependent Variable: *Criminal Activity Index*					
	(1)	(2)	(3)	(4)	(5)	(6)
Crime Technology from Leader, ω	8.173***	9.481***	8.135***	9.956***	8.014***	10.284***
	(0.354)	(0.396)	(0.364)	(0.429)	(0.375)	(0.461)
Distance Dummies						
Distance to Leader = 1, θ_1	7.024***	7.981***				
	(0.228)	(0.221)				
Distance to Leader = 2, θ_2	7.154***	8.687***	7.183***	9.178***		
	(0.230)	(0.240)	(0.230)	(0.282)		
Distance to Leader = 3, θ_3	7.462***	8.912***	7.512***	9.416***	7.557***	9.723***
	(0.237)	(0.252)	(0.238)	(0.292)	(0.240)	(0.325)
Distance to Leader = 4, θ_4	7.541***	9.053***	7.597***	9.564***	7.639***	9.886***
	(0.247)	(0.268)	(0.249)	(0.307)	(0.253)	(0.340)
Distance to Leader = 5, θ_5	7.806***	9.233***	7.873***	9.722***	7.943***	10.028***
	(0.259)	(0.294)	(0.262)	(0.329)	(0.266)	(0.364)
Distance to Leader = 6, θ_6	7.897***	9.398***	7.939***	9.874***	7.988***	10.187***
	(0.281)	(0.328)	(0.283)	(0.365)	(0.288)	(0.401)
Distance to Leader = 7, θ_7	8.173***	9.481***	8.135***	9.956***	8.014***	10.284***
	(0.354)	(0.396)	(0.364)	(0.429)	(0.375)	(0.461)
Individual Moral Cost	-0.133***	-0.061	-0.116**	-0.046	-0.085	-0.000
	(0.045)	(0.045)	(0.054)	(0.052)	(0.067)	(0.064)
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Leader Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	2,892	2,892	2,247	2,247	1,593	1,593
Networks	353	353	353	353	353	353

Notes. We are testing equation (12). We consider students inside a leader network, i.e. d<8. Columns (2), (4) and (6) include school fixed effects. Columns (3) and (4) do not include individuals at distance 1, whereas columns (5) and (6) do not include individuals at distance 1 and 2. Robust standard errors in parenthesis. Leader networks are undirected. * p < .10, ** p < .05 and *** p < .01. For a description of the categories included in the *Criminal Activity Index*, see Table 1.

Table 2b: Criminal Activity and Distance to Leader

Null Hypothesis	*F-test P-values*					
	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$	$\tau=6$	$\tau=7$
H_0: $\theta_1 = \theta\tau$	0.000	0.000	0.000	0.000	0.000	0.000
H_0: $\theta_2 = \theta\tau$		0.068	0.016	0.006	0.003	0.000
H_0: $\theta_3 = \theta\tau$			0.353	0.093	0.037	0.000
H_0: $\theta_4 = \theta\tau$				0.353	0.146	0.000
H_0: $\theta_5 = \theta\tau$					0.518	0.000
H_0: $\theta_6 = \theta\tau$						0.000

Notes. F-tests for equality of the estimated coefficients of distance-to-leader dummies (column 2, Table 2a) are performed. P-values of the corresponding F-tests are reported.

Table 3: Criminal Activity, Distance to Leader and Moral Cost

		Dependent Variable: *Criminal Activity Index*						Dependent Variable: *Probability to Become a Criminal*		
		Panel I - *NLS*			Panel II - *Heckman 2nd Stage*			Panel III - *Heckman 1st Stage*		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Technology from Leader										
	ω	6.545***	7.646***	7.019***	8.981***	8.646***	7.421***	1.506***	1.617***	1.914***
		(0.817)	(0.807)	(1.183)	(1.303)	(1.574)	(2.569)	(0.294)	(0.279)	(0.288)
Distance to Leader										
	θ	0.935***	1.092***	1.003***	1.283***	1.235***	1.060***	0.215***	0.231***	0.273***
		(0.117)	(0.115)	(0.169)	(0.186)	(0.225)	(0.367)	(0.042)	(0.040)	(0.041)
	ϕ	0.898***	0.845***	0.868***	0.890***	0.874***	0.929***	0.881***	0.851***	0.777***
		(0.024)	(0.030)	(0.044)	(0.024)	(0.042)	(0.045)	(0.048)	(0.051)	(0.106)
	Marginal Effect	-0.276***	-0.154***	-0.204***	-0.421***	-0.326***	-0.472***	-0.030***	-0.025**	-0.012
		(0.043)	(0.059)	(0.074)	(0.097)	(0.130)	(0.113)	(0.010)	(0.010)	(0.010)
Individual Moral Cost		-0.060	-0.069	-0.027				-0.075***	-0.089***	-0.047*
		(0.045)	(0.050)	(0.063)				(0.016)	(0.023)	(0.028)
Mills' Ratio					3.273***	2.321**	3.461**			
					(0.843)	(0.892)	(1.410)			
Controls										
Individual Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Leader Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Observations		2,892	2,247	1,593	1,740	1,388	1,060	2,892	2,247	1,593

Notes: We are testing equation (11). We consider students inside a leader network, i.e. d<8. Of the total 6,557 observations, 2,892 represent individuals at a distance 7, or less, from the leader (i.e. they are inside a leader network). Column (1) includes all individuals in a leader network, whereas (2) does not include individuals at distance 1, and (3) does not include individuals at distances 1 and 2. Marginal effects are calculated at the mean value of the leader characteristics and include the baseline effect (i.e. distance to leader coefficient). Robust standard errors in parenthesis. * $p < .10$, ** $p < .05$ and *** $p < .01$. For a description of the categories included in the *Criminal Activity Index*, see Table 1.

Table 4a: Violent Criminal Activity, Distance to Leader and Moral Cost

		Dependent Variable: *Violent Criminal Activity Index*						Dependent Variable: *Probability to Become a Violent Criminal*		
		Panel I - *NLS*			Panel II - *Heckman 2nd Stage*			Panel III - *Heckman 1st Stage*		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Technology from Leader										
	ω	4.052***	4.511***	4.886***	5.356***	5.780***	6.210***	2.876***	3.274***	3.078***
		(1.106)	(1.136)	(1.194)	(1.295)	(1.405)	(1.503)	(0.275)	(0.328)	(0.414)
Distance to Leader										
	θ	0.534***	0.636***	0.591***	0.629***	0.565***	0.380	0.237***	0.233***	0.255***
		(0.079)	(0.074)	(0.110)	(0.165)	(0.188)	(0.253)	(0.043)	(0.054)	(0.057)
	ϕ	0.897***	0.837***	0.852***	0.906***	0.896***	0.949***	0.838***	0.839***	0.792***
		(0.029)	(0.035)	(0.053)	(0.042)	(0.058)	(0.062)	(0.063)	(0.092)	(0.128)
	Marginal Effect	-0.387***	-0.528***	-0.495***	-0.198***	-0.113	-0.135	-0.026***	-0.028**	-0.027
		(0.082)	(0.094)	(0.142)	(0.063)	(0.084)	(0.106)	(0.009)	(0.012)	(0.017)
Individual Moral Cost		-0.075**	-0.096***	-0.092**				-0.071***	-0.083***	-0.058*
		(0.030)	(0.033)	(0.043)				(0.023)	(0.029)	(0.035)
Mills' Ratio					1.618**	1.107	1.163			
					(0.730)	(0.764)	(1.065)			
Controls										
Individual Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Leader Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Observations		2,892	2,247	1,593	1,515	1,226	950	2,892	2,247	1,593

Notes: We are testing equation (11) for violent crimes. Violent crime includes *fight*, *group fight*, *threat*, *hurt*, *shot/stab*, *knife/gun* and *damage* (see Table 1). We consider students inside a leader network, i.e. d<8. Of the total 6,557 observations, 2,892 represent individuals at a distance 7, or less, from the leader (i.e. they are inside a leader network). Column (1) includes all individuals in a leader network, whereas (2) does not include individuals at distance 1, and (3) does not include individuals at distances 1 and 2. Marginal effects are calculated at the mean value of the leader characteristics and include the baseline effect (i.e. distance to leader coefficient). Robust standard errors in parenthesis. * p < .10, ** p < .05 and *** p < .01.

Table 4b: Non-Violent Criminal Activity, Distance to Leader and Moral Cost

		Dependent Variable: *Non-Violent Criminal Activity Index*						Dependent Variable: *Probability to Become a Non-Violent Criminal*		
		Panel I - *NLS*			Panel II - *Heckman 2nd Stage*			Panel III - *Heckman 1st Stage*		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Technology from Leader										
	ω	1.828**	1.853**	1.841*	0.930	1.609	0.928	0.559**	0.503*	0.164
		(0.883)	(0.896)	(0.958)	(1.317)	(1.509)	(1.965)	(0.256)	(0.298)	(0.360)
Distance to Leader										
	θ	0.404***	0.462***	0.416***	0.652***	0.249	0.485	0.210***	0.225***	0.244***
		(0.069)	(0.073)	(0.109)	(0.222)	(0.286)	(0.374)	(0.050)	(0.054)	(0.048)
	ϕ	0.896***	0.853***	0.884***	0.893***	0.983***	0.956***	0.854***	0.824***	0.725***
		(0.034)	(0.043)	(0.060)	(0.048)	(0.061)	(0.089)	(0.073)	(0.101)	(0.139)
	Marginal Effect	-0.300***	-0.379***	-0.338***	-0.235**	-0.187*	-0.326***	-0.020***	-0.019*	-0.007
		(0.073)	(0.090)	(0.130)	(0.092)	(0.113)	(0.105)	(0.008)	(0.010)	(0.015)
Individual Moral Cost		0.015	0.027	0.063				-0.053**	-0.046*	-0.018
		(0.027)	(0.030)	(0.039)				(0.022)	(0.026)	(0.031)
Mills' Ratio					2.403*	1.175	2.014			
					(1.382)	(1.523)	(1.833)			
Controls										
Individual Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Leader Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Observations		2,892	2,247	1,593	1,077	895	727	2,892	2,247	1,593

Notes: We are testing equation (11) for non-violent crimes. Non-violent crime includes *steal < $50*, *graffiti*, *drugs*, *burglary* and *steal > $50* (see Table 1). We consider students inside a leader network, i.e. d<8. Of the total 6,557 observations, 2,892 represent individuals at a distance 7, or less, from the leader (i.e. they are inside a leader network). Column (1) includes all individuals in a leader network, whereas (2) does not include individuals at distance 1, and (3) does not include individuals at distances 1 and 2. Marginal effects are calculated at the mean value of the leader characteristics and include the baseline effect (i.e. distance to leader coefficient). Robust standard errors in parenthesis. * p < .10, ** p < .05 and *** p < .01.

Table 5: Robustness Check 1 (measuring the leader as the top 10 percentile of the individual crime index distribution)

		Dependent Variable: *Criminal Activity Index*						Dependent Variable: *Probability to Become a Criminal*		
		Panel I - *NLS*			Panel II - *Heckman 2nd Stage*			Panel III - *Heckman 1st Stage*		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Technology from Leader										
	ω	7.299***	7.141***	7.348***	8.177***	8.152***	8.318***	2.864***	3.187***	3.135***
		(2.086)	(2.448)	(2.535)	(2.190)	(2.688)	(2.921)	(0.293)	(0.344)	(0.442)
Distance to Leader										
	θ	1.079***	1.249***	1.329***	1.396***	1.466***	1.536***	0.214***	0.227***	0.253***
		(0.113)	(0.107)	(0.123)	(0.168)	(0.172)	(0.239)	(0.071)	(0.072)	(0.071)
	ϕ	0.876***	0.806***	0.736***	0.869***	0.803***	0.789***	0.897***	0.841***	0.742***
		(0.022)	(0.029)	(0.050)	(0.028)	(0.039)	(0.066)	(0.088)	(0.111)	(0.147)
	Marginal Effect	-0.827***	-1.103***	-1.274***	-0.425***	-0.191*	-0.214	-0.041***	-0.023*	-0.009
		(0.128)	(0.134)	(0.146)	(0.097)	(0.130)	(0.140)	(0.010)	(0.014)	(0.022)
Individual Moral Cost		-0.060	-0.063	-0.093				-0.090***	-0.103***	-0.085**
		(0.045)	(0.050)	(0.060)				(0.023)	(0.029)	(0.038)
Mills' Ratio					3.326***	2.481**	3.417**			
					(0.979)	(1.012)	(1.486)			
Controls										
Individual Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Leader Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Observations		2,725	2,071	1,375	1,613	1,233	882	2,725	2,071	1,375

Notes: We are testing equation (11). We consider students inside a leader network, i.e. d<8. Of the total 6,557 observations, 2,725 represent individuals at a distance 7, or less, from the leader (i.e. they are inside a leader network). Column (1) includes all individuals in a leader network, whereas (2) does not include individuals at distance 1, and (3) does not include individuals at distances 1 and 2. Marginal effects are calculated at the mean value of the leader characteristics and include the baseline effect (i.e. distance to leader coefficient). Robust standard errors in parenthesis. * p < .10, ** p < .05 and *** p < .01. For a description of the categories included in the *Criminal Activity Index*, see Table 1.

Table 6: Robustness Check 2 (including non-leader peer effects)

		Dependent Variable: *Criminal Activity Index*						Dependent Variable: *Probability to Become a Criminal*		
		Panel I - *NLS*			Panel II - *Heckman 2nd Stage*			Panel III - *Heckman 1st Stage*		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Technology from Leader										
	ω	6.314***	6.662***	6.955***	6.842***	7.662***	8.027***	2.678***	2.933***	2.765***
		(1.641)	(1.709)	(1.789)	(1.816)	(0.202)	(2.085)	(0.283)	(0.342)	(0.431)
Distance to Leader										
	θ	0.921***	1.089***	1.004***	1.279***	1.279***	1.145***	0.218***	0.237***	0.284***
		(0.120)	(0.116)	(0.169)	(0.175)	(0.202)	(0.279)	(0.052)	(0.059)	(0.064)
	ϕ	0.903***	0.847***	0.868***	0.892***	0.880***	0.926***	0.870***	0.840***	0.786***
		(0.025)	(0.030)	(0.044)	(0.028)	(0.039)	(0.040)	(0.062)	(0.081)	(0.101)
	Marginal Effect	-0.675***	-0.896***	-0.826***	-0.424***	-0.356***	-0.479***	-0.025***	-0.017	-0.007
		(0.122)	(0.059)	(0.212)	(0.081)	(0.098)	(0.107)	(0.009)	(0.012)	(0.015)
Individual Moral Cost		-0.063	-0.069	-0.027				-0.070***	-0.078***	-0.037
		(0.045)	(0.050)	(0.063)				(0.023)	(0.028)	(0.037)
Mills' Ratio					3.412***	2.309***	4.738***			
					(1.019)	(1.057)	(1.387)			
Individual Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Leader Characteristics		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Peer Crime		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations		2,892	2,247	1,593	1,740	1,388	1,060	2,892	2,247	1,593

Notes: We are testing equation (22). We consider students inside a leader network, i.e. d<8. Of the total 6,557 observations, 2,892 represent individuals at a distance 7, or less, from the leader (i.e. they are inside a leader network). Column (1) includes all individuals in a leader network, whereas (2) does not include individuals at distance 1, and (3) does not include individuals at distances 1 and 2. Marginal effects are calculated at the mean value of the leader characteristics and include the baseline effect (i.e. distance to leader coefficient). Robust standard errors in parenthesis. * p < .10, ** p < .05 and *** p < .01. For a description of the categories included in the *Criminal Activity Index*, see Table 1.

Figure 1. Distribution of Students by Distance to the Leader

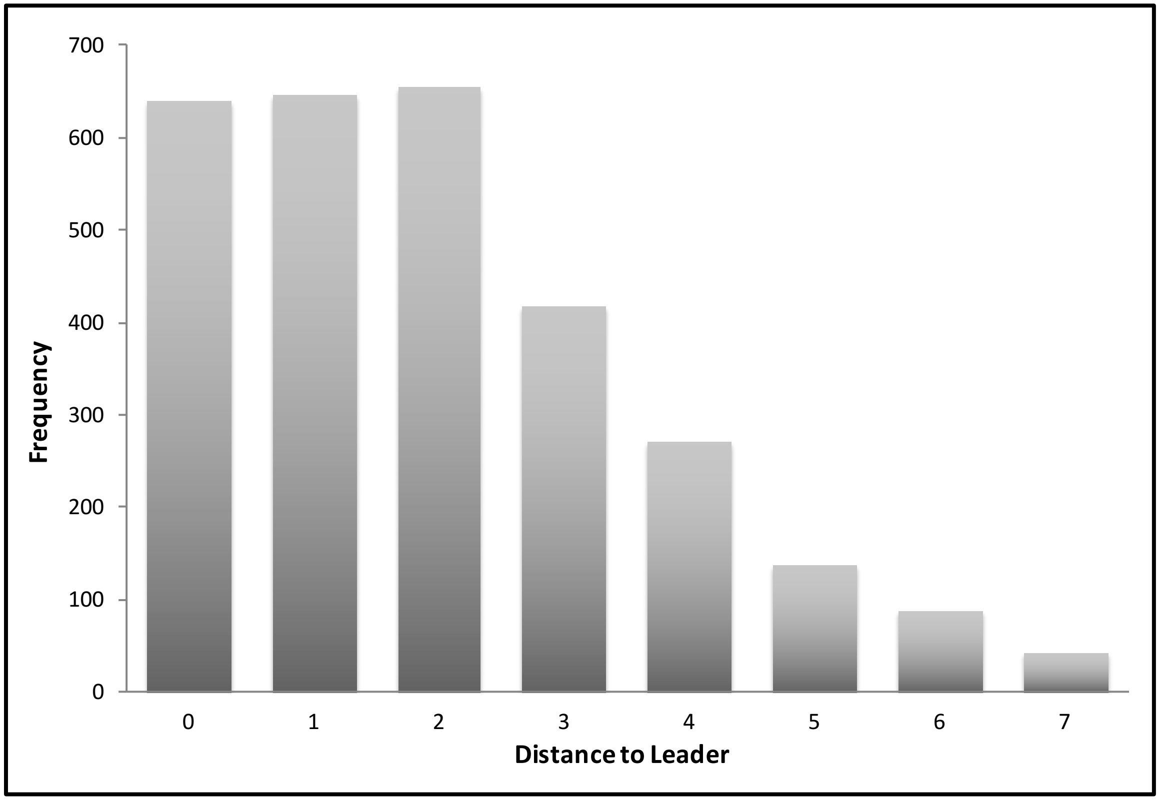


Figure 2. Distance to the Leader and Criminal Activity

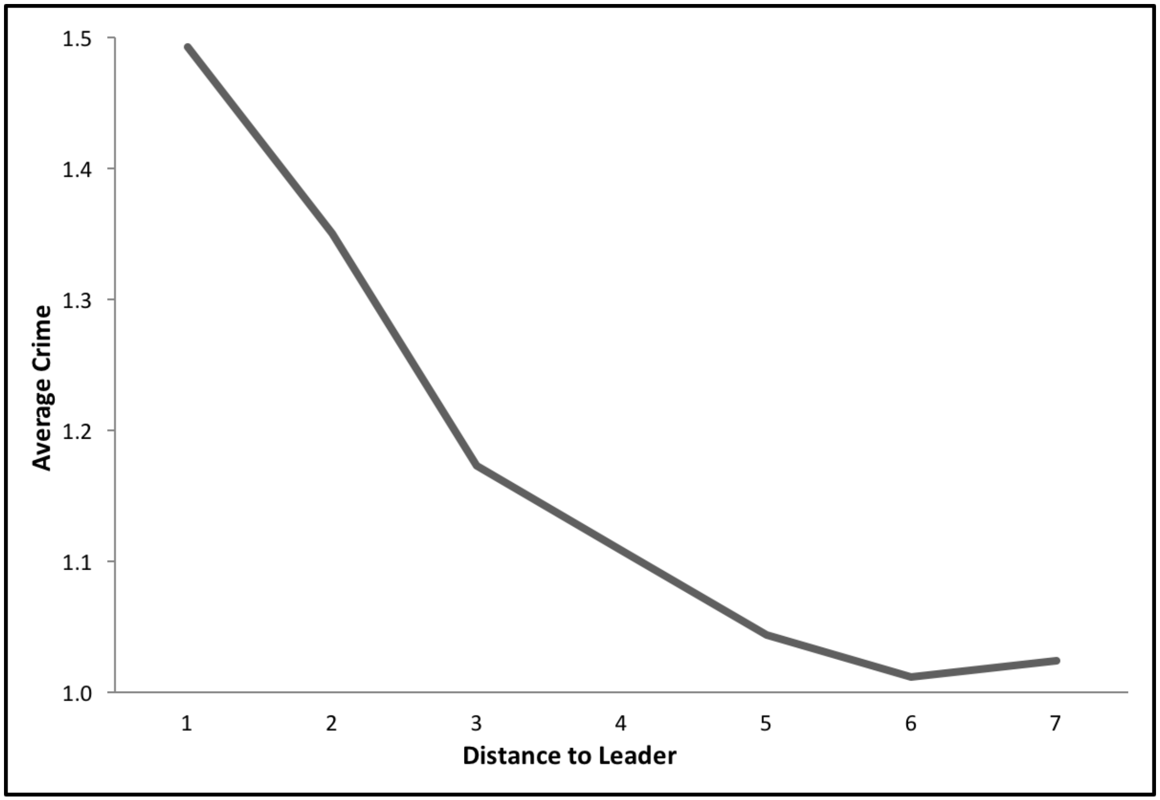
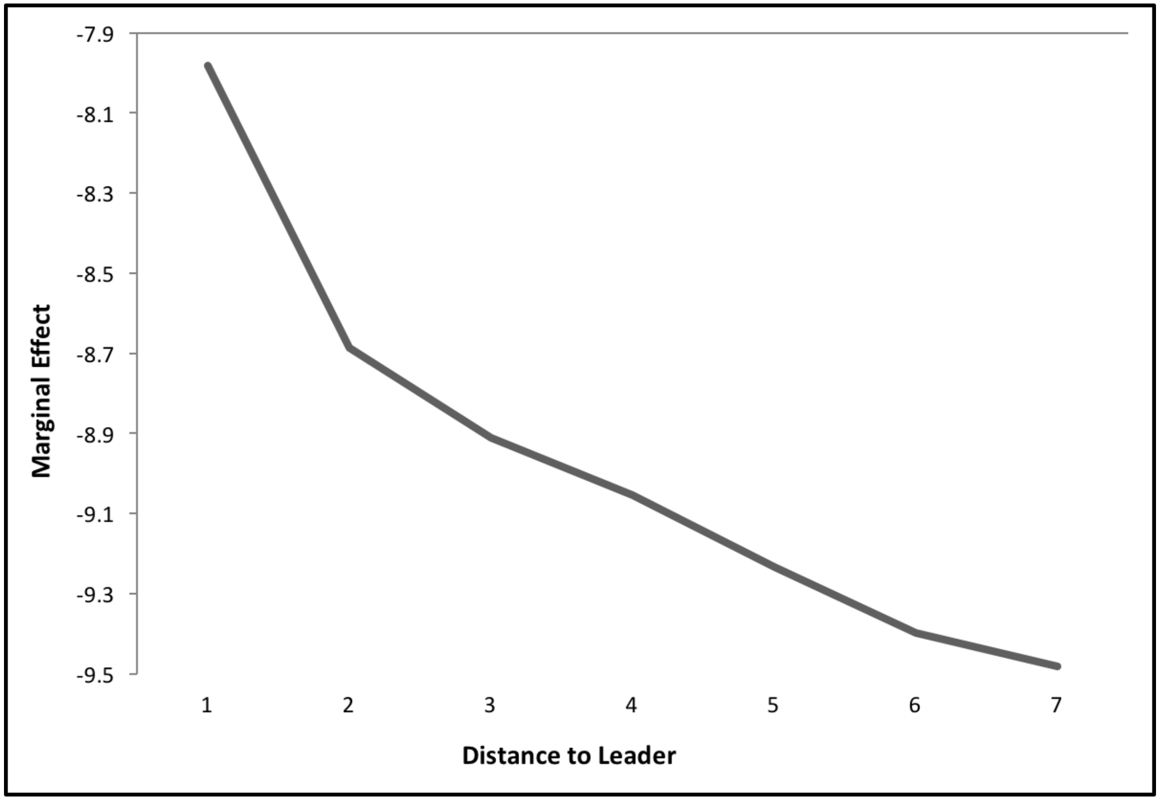


Figure 3. Marginal Effects



** Note: Estimation results from Table 2, column (2).*

Figure 4. Average Crime Reduction from Removing Leaders

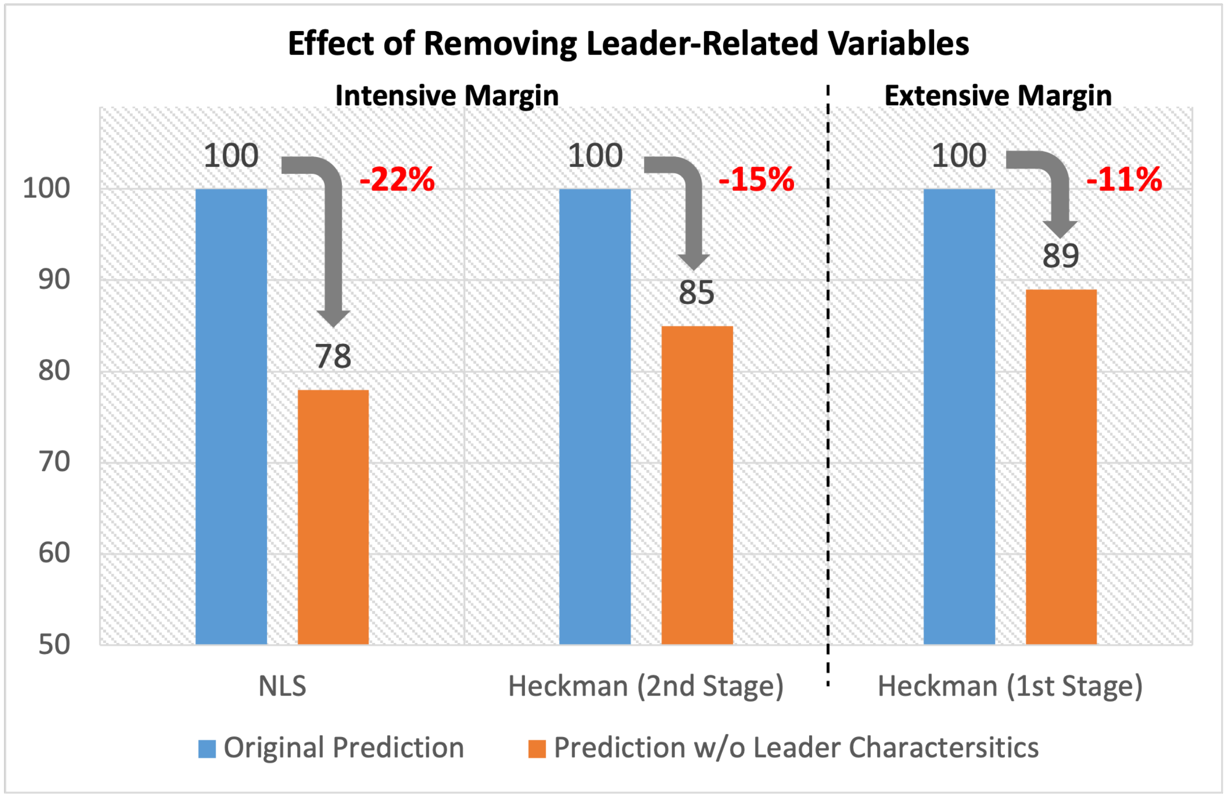


Figure 5. Crime Reduction from Removing Leaders by Distance to Leader

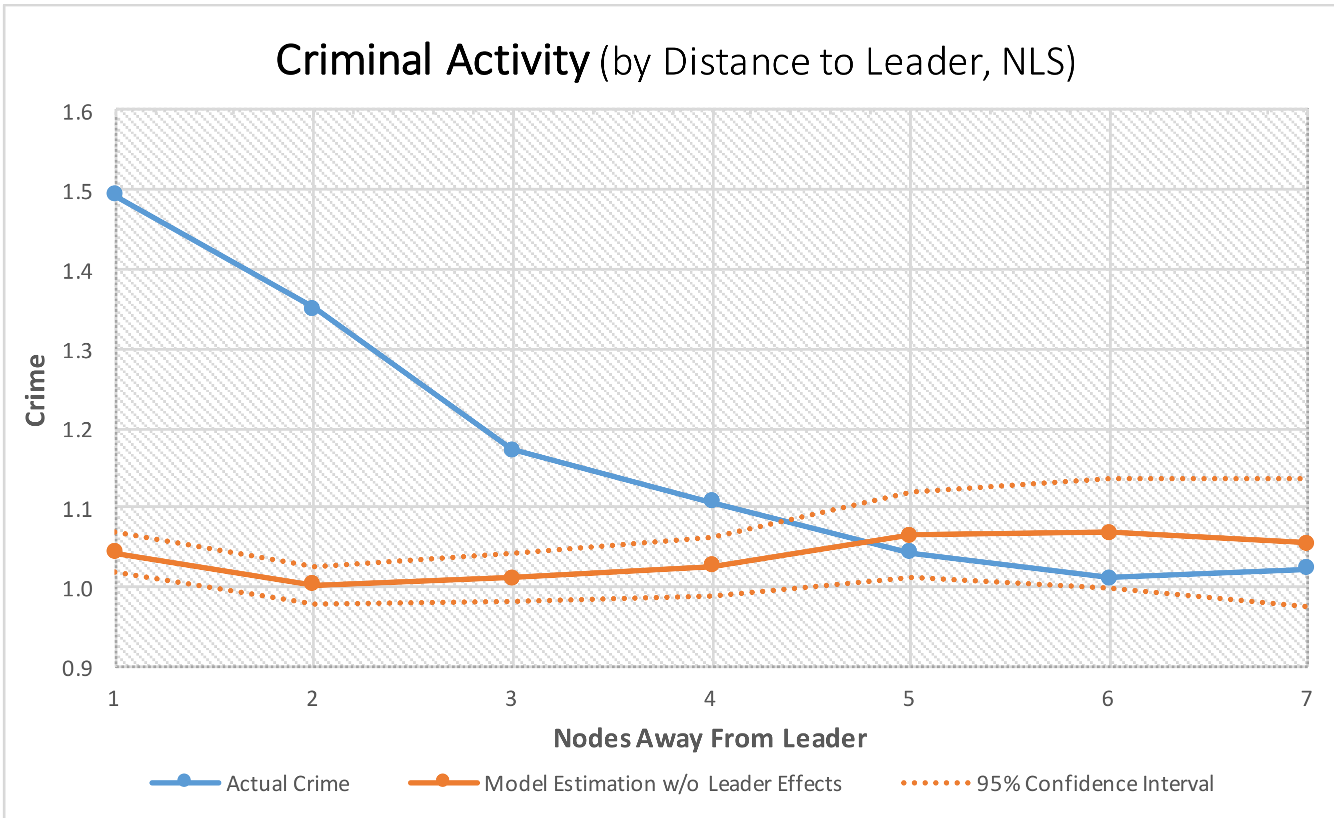
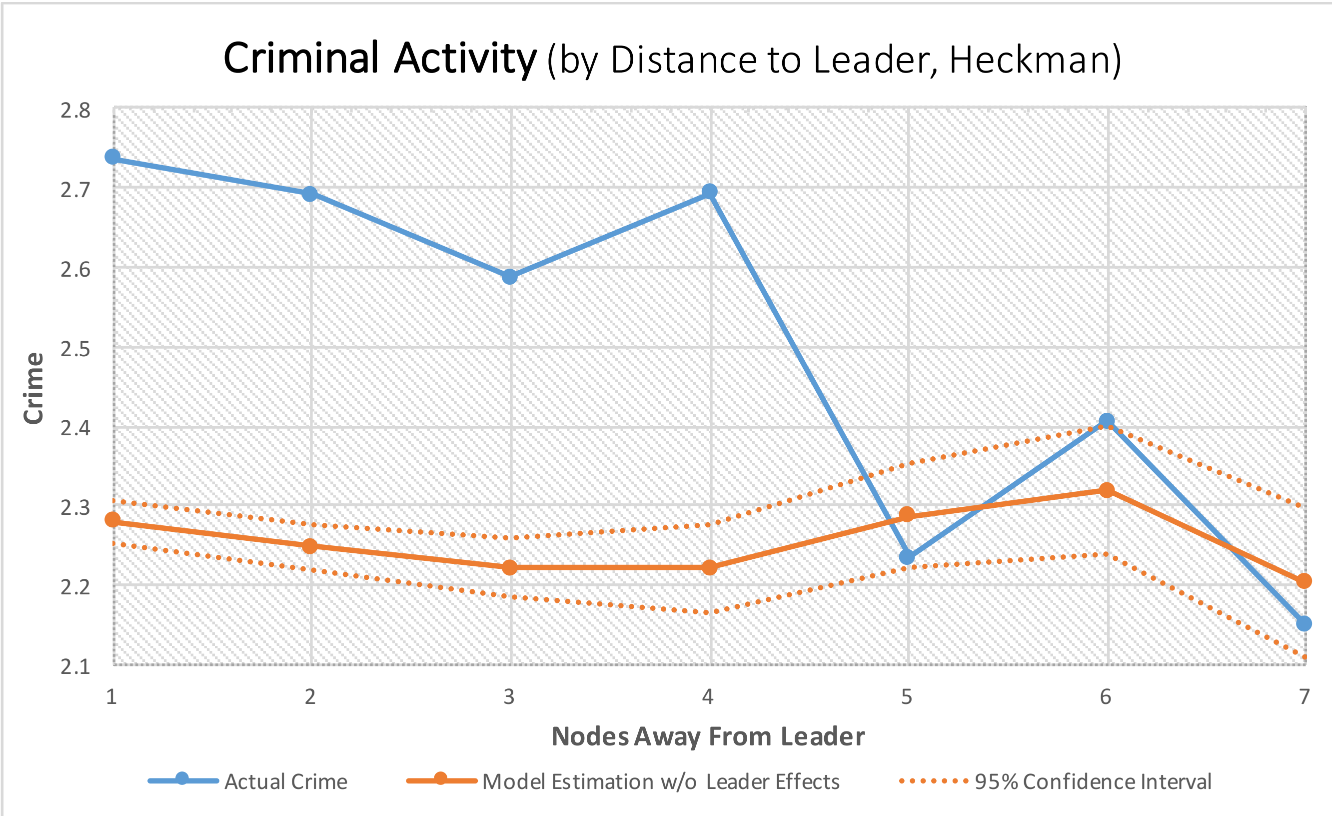


Figure 6. Crime Reduction from Removing Leaders by Distance to Leader



Appendix

A Extensions of the theoretical model

A.1 A different conformism behavior

In our model, the utility function is given by (4), where the conformist component is equal to $\frac{1}{4}\left(y_{i,gs} - \phi^{d_{iL,gs}} y_{L,gs}\right)^2$. We could have modeled the conformist component as $\frac{\phi^{d_{iL,gs}}}{4}\left(y_{i,gs} - y_{L,gs}\right)^2$, which implies that the conformist taste parameter is itself declining with the distance to the leader. The intuition of this new formulation is that the further away is a person from the leader, the lower is his or her taste for conformity. Utility is now given by

$$u_{i,gs}(\boldsymbol{y}_{gs}) = \left[\beta^1_{i,gs} + \omega^1_{gs} - \theta^1 d_{iL,gs} + \eta^1_s + \epsilon^1_{i,gs}\right] y_{i,gs} - \frac{1}{4} y^2_{i,gs} - \frac{\phi_1^{d_{iL,gs}}}{4}\left(y_{i,gs} - y_{L,gs}\right)^2$$

We still have

$$\begin{aligned} \frac{\partial u_{i,gs}}{\partial y_{i,gs}\partial d_{iL,gs}} &= -\theta^1 + \frac{1}{2}\left(y_{L,gs} - y_{i,gs}\right)\phi_1'(d_{iL,gs}) < 0, \ \frac{\partial u_{i,gs}}{\partial y_{i,gs}\partial \phi_1^{d_{iL,gs}}} = -\frac{1}{2}\left(y_{L,gs} - y_{i,gs}\right) < 0 \\ \frac{\partial u_{i,gs}}{\partial y_{i,gs}\partial y_{L,gs}} &= \frac{1}{2}\phi_1^{d_{iL,gs}} > 0 \end{aligned}$$

where $\phi_1(d_{iL,gs}) \equiv \phi_1^{d_{iL,gs}}$ with $\phi_1'(d_{iL,gs}) < 0$ and $y_{L,gs} > y_{i,gs}$. Hence, our model and this one see the same impact of the distance to the leader on an individual's own utility.

Assuming that $\omega^1_{gs} - \theta^1 d_{iL,gs} > 0$, we easily obtain the equilibrium criminal activity for each individual i, gs:

$$y_{i,gs} = \frac{2\left(\beta^1_{i,gs} + \omega^1_{gs} + \eta^1_{gs} + \epsilon^1_{i,gs}\right)}{\left(1 + \phi_1^{d_{iL,gs}}\right)} - \frac{2\theta^1}{\left(1 + \phi_1^{d_{iL,gs}}\right)} d_{iL,gs} + \frac{\phi_1^{d_{iL,gs}}}{\left(1 + \phi_1^{d_{iL,gs}}\right)} \beta_{L,gs}$$

where $y_{L,gs} = \beta_{L,gs}$. If we adopt the following normalizations,

$$\alpha \equiv \frac{2}{\left(1 + \phi_1^{d_{iL,gs}}\right)}, \ \ \beta_{i,gs} \equiv \alpha\beta^1_{i,gs}, \ \ \omega_{gs} \equiv \alpha\omega^1_{gs}, \ \ \eta_{gs} \equiv \alpha\eta^1_{gs}, \ \ \epsilon_{i,gs} \equiv \alpha\epsilon^1_{i,gs}, \ \ \theta \equiv \alpha\theta^1,$$

and $\phi^{d_{iL,gs}} \equiv \frac{\phi_1^{d_{iL,gs}}}{\left(1+\phi_1^{d_{iL,gs}}\right)}$, then, we obtain an equation similar to (11), which is given by

$$y_{i,gs} = \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \phi^{d_{iL,gs}} \beta_{L,gs} + \eta_{gs} + \epsilon_{i,gs}$$

A.2 The leader's crime decision is a function of the total crime in the network

In our model, we assume that the leader does not compare his or her criminal activity with anybody else; in other words, his or her crime productivity is only influenced by his or her own characteristics (see (5)). Hence, we assume that the leader's crime decision is *independent* of the crime decisions of the other individuals in the network. Let us relax this assumption and assume that the utility function of the leader is no longer given by (5) but by

$$u_{L,gs}(y_{L,g}) = \left(\beta_{L,gs} + \sum_{i \in g} y_{i,gs}\right) y_{L,gs} - \frac{1}{2} y_{L,gs}^2 \tag{A.1}$$

where $\beta_{L,gs}$ is still defined by (6). In (A.1), we have added the term $\sum_{i \in gs} y_{i,gs} y_{L,gs}$ into the utility function of the leader, which captures the (endogenous) total crime activity in the leader's network. The first-order condition yields

$$y_{L,gs}^* = \beta_{L,gs} + \sum_{i \in gs} y_{i,gs}^* \tag{A.2}$$

where $y_{i,gs}^*$ is given by (8), which we rewrite for the sake of presentation as follows:

$$y_{i,gs}^* = \pi_{i,gs} - \theta d_{iL,gs} + \frac{\phi^{d_{iL,gs}}}{2} y_{L,gs}^* \text{ for all } i \in gs$$

where $\pi_{i,gs} \equiv \beta_{i,gs} + \omega_{gs} + \eta_{gs} + \epsilon_{i,gs}$. Plugging this value of $y_{i,gs}^*$ into (A.2) leads to

$$\begin{aligned} y_{L,gs}^* &= \frac{\beta_{L,gs} + \sum_{i \in gs} \left(\pi_{i,gs} - \theta d_{iL,gs}\right)}{\left(1 - \frac{1}{2} \sum_{i \in gs} \phi^{d_{iL,gs}}\right)} \\ &= \frac{\beta_{L,gs} + \Omega_{gs}}{K_{gs}} \end{aligned}$$

where

$$\Omega_{gs} \equiv \sum_{i \in gs} \left(\pi_{i,gs} - \theta d_{iL,gs}\right) \text{ and } K_{gs} \equiv 1 - \frac{1}{2} \sum_{i \in gs} \phi^{d_{iL,gs}}$$

As a result,

$$\begin{aligned} y^*_{i,gs} &= \pi_{i,gs} - \theta d_{iL,gs} + \frac{\phi^{d_{iL,gs}}}{2}\left(\frac{\beta_{L,gs}+\Omega_{gs}}{K_{gs}}\right) \\ &= \pi_{i,gs} \underbrace{-\theta d_{iL,gs} + \frac{\phi^{d_{iL,gs}}}{2}\frac{\Omega_{gs}}{K_{gs}}}_{\text{direct effect of leader distance}} + \underbrace{\frac{\phi^{d_{iL,gs}}}{2}\frac{\beta_{L,gs}}{K_{gs}}}_{\text{cross effect of distance and leader characteristics}} \end{aligned}$$

Although the model is more complicated, we obtain similar results to those in the main text, namely showing the direct effect of the distance to the leader and the cross-effect of the distance to the leader and the leader's characteristics on the delinquent's criminal activities.

B Endogenous leaderhip

In this main text, we assume that each network has a crime leader. Here, we introduce the possibility of endogenous crime leadership. That is, we explain how one becomes a leader.

Consider a given network g and assume two stages. In the first stage, i.e., stage 0, the criminal individual $i^{\max}_{gs}$ with the highest value of $\pi^0_{i,gs} = \beta_{i,gs} + \eta_{gs} + \epsilon_{i,gs}$ (i.e., $i^{\max}_{gs} = \arg\max \pi^0_{i,gs}$) in the network decides whether or not to become a leader in the network. In the second stage, i.e., stage 1, when there is no leader, individuals decide their crime effort with the same crime technology $\pi^0_{i,gs}$ and no role model. However, when there is a leader, the leader implements at a cost C in the network a more efficient technology of crime $\pi^0_{i,gs} + \omega_{gs}$ shared by her network fellows. Further, being a role model, the leader obtains in return some "social prestige" from her peers, which is proportional to the level of crime under her leadership.

To be more precise, suppose that individual $i^{\max}_{gs} = L$ becomes a leader in the second stage $t = 1$. Then, her utility can be written as

$$u^1_{L,gs}(y^1_{L,gs}) = \pi^0_{i,gs}\, y^1_{L,gs} - \frac{1}{2}\left(y^1_{L,gs}\right)^2 + s\sum_{\substack{i\in g\\ i\neq L}} y^1_{i,gs} - C \qquad \text{(B.1)}$$

where $s\sum_{i\in g, i\neq L} y^1_{i,gs}$ is the "prestige" component of the leader ($y^1_{i,gs}$ is the crime effort of any delinquent who is not the leader), which is increasing in the level of crime in the network under her leadership and C is the cost to implement the leadership technology. As usual, we solve the model by backward induction.

B.1 Second stage

(i) **Crime effort with a leader**

We have already solved this stage. If person i is a criminal in a network with a leader L, then her equilibrium effort is given by equation (8), or[1]

$$y_{i,gs}^{1,L} = \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \eta_s + \epsilon_{i,gs} + \frac{\phi^{d_{iL,gs}}}{2} y_{L,gs}^1.$$

Her equilibrium utility is the given by (see (9)):

$$u_{i,gs}^{1,L}(\boldsymbol{y}_{gs}) = \frac{1}{2}\left(\pi_{i,gs}^0 + \omega_{gs} - \theta d_{iL,gs}\right)^2 - \frac{1}{8}\left(\phi^{d_{iL,gs}} y_{L,gs}^1\right)^2 + \frac{1}{2}\left(\pi_{i,gs}^0 + \omega_{gs} - \theta d_{iL,gs}\right)\phi^{d_{iL,gs}} y_{L,gs}^1. \tag{B.2}$$

Further, maximizing $u_{L,gs}^1(y_{L,gs}^1)$ given in (B.1) with respect to $y_{L,gs}^1$ leads to: $y_{L,gs}^1 = \pi_{i,gs}^0 = \pi_{L,gs}^0$. Plugging this value in the leader's utility function (B.1) and using (8) leads to:

$$\begin{aligned} u_{L,gs}^1(\boldsymbol{y}_{gs}) &= \frac{1}{2}\left(\pi_{L,gs}^0\right)^2 + s\sum_{\substack{i\in g\\ i\neq L}}\left(\beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \eta_s + \epsilon_{i,gs} + \frac{\phi^{d_{iL,gs}}}{2} y_{L,gs}^1\right) - C \\ &= \frac{1}{2}\left(\pi_{L,gs}^0\right)^2 + s\left[\Omega_{gs}^L + \omega_{gs} N_{gs}\right] + (1 - K_{gs}^L) y_{L,gs}^1) - C \\ &= \frac{1}{2}\left(\pi_{L,gs}^0\right)^2 + s\left[\Omega_{gs}^L + \omega_{gs} N_{gs}\right] + s(1 - K_{gs}^L)(\pi_{L,gs}^0) - C \end{aligned}$$

where $\Omega_{gs}^L \equiv \sum_{i\in gs}\left(\pi_{i,gs}^0 - \theta d_{iL,gs}\right)$, $K_{gs}^L \equiv 1 - \frac{1}{2}\sum_{i\in gs}\phi^{d_{iL,gs}}$, and N_{gs} is the size of network g in school s.

(ii) **Crime effort without a leader**

For each individual in a network without a leader, her equilibrium effort is given by:

$$y_{i,gs} = \beta_{i,gs} + \eta_s + \epsilon_{i,gs} = \pi_{i,gs}^0$$

and the equilibrium utility is equal to

$$u_{i,gs}^0(\boldsymbol{y}_{gs}) = \frac{1}{2}\left(\pi_{i,gs}^0\right)^2 \tag{B.3}$$

[1] The subscript L refers to the leader while the superscript L refers to a non-leader criminal under the regime when a leader exists.

B.2 First stage: Leadership entry

Individual $i_{gs}^{\max} = L$ becomes a leader if and only if $u_{L,gs}^1(\boldsymbol{y}_{gs}) \geq u_{L,gs}^0(\boldsymbol{y}_{gs})$. That is,

$$\frac{1}{2}\left(\pi_{L,gs}^0\right)^2 + s\left[\Omega_{gs}^L + \omega_{gs}N_{gs}\right] + s(1-K_{gs}^L)(\pi_{L,gs}^0 + \omega_{gs}) - C \geq \frac{1}{2}\left(\pi_{L,gs}^0\right)^2$$

There is a value of $\pi_{L,gs}^0$, denoted by $\widetilde{\pi}_{L,gs}^0$, for which individual $i_{gs}^{\max} = L$ is indifferent between being a leader in her network and not being a leader:

$$\widetilde{\pi}_{L,gs}^0 = \frac{C - s\left[\Omega_{gs}^L + \omega_{gs}N_{gs}\right]}{s(1-K_{gs}^L)} - \omega_{gs} \tag{B.4}$$

A network with a leader will be observed when $\pi_{L,gs}^0 \geq \widetilde{\pi}_{L,gs}^0$ and, thus, the crime efforts of all individuals in the network can be written as

$$y_{i,gs}^* = \begin{cases} \beta_{i,gs} + \omega_{gs} - \theta d_{iL,gs} + \eta_s + \epsilon_{i,gs} + \frac{\phi^{d_{iL,gs}}}{2} y_{L,gs}^* & \text{if} \quad \pi_{L,gs}^0 \geq \widetilde{\pi}_{L,gs}^0 \\ \beta_{i,gs} + \eta_s + \epsilon_{i,gs} & \text{if} \quad \pi_{L,gs}^0 < \widetilde{\pi}_{L,gs}^0 \end{cases} \tag{B.5}$$

Thus, the model we develop in the main text is valid under the assumption that $\pi_{L,gs}^0 \geq \widetilde{\pi}_{L,gs}^0$.

Appendix C: Data Appendix

Table C.1: Description of Data (Total Sample and Criminals, N=6,557)

		Total 6,557 (100%)				Criminals 3,360 (51%)			
	Variable definition	mean	sd	min	max	mean	sd	min	max
Criminal Activity Index	Sum of the total criminal activity during the last 12 months.	1.8	3.3	0	34	3.6	3.8	1	34
Criminal	Dummy variable taking value one if the respondent is a criminal.	0.5	0.5	0	1	1.0	0.0	1	1
Distance to Leader*	Geodesic distance (length of the shortest path) between the individual and the criminal network leader (individual whose criminal activity exceeds the school median in 3 median average deviations or more).	2.0	1.7	0	7	1.6	1.7	0	7
Moral Cost	Parental religion practice, where 0 = "no religion", 1 = "never", 2 = "less than once a month", 3 = "less than once a week, but at least once a month", and 4 = "once a week or more". Response to the question to parents: "How often have you gone to religious services in the past year?"	2.8	1.2	0	4	2.7	1.3	0	4
Female	Dummy variable taking value one if the respondent is female.	0.5	0.5	0	1	0.4	0.5	0	1
Student Grade	Grade of student in the current year, ranging from 7 to 12.	9.5	1.6	7	12	9.5	1.6	7	12
Non-White	Dummy variable taking value one if race is not white.	0.4	0.5	0	1	0.4	0.5	0	1
Math Score	Grade in mathematics, ranging from 1 (D or lower) to 4 (A).	2.7	1.0	1	4	2.5	1.0	1	4
Parent Education	Schooling level of the (biological or non-biological) father (mother, in case father education is not available), coded as 9 - "some school", 11 - "no completed high school, or a business, trade, or vocational school instead of high school", 13 - "high school graduate", 15 - "completed GED; or business, trade, or vocational school after high school; or some college", 17 - "graduated from college or a university", 19 - "professional training beyond a four-year college".	14.5	2.6	9	19	14.3	2.6	9	19
Residential Building Quality	Interviewer's response to the question "How well kept is the building in which the respondent lives?", coded as 1 - "very poorly kept (needs major repairs)", 2 - "poorly kept (needs minor repairs)", 3 - "fairly well kept (needs cosmetic work)", 4 - "very well kept".	3.4	0.8	1	4	3.4	0.8	1	4
Network Size	Number of individuals in the leader network only.	353	363	2	860	328	361	2	860

* *Includes observations inside the leader network only.*

Table C.2: Description of Data (Non-Criminals and Criminals, N=6,557)

		Non-Criminals 3,197 (49%)				**Criminals 3,360 (51%)**			
	Variable definition	mean	sd	min	max	mean	sd	min	max
Criminal Activity Index	Sum of the total criminal activity during the last 12 months.	0.0	0.0	0	0	3.6	3.8	1	34
Criminal	Dummy variable taking value one if the respondent is a criminal.	0.0	0.0	0	0	1.0	0.0	1	1
Distance to Leader*	Geodesic distance (length of the shortest path) between the individual and the criminal network leader (individual whose criminal activity exceeds the school median in 3 median average deviations or more).	2.7	1.5	1	7	1.6	1.7	0	7
Moral Cost	Parental religion practice, where 0 = "no religion", 1 = "never", 2 = "less than once a month", 3 = "less than once a week, but at least once a month", and 4 = "once a week or more". Response to the question to parents: "How often have you gone to religious services in the past year?"	2.8	1.2	0	4	2.7	1.3	0	4
Female	Dummy variable taking value one if the respondent is female.	0.6	0.5	0	1	0.4	0.5	0	1
Student Grade	Grade of student in the current year, ranging from 7 to 12.	9.6	1.6	7	12	9.5	1.6	7	12
Non-White	Dummy variable taking value one if race is not white.	0.3	0.5	0	1	0.4	0.5	0	1
Math Score	Grade in mathematics, ranging from 1 (D or lower) to 4 (A).	2.8	1.0	1	4	2.5	1.0	1	4
Parent Education	Schooling level of the (biological or non-biological) father (mother, in case father education is not available), coded as 9 - "some school", 11 - "no completed high school, or a business, trade, or vocational school instead of high school", 13 - "high school graduate", 15 - "completed GED; or business, trade, or vocational school after high school; or some college", 17 - "graduated from college or a university", 19 - "professional training beyond a four-year college".	14.7	2.7	9	19	14.3	2.6	9	19
Residential Building Quality	Interviewer's response to the question "How well kept is the building in which the respondent lives?", coded as 1 - "very poorly kept (needs major repairs)", 2 - "poorly kept (needs minor repairs)", 3 - "fairly well kept (needs cosmetic work)", 4 - "very well kept".	3.5	0.8	1	4	3.4	0.8	1	4
Network Size*	Number of individuals in the leader network.	391	363	2	860	328	361	2	860

** Includes observations inside the leader network only.*

Table C.3: Description of Data (Non-Leaders and Leaders, N=2,892)

		Non-Leaders 2,253 (78%)				**Leaders 639 (22%)**			
	Variable definition	mean	sd	min	max	mean	sd	min	max
Criminal Activity Index	Sum of the total criminal activity during the last 12 months.	1.3	1.8	0	10	8.8	5.6	1	34
Criminal	Dummy variable taking value one if the respondent is a criminal.	0.5	0.5	0	1	1.0	0.0	1	1
Distance to Leader	Geodesic distance (length of the shortest path) between the individual and the network leader (individual whose criminal activity exceeds the school median in 3 median average deviations or more).	2.6	1.5	1	7	0.0	0.0	0	0
Moral Cost	Parental religion practice, where 0 = "no religion", 1 = "never", 2 = "less than once a month", 3 = "less than once a week, but at least once a month", and 4 = "once a week or more". Response to the question to parents: "How often have you gone to religious services in the past year?"	2.8	1.2	0	4	2.7	1.3	0	4
Female	Dummy variable taking value one if the respondent is female.	0.5	0.5	0	1	0.3	0.5	0	1
Student Grade	Grade of student in the current year, ranging from 7 to 12.	9.7	1.5	7	12	9.2	1.5	7	12
Non-White	Dummy variable taking value one if race is not white.	0.4	0.5	0	1	0.4	0.5	0	1
Math Score	Grade in mathematics, ranging from 1 (D or lower) to 4 (A).	2.7	1.0	1	4	2.4	1.1	1	4
Parent Education	Schooling level of the (biological or non-biological) father (mother, in case father education is not available), coded as 9 - "some school", 11 - "no completed high school, or a business, trade, or vocational school instead of high school", 13 - "high school graduate", 15 - "completed GED; or business, trade, or vocational school after high school; or some college", 17 - "graduated from college or a university", 19 - "professional training beyond a four-year college".	14.4	2.6	9	19	14.2	2.6	9	19
Residential Building Quality	Interviewer's response to the question "How well kept is the building in which the respondent lives?", coded as 1 - "very poorly kept (needs major repairs)", 2 - "poorly kept (needs minor repairs)", 3 - "fairly well kept (needs cosmetic work)", 4 - "very well kept".	3.4	0.8	1	4	3.3	0.8	1	4
Network Size	Number of individuals in the leader network.	414	361	2	860	139	283	2	860

Table C.4: Description of Data by Distance to the Leader (Non-Leaders, N=2,253)

		Distance to the Leader (mean value)						
	Variable definition	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Criminal Activity Index	Sum of the total criminal activity during the last 12 months.	1.5	1.4	1.2	1.1	1.0	1.0	1.0
Criminal	Dummy variable taking value one if the respondent is a criminal.	0.5	0.5	0.5	0.4	0.5	0.4	0.5
Moral Cost	Parental religion practice, where 0 = "no religion", 1 = "never", 2 = "less than once a month", 3 = "less than once a week, but at least once a month", and 4 = "once a week or more". Response to the question to parents: "How often have you gone to religious services in the past year?"	2.6	2.8	2.9	3.0	3.0	3.0	3.1
Female	Dummy variable taking value one if the respondent is female.	0.5	0.5	0.5	0.5	0.5	0.5	0.4
Student Grade	Grade of student in the current year, ranging from 7 to 12.	9.4	9.6	9.9	9.6	10.1	10.3	10.4
Non-White	Dummy variable taking value one if race is not white.	0.3	0.3	0.4	0.5	0.6	0.6	0.7
Math Score	Grade in mathematics, ranging from 1 (D or lower) to 4 (A).	2.6	2.7	2.8	2.7	2.5	2.5	2.6
Parent Education	Schooling level of the (biological or non-biological) father (mother, in case father education is not available), coded as 9 - "some school", 11 - "no completed high school, or a business, trade, or vocational school instead of high school", 13 - "high school graduate", 15 - "completed GED; or business, trade, or vocational school after high school; or some college", 17 - "graduated from college or a university", 19 - "professional training beyond a four-year college".	14.4	14.5	14.4	14.2	14.2	14.4	14.1
Residential Building Quality	Interviewer's response to the question "How well kept is the building in which the respondent lives?", coded as 1 - "very poorly kept (needs major repairs)", 2 - "poorly kept (needs minor repairs)", 3 - "fairly well kept (needs cosmetic work)", 4 - "very well kept".	3.4	3.5	3.4	3.4	3.3	3.4	3.7
Network Size	Number of individuals in the leader network.	270	415	497	444	563	599	723

Table C.5: Description of Data (Inside and Outside a Leader Network, N=6,557)

		Inside a Leader Network 2,892 (44%)				**Outside a Leader Network 3,665 (56%)**			
	Variable definition	mean	sd	min	max	mean	sd	min	max
Criminal Activity Index	Sum of the total criminal activity during the last 12 months.	2.9	4.4	0	34	0.9	1.4	0	7
Criminal	Dummy variable taking value one if the respondent is a criminal.	0.6	0.5	0	1	0.4	0.5	0	1
Distance to Leader	Geodesic distance (length of the shortest path) between the individual and the network leader (individual whose criminal activity exceeds the school median in 3 median average deviations or more).	2.0	1.7	0	7	15.9	0.9	8	16
Moral Cost	Parental religion practice, where 0 = "no religion", 1 = "never", 2 = "less than once a month", 3 = "less than once a week, but at least once a month", and 4 = "once a week or more". Response to the question to parents: "How often have you gone to religious services in the past year?"	2.8	1.3	0	4	2.8	1.2	0	4
Female	Dummy variable taking value one if the respondent is female.	0.5	0.5	0	1	0.6	0.5	0	1
Student Grade	Grade of student in the current year, ranging from 7 to 12.	9.6	1.5	7	12	9.5	1.6	7	12
Non-White	Dummy variable taking value one if race is not white.	0.4	0.5	0	1	0.3	0.5	0	1
Math Score	Grade in mathematics, ranging from 1 (D or lower) to 4 (A).	2.6	1.0	1	4	2.8	1.0	1	4
Parent Education	Schooling level of the (biological or non-biological) father (mother, in case father education is not available), coded as 9 - "some school", 11 - "no completed high school, or a business, trade, or vocational school instead of high school", 13 - "high school graduate", 15 - "completed GED; or business, trade, or vocational school after high school; or some college", 17 - "graduated from college or a university", 19 - "professional training beyond a four-year college".	14.4	2.6	9	19	14.6	2.7	9	19
Residential Building Quality	Interviewer's response to the question "How well kept is the building in which the respondent lives?", coded as 1 - "very poorly kept (needs major repairs)", 2 - "poorly kept (needs minor repairs)", 3 - "fairly well kept (needs cosmetic work)", 4 - "very well kept".	3.4	0.8	1	4	3.5	0.8	1	4
Network Size	Number of individuals in the leader network.	353	363	2	860	-	-	-	-