The Value of Connections: Evidence from the Italian-American Mafia
`The Value of Criminal Connections` *

Giovanni Mastrobuoni

Submitted , Accepted July 2014†

Abstract

This study uses declassified data on US Mafia members of the 1950s and 1960s to estimate the criminal network effect on their economic status. I measure economic status exploiting detailed information about their place of residence.

Housing values are reconstructed using current deflated transactions data. I deal with non-random sampling of mobsters modelling investigations on connections as Markov chains. Reverse causality between economic status and the gangster's position in the network is solved exploiting exogenous exposure to pre-immigration connections. A standard deviation increase in *closeness* centrality increases economic status by between one-forth and three-fourth of a standard deviation.

JEL classification codes: A14, C21, D23, D85, K42, Z13

In January 2011, exactly 50 years after Robert F. Kennedy's first concentrated attack on the American Mafia as the newly appointed attorney general of the United States, nearly 125 people were arrested on federal charges, leading to what federal officials called the "largest mob roundup in FBI history."[1]

---

*Giovanni Mastrobuoni, Department of Economics, Collegio Carlo Alberto, IZA, Netspar; Wivenhoe Park, CO43SQ Colchester, UK. Email: gmastrob@essex.ac.uk. I would like to thank the editor Steven Machin, as well as an anonymous referee, Nadia Campaniello, Edoardo Gallo, Giacomo de Giorgi, Gaia Narciso, Michele Pellizzari, Santiago Oliveros, David Reinstein, Rodrigo Soares, Fabian Waldinger, and seminar participants at Trinity College Dublin, the University of Warwick, at the University of Essex, and at the Brucchi Luchino conference for their comments. Martino Bernardi and Dominic Smith have provided excellent research assistance.

[1]See The New York Times, January 21, 2011, page A21 of the New York edition.

Over the last 50 years the Mafia has continued following the same rules, and is still active in many countries, including the United States.[2] Despite this, the illicit nature of organized crime activities has precluded empirical analysis and the literature has overwhelmingly been anecdotal or theoretical (Reuter, 1994, Williams, 2001).

This study uses declassified data on 800 Mafia members, who were active just before the 1961 crackdown, to study the importance of criminal connections inside such a secret society (linking the network position of mobsters to an economic measure of their success). The records are based on an exact facsimile of the Federal Bureau of Narcotics (FBN) secret files on American Mafia members in 1960 (MAF, 2007).[3]

The data contain information collected from FBN agents on the gangsters' closest criminal associates, which I use to reconstruct the criminal network.[4] Connections are believed to be the building blocks of secret societies and of organized crime groups, including the Mafia. Francisco Costiglia, alias Frank Costello, a Mafia boss who according to the data was connected to 34 gangsters, would say "he is connected" to describe someone's affiliation to the Mafia (Wolf and DiMona, 1974).

Indeed, the first rule in Mafia's decalogue states that "No one can present himself directly to another of our friends. There must be a third person to do it," who knows both affiliates (Maas, 1968).[5] As a consequence, gangsters who are on average closer to all the other gangsters need fewer interconnecting associates to expand their network. While gangsters who bridge connections across separate clusters of the network can maintain

[2]According to the FBI, in 2005 there were 651 pending investigations related to the Italian-American mafia; almost 1,500 mobsters were arrested, and 824 were convicted; of the roughly 1,000 "made" members of Italian organized crime groups estimated to be active in the US, 200 were in jail (see www.fbi.gov.

[3]The distribution of the year of first arrest of mobsters has almost full support within the range 1908-1960, so one can infer that the data refer to what the authorities knew in 1960.

[4]In the 1930s and up to the 1950s the FBN, which later merged with the Bureau of Drug Abuse Control to form the Bureau of Narcotics and Dangerous Drugs, was the main authority in the fight against the Mafia (Critchley, 2009). For example, in New York the Federal Bureau of Investigation had just four agents, mainly working in office, assigned to the mafia, while in the same office more than 400 agents were fighting domestic communists (Maas, 1968).

[5]These rules were listed both, in a 1963 testimony by the first FBN informant, Joe Valachi, and in a piece of paper that belonged to the Italian Mafia boss Salvatore Lo Piccolo during his 2007 arrest.

such monopoly power.

There are measures of importance of members inside networks that are based on the average closeness and on the bridging capacity, called *closeness* centrality and *betweenness* centrality.[6] But even just the number of connections, known as *degree* centrality, might be important to reach leadership positions, as in the Mafia these are not simply inherited. Soldiers elect their bosses using secret ballots (Falcone and Padovani, 1991, pg. 101).[7].

Three main empirical challenges emerge when estimating how a gangster's network centrality influences his economic prospects, called the network effect: i) the measurement of economic prospects in the absence of information about illegal proceeds, ii) the non-random and iii) endogenous nature of the network.

Regarding the first issue, since illicit transactions and criminal proceeds inside the Mafia are unobservable, I use the value of the house or the apartment where such criminals presumably resided (or nearby housing) to measure their economic success. Such value is reconstructed based on the deflated value of the current selling price of their housing based on the internet site *Zillow.com.* Prices are deflated using the Metropolitan Statistical Areas' (MSAs) average housing values from Gyourko *et al.* (2013). Given that most mobsters who were active in 1960 were born from very poor families (see Lupo, 2009), the value of the house where they resided, whether it was owned or rented, is arguably a reasonable measure of their illegal proceeds,[8] though reconstructing the original value is certainly prone to error.[9]

[6]The *closeness* centrality index measures the average distance between a node (a member) and all the other nodes, and its inverse is a good measure for how isolated members are. The *betweenness* centrality index measures the number of times a node is on the shortest path between two randomly chosen nodes. Ductor *et al.* (forthcoming) use both to predict research output of individual researchers. Kinnan and Townsend (2012) look at how network distance to a bank affects consumption smoothing.

[7]Degree is often used to measure network importance. For example, Kremer and Miguel (2007) use it study the diffusion of a deworming pill take-up, while Hochberg *et al.* (2007) show that venture capitalists' investment performance depends on how connected they are.

[8]A large literature has shown the link between housing demand and income (see Goodman, 1988). The appendix Figure 11 shows a correlation of 60 percent between Zip code-level median housing prices and median household income, with an elasticity that is very close to 1.

[9]Any classical measurement error would inflate the standard errors, making the inference more conservative.

Regarding the second issue, in the 1960s the total estimated number of mafia members was around 5,000 (Maas, 1968). Since almost all high-ranking members have a record, the 800 criminal profiles are clearly a potentially nonrepresentative sample of Mafia members. To deal with the incompleteness and non-randomness of the network, in Section 1 I model law enforcement's surveillance and detection of Mafia network nodes (mobsters) as a Markov chain (see also Mastrobuoni and Patacchini, 2012).

The final issue about the potential endogeneity of the network requires a longer discussion. Sparrow (1991) and Coles (2001) propose the use of network analysis to study criminal networks, however, apart from some event studies based on a handful of connections, empirical evidence on criminal networks is scarce and never addresses the potential endogeneity of the network.[10]

In non-experimental settings the variation that identifies the effect of networks may be partly driven by homophily (the tendency of individuals to be linked to others with similar characteristics), or unobserved characteristics which determine someone's position in the network as well as his or her outcomes. Mobsters might, for example, use their (unobserved) wealth to build connections and buy more expensive housing.

Since real networks can hardly be generated entirely through an intervention, there are three ways to estimate (causal) network effects:[11] i) modelling sequentially the network formation and the network effects (see Chandrasekhar and Lewis, 2011), ii) experimenting

[10]Morselli (2003) analyses connections within a single New York based family (the Gambino family), Krebs (2002) analyses connections among the September 2001 hijackers' terrorist cells, Natarajan (2000, 2006) analyzes wiretap conversations among drug dealers, and McGloin (2005) analyzes the connections among gang members in Newark (NJ). There is considerable more theoretical work. Most studies have focused on a market structure view of organized crime, where the Mafia generates monopoly power in legal (for a fee) and illegal markets. Among others, such a view is present in the collection of papers in Fiorentini and Peltzman (1997), and in Reuter (1983), Abadinsky (1990), Gambetta (1996), and Kumar and Skaperdas (2009). Only two theoretical papers have focused on the internal organisation of organized crime groups. Garoupa (2007) looks at the optimal size of these organisations, while Baccara and Bar-Isaac (2008) look at the optimal internal structure (cells versus hierarchies).

[11]Alternatively, one can avoid making any causal claims. Ductor *et al.* (forthcoming) focus on predictions, and show that researchers' network centralities help to predict future research output.

with networks,[12] and iii) instrumenting the position in the network.[13] [14]

Since connections, their number, as well as their quality, are potentially even more important in a world without enforceable contracts, a world where secrecy, reputation, and violence prevail, such bonds are even more likely to be endogenous.

Several factors might influence the decision to connect and do business with another gangster. When gangsters expand their network they are trading off the increased risk of whistleblowing with increased criminal proceeds (Bonanno, 1983). Criminal hierarchies, kinships, complementarity and substitutability in criminal as well as non-criminal activities are just some of the factors that are likely to influence the gangster's decision to expand his network, and thus his network centrality.

Instead of modelling the entire network formation mechanism, I rely on an instrumental variable that influences mobsters' centrality in the Mafia network. Such instrument is based on information collected from the pre-immigration communities (as in Munshi, 2003), that the mobsters left several decades before I observe their network and their housing wealth in 1960. The identification strategy is based on advantages in building connections that originate from the gangsters' or from his ancestors' place of origin

[12]Experimental studies on networks usually take the network as given and randomly assign information or other treatments to single network nodes (Alatas *et al.*, 2012, Fafchamps *et al.*, 2013, Kremer and Miguel, 2007). While such experimental variation does not solve the endogeneity issue, Banerjee *et al.* (2013) develop and test a model of information diffusion, validating such model based on unexploited variation in the data (see also Blume *et al.*, 2012).

[13]For example, Munshi (2003) uses rainfall in the origin-community as an instrument for the size of the network at the destination (the United States). Mexican immigrants with larger networks in the US face better labour conditions.

[14]The more commonly studied peer effects can be seen as specific network effects, where only direct links are assumed to matter. Researchers have used this restriction to estimate how criminals' behaviour depends on the behaviour of their peers (see Baker and Faulkner, 1993, Bayer *et al.*, 2009, Drago and Galbiati, 2012, Haynie, 2001, Patacchini and Zenou, 2008, Sarnecki, 1990, 2001, Sirakaya, 2006). An old and extensive literature in labour economics documents the importance of friends and relatives in providing job referrals (Bayer *et al.*, 2008, Glaeser *et al.*, 1996, Montgomery, 1991). In recent years the interest has shifted toward understanding not just peer influence, or the influence of direct links, but how the whole architecture of a network, thus including indirect links, influences behaviour and outcomes (Ballester *et al.*, 2006, Goyal, 2007, Jackson, 2008, Vega-Redondo, 2007). Empirical evidence on these "network effects" is scarce but growing (see, among others Alatas *et al.*, 2012, Angelucci *et al.*, 2010, Banerjee *et al.*, 2013, Hochberg *et al.*, 2007, Kinnan and Townsend, 2012), with the main burden being the endogeneity of the network (see Blume *et al.*, 2012).

(Italy). Conditional on the region of birth, a detailed description of their legal and illegal activities, and other individual characteristics of the mobsters these innate connections should not influence mobsters' housing wealth in the United States (other than through such connections).[15] While this exclusion restriction remains untestable, I address potential pitfalls; mainly endogenous migration driven by successful mobsters, and direct effects of potential innate connections on housing values.

Several factors might have helped building more connections when the gangster's ancestors were concentrated in more traditional Mafia territories: i) increased trust that spills over from known and reputable families,[16] ii) easier punishments through left-behind kinships, and iii) knowledge about the rules and traditions of the secret society (e.g. "*omertà*," which is a vow of silence).

Absent detailed information on the communities of origin for all the members born in the US, I use the informative content of surnames on the place of origin, more commonly known as *isonomy*. Such measure of potential innate interactions predicts the gangster's individual number and quality of connections. Exploiting differences in isonomy between Southern and Northern Italy, I perform falsification tests of the first stage regressions which validate the instrument's capability to proxy for the place of origin. The instrument is weak, though I show that restricting the analysis to ever arrested gangsters (making up 80 percent of the sample), for which surnames are arguably measured with greater precision strengthens the instrument without changing the results.

When using the instrument a one standard deviation increase in network *closeness* centrality (the inverse of the average network distance from all other gangsters) increases housing value by three-forth of a standard deviation, with the p-values on the endogeneity

[15]See Section 2.4 for a thorough discussion about the instrument. Such instrument is also related to the growing literature on trust and family values. Guiso *et al.* (2006) present an introduction to the importance of culture, defined as "customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation," on economic behaviour. The same applies to criminal behaviour.

[16]Karlan *et al.* (2009) build and estimate a model of trust in social networks between a lender and a borrower based on the collateral that is provided by their weakest links.

tests being usually close to 10 percent. The results are similar for *eigenvector* centrality (which is a function of the prestige of the connected members), while the results for *degree* (the simple count of connections) and *betweenness* centrality (the bridging capacity across different clusters of the network) tend to be weaker, but for different reasons. On one hand, *degree* appears to be a crude measure of someone's importance (the value of connections is increasing in the rank of the gangster). On the other hand, gangsters with high *betweenness* were more likely to be part of the *Commissione*, the governing body of the Mafia. These members of the Mafia often kept a lower profile by living in more humble housing, which unresolvably biases the corresponding results downward.

1. Sampling and Estimating Network effects

A quick look at record number one, Joe Bonanno (see the Online appendix Figure 10), reveals the kind of information that will be used to link a mobster's network centrality to his economic success. According to the FBN he was born on January 18, 1905 in Castellamare (Sicily), and resided in 1847 East Elm Street in Tucson (Arizona). He had interests in three legal businesses: *Grande Cheese Co.*, *Fond du Lac* (Wisconsin), *Alliance Realty & Insurance* (Tucson, Arizona), and *Brunswick Laundry Service* (Brooklyn, New York), etc.. Finally, his closest criminal associates were Lucky Luciano, Francisco Costiglia (Frank Costello), Giuseppe Profaci, Anthony Corallo, Thomas Lucchese, and Carmine Galante.

I use i) the value of the house where mobsters reside to measure economic success y (Section 2.1), ii) information on their associates to reconstruct the network $\mathbf{G}$ (Section 2.2),[17] iii) the informational content of surnames to build the instrumental variable (Section 2.4).

[17] I construct the undirected network, meaning that i and j with at least one of the two records lists the other surname among the associates.

But before analysing networks, particularly when such networks are hidden, it is important to take into account that the observed network $\overline{\mathbf{G}}$ represents a subset (subgraph) of the entire network $\mathbf{G}$, and not necessarily a random one.

More formally, the goal is to model an economic outcome y as a linear function of network centrality c (abstracting for simplicity from other covariates),

$$y = \alpha + c(\mathbf{G})\beta + \epsilon, \tag{1}$$

when the observed network $\overline{\mathbf{G}}$ is a subset of $\mathbf{G}$. In general the measurement error that biases the estimate $\widehat{\beta}$ will not be classical. Chandrasekhar and Lewis (2011) show that when c measures *degree* centrality, the network is exogenous, and the sampling is random (with sampling rate ψ), $plim\widehat{\beta} = \beta \cdot \psi^{-1} \cdot$ attenuation bias. The reason for the scaling factor ψ^{-1} is that under random sampling the expectation of $c(\mathbf{G})$ is approximately equal to $\psi c(\mathbf{G})$.

However, the observed Mafia records are unlikely to represent a random sample of mobsters. The 800 criminal files come from an exact facsimile of a Federal Bureau of Narcotics report of which fifty copies were circulated within the Bureau starting in the 1950s. They come from more than 20 years of investigations, and several successful infiltrations by undercover agents (McWilliams, 1990). The FBN data represent a snapshot of what the authorities knew in 1960, and thus do not contain exact information about the hierarchies within the organisation. Such information was revealed only a few years later, when Joe Valachi, a Mafia associate, became the first FBN and later FBI informant.[18]

Joe Valachi's testimony confirmed FBN's view that the Mafia had a pyramidal structure with connections leading toward every single member.[19] Indeed, the observed net-

[18] Jacobs and Gouldin (1999) provide a relatively short overview about law enforcement's unprecedented attack on Italian organized crime families following Valachi's hearings.

[19] Valachi revealed that the *Cosa Nostra* was made of approximately 25 Families. *Cosa Nostra* was governed by a *Commissione* of 7-12 bosses, which also acted as the final arbiter on disputes between Families. The remaining 10 to 15 families were smaller and not part of *Cosa Nostra*'s governing body.

work is connected (or ergodic), meaning that from each node (gangster) one can reach any other node. Moreover, it is a "small-world" network, as the average path length between gangsters is just 3.7 steps.

Given the hierarchical structure of the Mafia and the estimated 5,000 associates who were active during those years, the 800 gangsters are likely to be a non-random sample of *Cosa Nostra* members. More active, more important, and more connected mobsters were certainly more likely to be noticed and tracked. Indeed, all known Mafia bosses who were alive in 1960 are listed in the records.

This means that the observed 800 gangsters are likely to be more connected than the average one, and that part of the network is unobserved. There are no written records about how the FBN followed mobsters and constructed the network. Through surveillance posts and undercover agents, the agency was likely discovering previously unknown mobsters following known ones. Two surveillance photographs of Italian mobsters taken in 1980 and in 1988 show evidence of these patterns (see the Online Appendix Figure 9). As a consequence more connected gangsters were more likely to be "sampled" by the FBN. In order to produce a representative sample of mobsters, "sampling" weights should underweigh highly connected mobsters and vice versa.

This kind of sampling resembles a procedure that is used to sample hidden populations, called *snowball sampling* (see Frank, 1979, Goodman, 1961, Granovetter, 1976, Rothenberg, 1995, Snijders, 1992). Let us assume the FBN starts observing one or more mobsters out of N, and that such known mobsters are indicated with the number one in the $1 \times N$ vector of zeros and ones $\mathbf{p_0}$, called *the seed*. Following the initial mobsters' links the FBN will observe more and more mobsters. Starting with the $N \times N$ symmetric matrix $\mathbf{A} = [a_{ij}]$ with elements equal to 1 when mobsters i and j are connected and zero otherwise (called the *adjacency matrix*), one can obtain the transition matrix $\mathbf{T}$

Each Family was structured in hierarchies with a boss (*Capo Famiglia*) at the top, a second in command, called underboss (*Sottocapo*), a counselor (*Consigliere*) and several captains (*Caporegime*) who head a group of soldiers (*regime*) (Maas, 1968).

normalising the columns to sum up to one. The element t_{ij} of $\mathbf{T}$ measures the likelihood of discovering mobster j when mobster i is under surveillance.

After k steps the likelihood of discovering mobster j is equal to the j-th of the vector $\mathbf{p_k} = \mathbf{p_0}\mathbf{T}^k$. The corresponding stationary distribution $\mathbf{p}$, defined as a vector that does not change under application of the transition matrix, is independent of the seed, and is equal to $\mathbf{p} = \mathbf{p}\mathbf{T}$.[20]

Element p_i of the probability vector $\mathbf{p}$ can be interpreted as the likelihood of observing gangster i if one randomly picked and followed an edge of the network. The resampling weights are thus equal to the inverse of such probability $w_i = \frac{1}{p_i}$, with $0 < p_i < 1$. Since p_i is almost proportional to the number of connections, such weights are quite intuitive. Gangsters with fewer connections, who are less likely to be spotted by the FBN and as a result are under-represented, receive larger weights. The weighting corrects for the selection bias when describing the individual characteristics of the mobsters.[21]

Moreover, using a Monte Carlo simulation, Lee *et al.* (2006) show that under snowball sampling the scale bias is less than ψ^{-1} as for the most central nodes $c(\mathbf{G})$ and $c(\overline{\mathbf{G}})$ coincide.[22] Since exact re-scaling would only be applicable to *degree*, later I use logarithmic transformations of c to address the scaling bias ψ^{-1} and an instrumental variable approach to address the attenuation bias (as well as to address the potential endogeneity of the network).

Expressing the variables of equation 1 in logs has several advantages. First, we will see that the densities of $log(y)$ and $log(c)$ are approximately normal. Second, it deals

[20]The Perron-Frobenius theorem ensures that when the Markov chain is ergodic such a vector exists and is unique. I approximate $\mathbf{p}$ with $\mathbf{p_{40}}$, and compute such distribution multiplying a constant vector of size N (number of nodes) that sums up to one by the 40th power of $\mathbf{T}$. This result is related to the convergence in snowball sampling, called respondent-driven-sampling (Heckathorn, 1997).

[21]See Golub and Jackson (2010) for a discussion about selection bias in networks.

[22]Such bias would be even less severe under respondent-driven sampling.

with the scaling bias. The (weighted) regression model in *logs* becomes:

$$\log(y) = \widetilde{\alpha} + \log(c(\mathbf{G}))\widetilde{\beta} + \widetilde{\epsilon}, \tag{2}$$

where $\widetilde{\alpha}$ includes $-\log(\psi)$.

The summary statistics (Table 1) describe the individual characteristics as well as the network characteristics of the gangsters, with and without correcting for the non-random sampling design. The average number of identical surnames (possibly "extended family members") are about 1.5. Most gangsters migrated between the turn of the century and the 1930s (see the Online Section A), though the records contain no information on when the gangsters, or their families, migrated to the US. About 70 percent of mobsters who were active in 1960 were born in the US, while the rest was split between Sicily (about 20 percent) and the rest of Italy (about 10 percent). In 1960 the average age of the associates is about 48, and 76 percent of them are married (10 percent are divorced). About 15 percent of the times the wife's maiden name is shared with other gangsters, leading to a presumably "connected wife". The average number of known children is about 1, while the average number of known siblings is 2. The FBN records contain also some information about the mobsters' legal and illegal businesses. The average number of businesses is about 1. A little more than half of them are involved in drug dealing, but in line with long criminal careers (the average age at first arrest is 23) the average number of crime types listed in their records is close to 3. Only 14 percent of gangsters have no arrest record. For these gangsters some variables (e.g. the place of residence, or the exact surname) might be measured with more noise.

Weighting introduces larger changes in average centrality. For instance, the average *degree* drops from about 11 to about 6 when weighting. The remaining centrality indices are divided by their maximum value (multiplied by 100). The housing characteristics and the instrumental variable "potential innate interactions" are discussed later in Section

2.1 and 2.4.

2. Descriptive Evidence

2.1 Housing values and Number of Legal Businesses

There is no database on individual housing values of 1960 properties,[23] but feeding the exact residence address into *Zillow.com* produces 641 current real estate values, and for 561 homes (about 90 percent of the sample) there is also information on the year the house was built.[24] The remaining 159 mobsters were not residing in the US anymore (like Lucky Luciano, who had already been expelled from the country), or never lived in the US (most of these mobsters were living in Italy).[25]

Zillow's estimated price is calculated from public data and from data submitted by users (e.g. real estate agents or appraisers physically inspects the home and take special features, location, and market conditions into account). Comparing historic estimates based on Zillow with the actual transaction prices of homes that sold, about 80 to 90 percent of the time the measurement error is within 20 percent of the sale price (more than half of the time it is within 10 percent).[26]

In order to deflate contemporary housing values to 1960 ones, I use the average housing value in 1960 and in 2000 taken from Gyourko *et al.* (2013) for the 608 homes that are in a MSA, while I use State level Census data for the remaining 33 homes.[27]

[23] The only historical data on individual housing units that goes so far back in time are rental prices and sale prices from newspaper articles (see, for example, Margo (1996)), but it would be close to impossible to find those exact same housing units occupied by the mobsters.

[24] One third of the times the exact address did not produce an estimated value, and the nearest house with such information was selected. Since housing values tend to be highly geographically clustered such proxy is likely to reduces the precision of our estimates by a small amount.

[25] While these mobsters do not contribute directly to the analysis, they are part of the network and are used to construct measures of network centrality.

[26] See http://www.zillow.com/howto/DataCoverageZestimateAccuracy.htm

[27] See http://www.census.gov/hhes/www/housing/census/historic/values.html.

The left Panel of Figure 1 shows the relationship (truncated at the 90th percentile) between the current and the 1960 values. Housing prices have approximately doubled over the last 50 years, though they have increased almost 5 times in San Francisco, while they stayed almost constant in Binghampton, Utica/Rome, or Buffalo. In the New York MSA, where almost 300 gangsters reside, prices doubled.

The 5th, 10th, 25th, 50th, 75th, and 90th percentile of the housing value in 1960 were 39, 50, 95, 190, 325, and 662 thousand dollars. The 95th and 99th percentiles were 1.5 and 4.7 million dollars. The right Panel of Figure 1 shows the housing value density (truncated at the 90th percentile). The mean housing value in 1960 is $400,000 when not weighting the sample, and is smaller ($379,000) when weighting.[28]

As long as the remaining within-city variation is stable between 1960 and today, the deflated housing prices represent a good approximation of the gangsters' economic status. Unfortunately there is little empirical literature about the persistence in housing prices within cities over decades. A recent study (Villarreal, 2013) shows that natural drainage conditions influenced the desirability of the local environment during the settlement of New York City in the nineteenth century, and that, despite private residential redevelopment, those historical conditions explain contemporary variation in housing prices and household incomes.[29] *Zillow.com* publishes more than 10,000 zip code-level median housing prices going back 16 years. Using such data to compute the correlation between current (T) and lagged ($T-t$) prices, one can show that as t goes to 16 years the correlation declines from 100 to 92 percent.[30] Moreover, deflating prices using city-level inflation rates the decline drops from 8 to just 2 percentage points. That said, the correlation might decrease (and the measurement error increase) when moving from median prices to prices of individual housing units, and when reducing t for the remaining 38

[28]Table 1 shows that 10 percent of the houses found on *Zillow.com* were built after 1960. To control for the fact that these houses might have a different valuation I am going to control for a dummy variable equal to one when the houses were built after 1960.

[29]Bleakley and Lin (2012) show persistence across cities driven by now obsolete portage locations.

[30]See the Online appendix Figure 12.

years.

The construction of housing value is also subject to additional problems. Wealthy mobsters might own several houses and their place of residence might not perfectly represent their wealth whenever it is in the hands of figureheads.[31] All these measurement problems render the results more conservative.

Later in the network effects regressions I also control for the legitimate earnings opportunities, measured by the number of legal business that gangsters own.

2.2 Network-based Measures of Importance

Each criminal record contains a list of criminal associates. Figure 10 indicates, for example, that Joe Bonanno was associated with Luciano, Costello, Profaci, Corallo, Lucchese, and Galante. There is no evidence about how the FBN established such associations, but each record tends to list the most important (and connected) associates.

Indirected connections are clearly more numerous, as mobsters can be listed as associates in several records. As in Mastrobuoni and Patacchini (2012), I define two mobsters to be connected whenever the FBN lists one of the two mobsters as an associate in the other mobster's file. *Outdegree* (the number of associates listed in one file) is bounded by the available space on a record (the maximum is 13), while *indegree* (the number of times someone is listed in other records) is not. For this reason *degree* (the number of undirected connections) is only weakly correlated with *outdegree* (37 percent), and mainly depends on *indegree*.[32]

The number of connections is clearly the simplest but crudest way to measure the importance of members. In recent years social network theorists proposed different centrality measures to account for the importance of someone's connections (Borgatti, 2003,

[31] Joe Bonnano is a good example. While being a NYC boss his official residence was for strategic reasons in Tucson, Arizona.

[32] In other words, I construct a symmetric adjacency matrix of indirected connections between mobsters' last names. Dealing with changing first names would have been a complex task.

Wasserman and Faust, 1994).[33]

Unlike *degree*, which weights every contact equally, the *eigenvector* index weighs contacts according to their centralities.[34] The index takes the whole network into account (direct and indirect connections).[35]

Given the first Mafia rule that guarantees secrecy, gangsters who are closer to other gangsters need fewer interlinking associates to reach a randomly chosen gangster, while gangsters with important bridging capacities across clusters of the network have more monopoly power in establishing such new links. While *degree* centrality might in principle be inferior to the other centrality measures, I will estimate network effects based on all four measures of centrality.

All these measures are positively skewed (Figure 2), especially *betweenness* centrality, indicating that a few mobsters represent the bridges between subsets of the network. Given the skewness and the scaling bias that I discussed prior to Equation 2 in order to estimate the network effects all measures of network centrality are taken in logs (Ductor *et al.*, forthcoming, also use a specification in logs).[36] The corresponding (more symmetric) densities are plotted in the Online appendix Figure 13. Another Online appendix Figure shows that all centrality measures are positively related to each other (Figure 14). Plotting log *eigenvector* against log *closeness* generates a thick line ($\rho = 96$ percent), which shows that once one penalises the larger outliers the two centrality measures are quite similar.

But such large correlation masks a very different variability. The ratio between the

[33]See also Sparrow (1991) for a discussion on centrality indices in criminal networks.

[34]It equals the *eigenvector* of the largest positive eigenvalue of the adjacency matrix, the $N \times N$ matrix of zeros and ones (indicating whether gangster i and j are connected).

[35]As first noted by Granovetter (1973), weak ties (i.e. friends of friends) are important source of information. While Ballester *et al.* (2006, 2010) show that in non-cooperative games the activity of individuals is proportional to the *eigenvector* centrality (the "key-player" having the largest *eigenvector* centrality), several assumptions of that model would not hold for the Mafia: conditional on operating inside the same area and being part of the same Mafia "Family," the hierarchy (introducing cooperativeness) as well the implied risk such connections are likely to influence the activity of mobsters.

[36]For the *betweenness* centrality index, since 4.5 percent of observations have such index equal to zero, I take the inverse hyperbolic sine transformation $log(y + (y^2 + 1)^{1/2})$.

standard deviation of the log *eigenvector* index and the log *closeness* index is about 10 to 1. This has to be taken into account when interpreting standardised variations.

The correlations are lower with respect to the other two measures, especially in the lower tail. For *betweenness* the low correlation is driven by the fact that several mobsters, despite having many connections, have extremely low levels of *betweenness*. Hierarchies are the likely reason. For about 400 mobsters I managed to reconstruct their position within the mafia (not always referred to their status in 1960). Underbosses and captains who head several soldiers within one Mafia "Family," tend to have large *degree*s but low *betweenness*. Counsellors and bosses have the largest median *betweenness* measure (about 1/2), while those of captains and underbosses are half that large.[37] Differences in *closeness* and *eigenvector* centrality across ranks are less striking. With respect to *degree* the low correlation with other indices is driven by small values. At larger *degrees* most indices tend to be highly correlated with each other.

The next section describes the relationship between housing prices and network centrality.

2.3 Economic Status, Network Centrality, and Potential Biases

Figure 3 shows that the unconditional weighted local polynomial regression of log-housing value is increasing in all log-centrality measures, and is not far from being linear. The correlation is stronger when using the *eigenvector* index and the *closeness* index, than when using the simple *degree* or the *betweenness* index (approximately, 20 percent versus 10 percent).

On the one hand, *degree* is likely to be a poor measure of centrality when the connections are scarce but valuable. On the other hand, mobsters with larger *betweenness*

[37]Soldiers have a median *betweenness* index of about 0.18.

indices, represent bridges between clusters of the network, most likely different Mafia Families. Since the detection of a such high ranking bosses would have put the whole organisation in peril (see Baccara and Bar-Isaac, 2008), there is anecdotal evidence that some bosses kept a low profile. For example, Joseph Bonanno tells the story about when he decided not to join Lucky Luciano's very lucrative garment industry in New York to avoid being in the spotlights (Bonanno, 1983). Such members' true economic outcomes might also be hidden whenever their wealth is in the hands of figureheads.

There is indeed evidence that some Mafia leaders preferred to "officially" reside in unpretentious housing. Using the FBN's description of the associates, one can show that those described as "leaders" or "bosses" tend to be more central in the network (Table 2). They tend to have considerably larger *betweenness* centrality than lower-ranked gangster, about 40 percent larger, while other centrality measures differ less between bosses and lower-ranked gangsters. Despite this, such leaders tend to live in less expensive housing.[38]

Moreover, the Sicilian origin seems to influence the decision to keep a lower profile.[39] Figure 4 shows that despite the fact that Sicilians are usually more centrally located in the Mafia network, they tend to live in considerably cheaper housing; especially at the top of the distribution of housing values. Later in Section 3 I will test the robustness of the network effect when controlling for the place of origin.

Other potential omitted variables when regressing housing values on network centrality are the family composition and the gangsters' initial wealth. Larger families are instrumental to the bosses' success, but need also larger and more expensive housing. Wealth might be used to buy both, power inside the mafia and more expensive housing. While the data have no good proxy for initial wealth, in the next section I devise a

[38]Bosses also show an aversion to drug dealing, a very lucrative business. Later we will see that gangsters involved in drug dealing live is houses that are about 30 percent more expensive.

[39]Recently arrested bosses who were heading the entire Sicilian mafia, Totò Riina and Bernardo Provenzano, were living in very poor houses. Such cautious behaviour seems less present in other organized crime groups. A recently arrested boss from the Neapolitan Camorra, Francesco Schiavone, was living in a mansion, built after the house in the Hollywood movie "Scarface." This same pattern between the Italian region of birth and housing values is evident in the data.

presumably exogenous instrument that is based on the innate potential connection back in Italy. In particular, I exploit the joint spatial distribution of the gangsters' surnames in Italy to construct a measure of inherited (potential) connections, which is moderately correlated with network centrality.

2.4 Birthplace and Potential Innate Interactions

Several authors have highlighted the importance of familial, interpersonal, and communal relationships in determining criminals' success inside organized crime groups (see, among others, Coles, 2001, Falcone and Padovani, 1991, Ianni and Reuss-Ianni, 1972). Most of these relationships are also likely to influence housing decisions and would lead to implausible exclusion restrictions. For example, marrying a gangster's daughter is likely to boost someone's power inside the mafia, but might also change someone's housing budget directly.

Ideally, one would use mobsters' innate characteristics which influence his future chances to build connections, but are unrelated to his housing choice (other than through the derived centrality in the network). Proximity to other mobsters represents a natural choice. But such proximity should not be related to inherited wealth, as wealth might be used to acquire centrality. Moreover, geographic proximity based on the place of US residence is likely to be endogenous with respect to network centrality, as more powerful mobsters might decide to live in the middle of their sphere of influence.

For these reasons I use a measure of proximity that is based on Italian and not US residencies. Mafia associates born in Italy,[40] but also those born in the US from Italian immigrants often kept strong links with the Italian communities of origin.[41] At the be-

[40] About a quarter of mobsters were born in Sicily, 2/3 were born outside of Italy (mostly in the US) and the rest in other regions of Italy. Properly weighting the data, these fractions are 2/10, 7/10, and 1/10, indicating that Sicilians tend to have more connections.

[41] Table 1 shows that the average age is 48 years, which means that the average year of birth is 1912, right in the middle of the Italian migration wave (see the Online Section A). Most mobsters are either

ginning of this study, I already listed a number of reasons why such innate connections might influence the gangsters' connections back in the US (reputation, trust, punishability, and knowledge of the rules). The main untestable identification assumption is that such interactions at the origin do not shape the gangsters' housing preferences, at least not conditional on other covariates (including the region of birth).

The records do not contain pre-immigration information on the exact place of origin in Italy, but one can approximate such information using the informative content of surnames. For at least 30 years researchers in human biology have been exploiting the analogy between patrilineal surname transmission and the characterisation of families and communities (Lasker, 1977). For geographic, historical, as well as social reasons, surnames tend to be highly geographically clustered, particularly in countries with low internal mobility like Italy (see Allesina, 2011, Barrai *et al.*, 1999, Zei *et al.*, 1993).[42]. The geographic distribution of surnames, called *isonomy*, contains a strong signal about someone's origin.[43] The surname of record number 1, "Bonanno", is more widespread across the whole country, though, again, most Bonanno families live in Sicily, and a non-negligible fraction lives in Castellamare del Golfo, which is where Bonanno's family was coming from.[44]

Given that i) 30 percent of gangsters were born in Italy (including Sicily) and later moved to the US and even those who were born in the US were likely to keep links with Italy, and ii) surnames tend to be geographically clustered, the way the current distribution of a given surname overlaps with the distribution of all the other surnames represents one way, possibly the only way, to measure the connections stemming from

first or second generation immigrants. All but a handful of mobsters were of Italian origin, as this was a prerequisite to become a member. The few non-Italian gangsters in the data were either French gangsters from Marseille or Corsica, or part of the, so-called, Jewish Mafia.

[42]See Colantonio *et al.* (2003) for an overview on recent developments on the use of surnames in human population biology.

[43]One can try out surnames of Italian economists on the following Web sites: http://www.gens.info/ or http://www.paginebianche.it/.

[44]Guglielmino *et al.* (1991) show that in Sicily genetic *and* cultural transmission are revealed by surnames.

the gangsters' origin country (thus unrelated to US housing prices).

Since poor living conditions in Italy were the main driver to migrate to the US, such innate connections are hardly measuring innate wealth. Conditional on the region of birth and on all the other individual characteristics of the mobsters they are also unlikely to be related to individual preferences for housing. Since some of the mobsters might have migrated as a response to increased wealth levels by the potential innate connections (violating the exclusion restriction between potential innate connections and housing wealth), later in the robustness regressions I will control for the place of origin of the gangsters' direct associates (direct links).

Figure 5 visualises how I construct the index. Starting from the current zip code level distribution of the members' surnames,[45] I compute the probability that each members' surname shares a randomly chosen zip code located in Italy (as a robustness check I also limit the attention to the South) with other surnames from the list.[46] To be more precise, the index for member i is equal to 1,000,000 times the sum across zip codes j of the fraction of surnames of member i present in zip codes j times the fraction of surnames of the other members ($-i$) in the same zip code:

$$\text{Potential innate interactions index}_i = 10^9 \sum_j \frac{\#surname_{i,j}}{\sum_j \#surname_{i,j}} \frac{\#surname_{-i,j}}{\sum_j \#surname_{-i,j}}. \quad (3)$$

There are 4,748 zip codes for about 60 million Italians, thus each zip code covers a little more than 12,000 Italians, and an area of about 23 square miles, a reasonable area within which most relationships are likely to get established. In Figure 5 each circle is proportional to the number of surnames present within each zip code. Not surprisingly many surnames show up in Sicily, in Naples, and in Calabria. Many of these surnames

[45]Only four mobsters were neither born in Italy nor in the US. For two of these mobsters (Lansky and Genese) the *Potential Innate Interactions* index is zero.

[46]I use Italy's phone directory http://www.paginebianche.it/ in 2010. Ideally one would use the distribution of surnames in 1960, though previous research has shown how persistent such distribution is (Colantonio *et al.*, 2003).

appear also in large cities that were subject to immigratory flows from the south, like Milan, Rome, and Turin. Such migration patterns introduce some noise in the instrument, which is why later I also compute a *Potential Innate Interactions* measure that is just based on Southern regions (*Campania, Molise, Calabria, Basilicata, Sicilia, Sardegna, Puglia*), with Northern regions acting as imperfect falsification samples (as migration might depend on ethnic networks).

The *Potential innate interactions* index tends to be small when the fraction of surnames i overlaps little with the fraction of all other surnames $-i$. Dividing by the total number of surnames (i and $-i$) takes into account that some surnames are more frequent than others.

The average index is equal to 36 per one million, though taking the sampling into account it drops to 26, already indicating that more connected mobsters have more interactions (see Table 1). About ten percent of the times the index is zero, either because the zip codes do not overlap or because the surname is not in the phone directory.[47]

Figure 6 shows the distribution of the *interaction index*. Mobsters born in Italy tend to have more innate potential interactions than those born outside of Italy (Figure 7). Most mobsters with very large interactions were born in the Western part of Sicily (the 10 cities of birth corresponding to the largest interaction indices are in major US cities (Chicago, NYC, St. Luis), but also Palermo (Sicily), Cerda (Sicily), Trapani (Sicily), Amantea (Calabria). Most of these are well-known mafia enclaves.

When measuring the interactions focussing on zip codes in Southern Italy the shape of the bars is very similar, while when focussing on zip codes in Central and Northern Italy the interaction probabilities are considerably lower, and are lowest among Sicilians, which is consistent with the index capturing interactions in the place of origin. Since Northern interactions are likely to be the product of recent migrations from the South to

[47]In the regressions I allow the zeros to have an independent effect. Moreover, ideally one would us the mother's surname as well, though such information is not always available.

the North of Italy rather than true potential connections in 1960 they represent a useful placebo case.[48]

Potential interactions influence the centrality measures (see Figure 8 and Table 3). Correlations are between 15 (*eigenvector*, *closeness*, and *degree*) and 30 percent (*betweenness*), indicating that such initial interactions are important, though neither necessary, nor sufficient, to reach the top positions in the organisation (those positions that generate bridges across Families). Moreover, the correlation is larger when restricting the interactions to Southern Italy and considerably weaker when restricting the interactions to Northern Italy. For this reason, in most two-stage regressions I will use the interactions based on Southern regions as instruments.

The next Section assigns confidence intervals to the unconditional network effects shown in Figure 3, and adds further controls and an IV strategy to pinpoint such effects.

3. Regression Results

3.1 Evidence based on Ordinary Least Squares regressions

When estimating Eq. 2 using *closeness* centrality by ordinary least squared, I find that doubling the centrality measure increases housing values by about 200 percent (Table 4).[49] Such large elasticities are driven by a very compressed distribution of *closeness* centrality. In terms of standard deviations (SDs), a SD increase in log *closeness* (0.12)

[48]Pull factors of migration that are driven by social networks would preserve some of the original predictive power.

[49]Given that there is one large network, the residuals might be correlated across mobsters. Assuming that such correlation depends on the shortest distance between pairs of gangsters the variance of the residual of gangster i is going to be a function of the sum of all such distances d_{ij} over j. In particular, $\sigma_i = \sum_j \sigma_{ij}(d_{ij})$. Approximating such variance with either a linear function or an exponential function in average distances (the inverse of closeness) one always rejects that distance influences the squared residuals, and thus the correlations across mobsters (see appendix Table 9). Nevertheless, in all regressions the standard errors are clustered at the surname level, which is the level at which the centrality measures are calculated.

increases housing values by one quarter.

The first Column controls only for variables collected from *Zillow.com*, in particular whether the housing unit has been built after 1960 (the year of the FBN records) and whether such information is available. The negative coefficient on this variable is capturing that for more expensive housing Zillow.com is more likely to collect information on the year the house was built. Controlling for additional variables (Column 2) increases somehow the effect, while adding US State of residence fixed effects reduces them (Column 3); partly because the more influential mobsters resided in New York and New Jersey (where housing prices tend to be high).[50] Given that the State of residence measures part of the centrality of mobsters in the remainder of the study I will not control for it. The instrument used in Section 3.2 is by construction unrelated to the State of residence (other than through connections). I will also show that the results change little when controlling for the log state-level average housing price (thus *de facto* using the log of relative housing prices).

The last two columns show that similar results are found when using *closeness* in levels rather than in logs, though in such case the predictive power of centrality drops slightly (in line with the scaling bias). As for the log, a standard deviation increase in *closeness* centrality (equal to 8.27) increases housing values by about one quarter.

Before showing the results for other centrality measures, let me briefly discuss the coefficients on the other regressors. Gangsters born in Italy, in particular those born in Sicily tend to live in cheaper housing, which might either be due to housing preferences or to a more pronounced avoidance to attract attention (though the effects are not significantly different from 0).[51] Age at first arrest, which might represent a (negative) measure of career experience within the organisation tends to be negatively related to housing values, meaning that earlier arrests tend to be related to increased housing val-

[50]Measurement error might also be co-responsible for such drop.

[51]Appendix Table 10 shows that the place of birth, while influencing the housing value, does not introduce heterogeneity in the effect of centrality.

ues. Not just experience, but also being involved with drug dealing seems to be related to higher housing values, though the coefficient stops being significant once state of residence effects are added to the regression (indicating that the drug dealing business used to be geographically clustered). The number of legal businesses tend to be positively associated with housing values, which is not surprising.

Going back to the centrality measures, in Table 5 I substitute *closeness* centrality with the other measures, with and without controlling for additional regressors. A quick look at the R-squared reveals that *closeness* centrality is the strongest predictor of housing value. This is coherent with the first rule of the Mafia, which states that only interlinking affiliates, for example B in a network A-B-C, have the power to introduce a direct link between A and B. Mobsters who are on average closer to all other mobsters will thus need fewer interlinking associates to establish new connections. *Eigenvector centrality* has similar predictive power, which is not surprising given how strongly correlated the two measures are.[52]

The coefficient on *degree*, instead, has relatively larger standard errors, and a standard deviation increase in log *degree* increases housing value by about 11 percent; in line with what emerged in Figure 3, the flattest relationship is the one between log housing value and log *betweenness* centrality. Statistically speaking, the slope is 0.[53]

This is likely driven by the bosses' preference for keeping a low profile and, as argued before, such bias cannot be eliminated.

[52]The elasticity is larger for *closeness* centrality only because its variation is about 10 times smaller than the one of *eigenvector* centrality. One standard deviation increase in log-*closeness* centrality has about the same impact on housing value as a one standard deviation increase in log *eigenvector* centrality.

[53]All the regressions use the weighting strategy developed earlier on, but the results based on unweighted regressions can be seen in the Online appendix Table 11.

3.2 Evidence based on Instrumental Variable regressions

As previously discussed, for a number of reasons (including measurement error) OLS coefficients on centrality measures might be inconsistent. Table 6 presents the first stage (sketched in Figure 8) and the main regressions in a compact form, focussing on the instruments and on the endogenous variable.[54]

Having no potential innate connections is allowed to have its own influence on centrality. All centrality measures are positively correlated with the potential connections, but the coefficient is more significant for *closeness* and *betweenness* centralities, though the instrument is not very strong. In particular, for *closeness* centrality the F-statistic for the excluded instruments is around 3.5 when considering the joint significance of interactions and zero interactions (otherwise it is close to 5). The F-statistic is largest and above 10 for the *betweenness* index. While there are no Montecarlo simulations to determine the bias for clustered standard errors, we should take into account that the estimates are likely to be biased toward OLS. Later I will increase the strength of the instrument by focussing on the 86 percent of gangsters who have been arrested at least once, and for whom all variables are more likely to be measured with precision. But test statistics related to the reduced form regressions, which do not suffer from weak-instrument bias (the Anderson and Rubin p-value as well as the p-value on potential innate interactions alone) tend to be below 10 percent, indicating that the instrument does indeed influence housing values.

Instrumenting the centrality measures the coefficients on *closeness*, *eigenvector*, as well as on *degree* centrality tend to be three times as large as the OLS equivalents, meaning that a standard deviation increase in either *closeness* or *eigenvector* centrality almost doubles the housing value. Such differences are likely to be partly explained by measurement error bias, while there is little evidence that the weakness of the instrument

[54] Appendix Tables 12 and 13 show all the coefficients.

is driving these differences. As in the OLS case, *betweenness* does not appear to be significantly different from 0, confirming that the gangsters with large bridging capacity tend to keep a lower profile, no matter whether the centrality has been reached exploiting potential innate connections or not. Overall the relative precision of the estimates tends to be smaller than for the OLS regressions, which is in part due to the instrument's weakness. Indeed, the standard errors are large, and endogeneity test p-values fluctuate around 10 percent.

Table 7 improves upon the previous results by focussing on samples where variables are measured with greater care. Surnames, which represent the core information for the instrumental variable strategy, are not easily measured inside the Mafia. For secrecy reasons gangsters are typically known by their nickname. This is why the FBN was also collecting the gangsters' aliases.[55] Knowing the exact name is clearly important to reconstruct its geographical distribution in Italy, and for arrested gangsters such information the FBN could be double-checked. Column 1 shows the baseline results. In Column 2 I restrict the analysis to gangsters who have at least one arrest record. These represent about 85 percent of the total, and are potentially a more selected sample. The Kleibergen-Paap Wald F-statistic does indeed improve, and the IV coefficients increase slightly.

Column 3 shows that the results do not change much when computing the potential interactions using all Italian regions, though the power of the instrument decreases considerably.

Another sample for which the information might be more precise are the mobsters born in the United States. Focussing on this group of criminals addresses also the concern that innate connections might still have independent (direct) influence on current wealth,

[55] Some of the aliases for gangsters mentioned before were: Don Vitone, The Old Man (Vito Genovese); Francisco Castiglia, Frank Costello, Frank Saverio, Saveria (Francisco Costiglia); Joe Bananas, Joe Bononno, Joe Bonnano, Joe Bouventre (Joseph Bonanno), Joe Proface (Giuseppe Profaci); Carlo Gambrino, Carlo Gambrieno, Don Carlo (Carlo Gambino).

as this should be less likely to happen for US born ones. The smaller sample reduces the power of the instrument but the instrumental variable coefficient does not change much.

In Table 8 I perform a whole series of additional robustness checks, always keeping the entire sample. The first Column adds the average log housing value at the MSA level to control for the possibility that the results are simply due to sorting of the most influential mobsters in cities that have the highest housing values. The network effect becomes just a little smaller when controlling for such relative differences in housing prices (indicating that centrality is positively correlated with average housing values).

Column 2 does not control for the number of legal businesses owned by the mobster, which reduces slightly the coefficient. Column 3 deals again with the potential violations of the exclusion restriction. Since some of the mobsters, with their surnames, might have migrated as a response to increased wealth levels, I add the fraction of direct links (associates) that were born in Sicily and in peninsular Italy to the regression. Doing so leaves the results almost unchanged.

One could argue that even controlling for the sum of the types of crimes for which the criminals have been arrested and for the sum of legal businesses that the gangster owns, the exclusion restrictions might not hold whenever the potential innate interactions not only influence the gangsters' centrality but also additional remunerative abilities that are not channelled through connections. While there is no definite test to exclude such a possibility, in columns 4 and 5 I control for a series of dummies that describe the gangster's crimes, *modus operandi*, and legal jobs, and potentially control for such unobserved remunerative ability. The potential downside is that adding these controls might capture part of the network effect.

There are 13 crime categories with at least 3 percent of arrested gangsters (murder, robbery, burglary, larceny, drug offence, prohibition offence, illegal detention of weapons, menace, theft, assault, fraud, disorderly behaviour, and other crimes), 6 *modus operandi* that come from FBN descriptions of the gangster's illegal activities (racketeering, drug dealing, gambling, bookmaking, murder, and other), and finally 14 job descriptions with at least 3 percent of mobsters involved (restaurant, drugstore, food industry, real estate, transportation, hotel, import-export, automotive, clothing, manual labour, row materials, amusement, casino, and others). While some of these dummy variables are correlated with both the centrality of mobsters and the log housing value, the network effect changes little when adding these controls and, if anything, becomes even larger.

Column 6 tests the robustness of the results when getting rid of the zero interaction dummy (thus assuming continuity and linearity at 0) and allowing the instrument to have non-linear effects on log *closeness*. The network effect because smaller but is still significant at the 10 percent level. Finally, in the last two columns I add a variable that is clearly endogenous, whether the mobster is married to a "connected" wife, meaning a wife whose maiden name is also the surname of another mobster in the data.

These marriages tend to be arranged for strategic reasons, and allow gangsters to gain additional power (connections) as well as wealth. Adding this variable increases the OLS estimates while keeping the 2SLS estimate almost unchanged, suggesting that omitted variables (which the instrument takes care of) are indeed biasing the OLS estimate toward zero.[56]

Overall, no matter the specification a standard deviation increase in *closeness* centrality–for instance, such difference would mean moving from a median *closeness* to the 90th percentile–increases housing values by about 75 percent.

4. Concluding remarks

This paper estimates how, in 1960, network centrality inside the Italian-American Mafia influenced gangsters' economic prosperity, measured by the value of their homes. In the

[56] Given the potential endogeneity of such marriage one cannot draw more definite conclusions.

overground world, the whole geometry of connections has been shown to be related to a variety of economic outcomes. In the underground world such connections are presumably even more important, and yet evidence on this has mainly been based on ethnographic studies.

Moreover, even in the overground world researchers have rarely gone beyond just documenting correlations, as networks tend to emerge endogenously out of complicated bilateral and multilateral decision processes, and researchers often observe only subsets of the entire network. I deal with the non-random sampling design modelling the law-enforcement data collection process within the network as a Markov process.

With respect to the endogenous nature of networks, social scientists have been able to exploit the geometry of the network to develop identification strategies for direct connections (peer effects), but not yet for measures of how central agents are inside networks. Instruments for networks with reasonable exclusion restrictions are in short supply. Any characteristic that determines someone's position inside his/her network is also likely to directly influence a multitude of other outcomes. In the Mafia, for example, family relationships, wealth, place of birth, etc. might help securing a centrality in the network, but could also be related to the demand for housing.

For migrants with strong ethnic identities, instruments naturally evolve from shocks that happen in the country of origin and are thus less likely to influence economic outcomes in the country of destination. Munshi (2003) uses rainfall in Mexico to instrument for the network size of Mexican immigrants to study how such size influences labour market outcomes. Similarly, this paper instruments network centrality using the potential exposure to connections in the gangster's place of origin (30 to 50 years earlier). In the absence of pre-immigration data, I use the informational content of surnames (*isonomy*) to measure such place. Individual home values in 1960 are reconstructed starting from today's values.

Notwithstanding the distant past and the approximation lead to weak instruments and larger standard errors, the evidence suggests that centrality measures have a sizeable influence on housing values. Mobsters who are on average closer to their peers tend to live in more expensive housing. This is consistent with the importance of interconnection capacity in a secret society where two unlinked mobsters need a common associate to generate a new (direct) link.

But then, mobsters who act as bridges across clusters of the larger network[57] tend to keep a lower profile, and officially reside in less expensive housing (either because they hide their true residence through the use of figureheads or because they really prefer so). The evidence suggests that these tend to be the more important bosses, those who most likely form the governing body of the Mafia (la *Commissione*). In line with Bonnano's autobiography such bosses were less likely to be directly involved in the narcotics businesses, which might be part of the same "attention avoidance strategy" (Bonanno, 1983).

The geometry of the network is likely to influence not just the economic prosperity of the "made" men but also their criminal activity (and viceversa); and therefore the overall crime rates.[58] Understanding how network effects shape the portfolio of crimes that gangsters commit represents a natural extension of this work.

Finally, while data restrictions prevent researchers from performing similar analyses based on more recent organized crime networks, this might hopefully change in the near future. Understanding how central figures grow up inside criminal networks is fundamental to the design of targeted law enforcement strategies.

Giovanni Mastrobuoni, Department of Economics, University of Essex, Collegio Carlo Alberto, Netspar, IZA; submitted August 2013, accepted July 2014.

[57]Given what is known about the Mafia these are Mafia Families, clans, or *mandamenti*.

[58]Mastrobuoni and Patacchini (2012) show that cities with larger Mafia densities (measured by the average *eigenvector* index) exhibit larger violent crime rates.

References

(2007). *MAFIA: The Government's Secret File on Organized Crime. By the United States Treasury Department, Bureau of Narcotics*, HarperCollins Publishers.

Abadinsky, H. (1990). *Organized Crime*, Wadsworth Publishing.

Alatas, V., Banerjee, A., Arun G. Chandrasekhar, R.H. and Olken, B.A. (2012). 'Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia', .

Allesina, S. (2011). 'Measuring nepotism through shared last names: the case of Italian academia', *PLoS one*, vol. 6(8), p. e21160.

Angelucci, M., De Giorgi, G., Rangel, M.A. and Rasul, I. (2010). 'Family networks and school enrolment: Evidence from a randomized social experiment', *Journal of Public Economics*, vol. 94(3), pp. 197–221.

Baccara, M. and Bar-Isaac, H. (2008). 'How to Organize Crime', *Review of Economic Studies*, vol. 75(4), pp. 1039–1067.

Baker, W.E. and Faulkner, R.R. (1993). 'The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry', *American Sociological Review*, vol. 58(6), pp. pp. 837–860.

Ballester, C., Zenou, Y. and Calvó-Armengol, A. (2006). 'Who's who in networks. wanted: The key player', *Econometrica*, vol. 74(5), pp. 1403–1417.

Ballester, C., Zenou, Y. and Calvó-Armengol, A. (2010). 'Delinquent networks', *Journal of the European Economic Association*, vol. 8(1), pp. 34–61.

Banerjee, A., Banerjee, A., Chandrasekhar, A.G., Duflo, E. and Jackson, M.O. (2013). 'The Diffusion of Microfinance', *Science*, vol. 341(1236498).

Banfield, E.C. (1967). 'The moral basis of a backward society.', .

Barrai, I., Rodriguez-Larralde, A., Mamolini, E. and Scapoli, C. (1999). 'Isonymy and isolation by distance in italy', *Human Biology*, pp. 947–961.

Bayer, P., Hjalmarsson, R. and Pozen, D. (2009). 'Building criminal capital behind bars: Peer effects in juvenile corrections', *The Quarterly Journal of Economics*, vol. 124(1), pp. 105–147.

Bayer, P., Ross, S. and Topa, G. (2008). 'Place of work and place of residence: Informal hiring networks and labor market outcomes', *Journal of Political Economy*, vol. 116, pp. 1150–1196.

Bleakley, H. and Lin, J. (2012). 'Portage and path dependence', *The Quarterly Journal of Economics*.

Blume, L.E., Brock, W.A., Durlauf, S.N. and Jayaraman, R. (2012). 'Linear Social Network Models', .

Bonanno, J. (1983). *A Man of Honor: The Autobiography of Joseph Bonanno*, St. Martin's.

Borgatti, S.P. (2003). *The Key Player Problem*, R. Breiger, K. Carley and P. Pattison, Committee on Human Factors, National Research Council, pp. 241–252.

Buonanno, P., Durante, R., Prarolo, G. and Vanin, P. (forthcoming). 'Poor institutions, rich mines: Resource curse and the origins of the sicilian mafia', *The Economic Journal*, forthcoming.

Chandrasekhar, A.G. and Lewis, R. (2011). 'Econometrics of sampled networks', .

Colantonio, S., Lasker, G.W., Kaplan, B.A. and Fuster, V. (2003). 'Use of surname models in human population biology: a review of recent developments', *Human Biology*, vol. 75(6), pp. 785–807.

Coles, N. (2001). 'It's Not What You Know-It's Who You Know That Counts. Analysing Serious Crime Groups as Social Networks', *British Journal of Criminology*, vol. 41(4), p. 580.

Critchley, D. (2009). *The Origin of Organized Crime in America: The New York City Mafia, 1891–1931*, Routledge.

Drago, F. and Galbiati, R. (2012). 'Indirect effects of a policy altering criminal behavior: Evidence from the italian prison experiment', *American Economic Journal: Applied Economics*, vol. 4(2), pp. 199–218.

Ductor, L., Fafchamps, M., Goyal, S. and van der Leij, M.J. (forthcoming). 'Social networks and research output', *Review of Economics and Statistics*.

Fafchamps, M., Vaz, A. and Vicente, P.C. (2013). 'Voting and Peer Effects : Experimental Evidence from Mozambique', .

Falcone, G. and Padovani, M. (1991). *Cose di Cosa Nostra*, Rizzoli.

Fiorentini, G. and Peltzman, S. (1997). *The Economics of Organised Crime*, Cambridge University Press.

Frank, O. (1979). 'Sampling and estimation in large social networks', *Social networks*, vol. 1(1), pp. 91–101.

Gambetta, D. (1996). *The Sicilian Mafia: the business of private protection*, Harvard Univ Press.

Garoupa, N. (2007). ‘Optimal law enforcement and criminal organization’, *Journal of Economic Behavior & Organization*, vol. 63(3), pp. 461–474.

Glaeser, E.L., Sacerdote, B. and Scheinkman, J.A. (1996). ‘Crime and social interactions’, *The Quarterly Journal of Economics*, vol. 111(2), pp. 507–548.

Golub, B. and Jackson, M.O. (2010). ‘Using selection bias to explain the observed structure of internet diffusions’, *Proceedings of the National Academy of Sciences*, vol. 107(24), pp. 10833–10836.

Goodman, A.C. (1988). ‘An econometric model of housing price, permanent income, tenure choice, and housing demand’, *Journal of Urban Economics*, vol. 23(3), pp. 327–353.

Goodman, L.A. (1961). ‘Snowball sampling’, *The Annals of Mathematical Statistics*, vol. 32(1), pp. 148–170.

Gosch, M.A. and Hammer, R. (1975). *The Last Testament of Lucky Luciano*, Little, Brown.

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

Granovetter, M. (1976). ‘Network sampling: Some first steps’, *American Journal of Sociology*, pp. 1287–1303.

Granovetter, M.S. (1973). ‘The strength of weak ties’, *American Journal of Sociology*, vol. 78, pp. 1360–1380.

Guglielmino, C., Zei, G. and Cavalli-Sforza, L. (1991). ‘Genetic and cultural transmission in sicily as revealed by names and surnames’, *Human biology*, pp. 607–627.

Guiso, L., Sapienza, P. and Zingales, L. (2006). 'Does Culture Affect Economic Outcomes?', *The Journal of Economic Perspectives*, vol. 20(2), pp. 23–48.

Gyourko, J., Mayer, C. and Sinai, T. (2013). 'Superstar cities', *American Economic Journal: Economic Policy*, vol. 5(4), pp. 167–99.

Haynie, D.L. (2001). 'Delinquent peers revisited: Does network structure matter? 1', *American journal of sociology*, vol. 106(4), pp. 1013–1057.

Heckathorn, D.D. (1997). 'Respondent-driven sampling: a new approach to the study of hidden populations', *Social problems*, pp. 174–199.

Hochberg, Y.V., Ljungqvist, A. and Lu, Y. (2007). 'Whom you know matters: Venture capital networks and investment performance', *The Journal of Finance*, vol. 62(1), pp. 251–301.

Ianni, F.A. and Reuss-Ianni, E. (1972). *A Family Business: Kinship and Social Control in Organized Crime*, Russell Sage Foundation.

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

Jacobs, J.B. and Gouldin, L.P. (1999). 'Cosa nostra: The final chapter?', *Crime and Justice*, vol. 25, pp. pp. 129–189.

Karlan, D., Mobius, M., Rosenblat, T. and Szeidl, A. (2009). 'Trust and social collateral', *The Quarterly Journal of Economics*, vol. 124(3), pp. 1307–1361.

Kinnan, C. and Townsend, R. (2012). 'Kinship and financial networks, formal financial access, and risk reduction', *The American Economic Review*, vol. 102(3), pp. 289–293.

Krebs, V.E. (2002). 'Mapping Networks of Terrorist Cells', *Connections*, vol. 24(3), pp. 43–52.

Kremer, M. and Miguel, E. (2007). 'The illusion of sustainability', *The Quarterly Journal of Economics*, vol. 122(3), pp. 1007–1065.

Kumar, V. and Skaperdas, S. (2009). *Criminal Law and Economics*, chap. On The Economics of Organized Crime, Edward Elgar Pub.

Lasker, G.W. (1977). 'A coefficient of relationship by isonymy: a method for estimating the genetic relationship between populations', *Human Biology*, pp. 489–493.

Lee, S.H., Kim, P.J. and Jeong, H. (2006). 'Statistical properties of sampled networks', *Physical Review E*, vol. 73(1), p. 016102.

Lupo, S. (2009). *Quando la Mafia Trovò l'America*, Einaudi.

Maas, P. (1968). *The Valachi Papers*, Putnam, New York.

Margo, R.A. (1996). 'The rental price of housing in new york city, 1830-1860', *The Journal of Economic History*, vol. 56, pp. 605–625.

Mastrobuoni, G. and Patacchini, E. (2012). 'Organized Crime Networks: an Application of Network Analysis Techniques to the American Mafia', *Review of Network Economics*, vol. 3.

McGloin, J.M. (2005). 'Policy And Intervention Considerations of a Network Analysis of Street Gangs', *Criminology & Public Policy*, vol. 4(3), pp. 607–635.

McWilliams, J.C. (1990). *The Protectors: Harry J. Anslinger and the Federal Bureau of Narcotics, 1930-1962*, Univ of Delaware Press.

Montgomery, J.D. (1991). 'Social networks and labor-market outcomes: Toward an economic analysis', *The American Economic Review*, vol. 81(5), pp. 1408–1418.

Morselli, C. (2003). 'Career opportunities and network-based privileges in the Cosa Nostra', *Crime, Law and Social Change*, vol. 39(4), pp. 383–418.

Munshi, K. (2003). 'Networks in the Modern Economy: Mexican Migrants in the US Labor Market', *Quarterly Journal of Economics*, vol. 118(2), pp. 549–599.

Natarajan, M. (2000). 'Understanding the Structure of a Drug Trafficking Organization: A Conversational Analysis', *Crime Prevention Studies*, vol. 11, pp. 273–298.

Natarajan, M. (2006). 'Understanding the Structure of a Large Heroin Distribution Network: A Quantitative Analysis of Qualitative Data', *Journal of Quantitative Criminology*, vol. 22(2), pp. 171–192.

Patacchini, E. and Zenou, Y. (2008). 'The strength of weak ties in crime', *European Economic Review*, vol. 52, pp. 209–236.

Pinotti, P. (forthcoming). 'The economic costs of organized crime: evidence from southern italy', *The Economic Journal*.

Putnam, R.D., Leonardi, R. and Nanetti, R.Y. (1994). *Making democracy work: Civic traditions in modern Italy*, Princeton University Press.

Reuter, P. (1983). *Disorganized Crime: the Economics of the Visible Hand*, MIT press.

Reuter, P. (1994). *Handbook of organized crime in the United States*, chap. Research on American Organized Crime, Westport, CT: Greenwood Press.

Rothenberg, R.B. (1995). 'Commentary: sampling in social networks', *Connections*, vol. 18(1), pp. 104–110.

Sarnecki, J. (1990). 'Delinquent Networks in Sweden', *Journal of Quantitative Criminology*, vol. 6(1), pp. 31–50.

Sarnecki, J. (2001). *Delinquent Networks: Youth Co-offending in Stockholm*, Cambridge Univ Pr.

Sirakaya, S. (2006). 'Recidivism and social interactions', *Journal of the American Statistical Association*, vol. 101(475), pp. 863–877.

Snijders, T.A. (1992). 'Estimation on the basis of snowball samples: How to weight?', *Bulletin de méthodologie sociologique*, vol. 36(1), pp. 59–70.

Sparrow, M.K. (1991). 'The Application of Network Analysis to Criminal Intelligence: an Assessment of the Prospects', *Social Networks*, vol. 13(3), pp. 251 – 274.

Vega-Redondo, F. (2007). *Complex Social Networks*, Econometric Society Monograph Series, Cambridge: Cambridge University Press.

Villarreal, C. (2013). 'Where the other half lives: Evidence on the origin and persistence of poor neighborhoods from new york city 1830-2012', Mimeo.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis. Methods and Applications*, cambridge: cambridge university press edn.

Williams, P. (2001). 'Transnational Criminal Networks', *Networks and netwars: the future of terror, crime, and militancy*, p. 61.

Wolf, G. and DiMona, J. (1974). *Frank Costello: Prime Minister of the Underworld*, Morrow.

Zei, G., Barbujani, G., Lisa, A., Fiorani, O., Menozzi, P., Siri, E. and Cavalli-Sforza, L.L. (1993). 'Barriers to gene flow estimated by surname distribution in italy', *Annals of Human Genetics*, vol. 57(2), pp. 123–140.

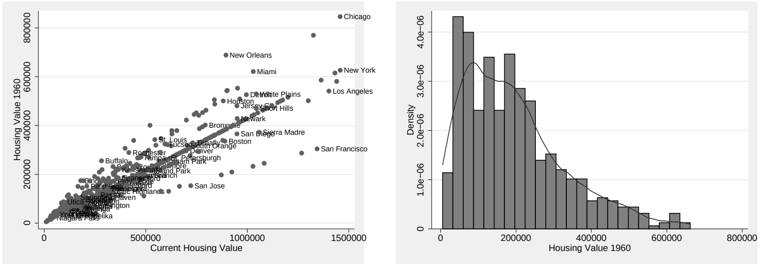


Fig. 1: *Housing Value Deflators and Distribution*

Notes: Current housing prices (2010) are taken from *Zillow.com*. Deflators use MSA or State level housing values from the Census (Gyourko *et al.*, 2013).

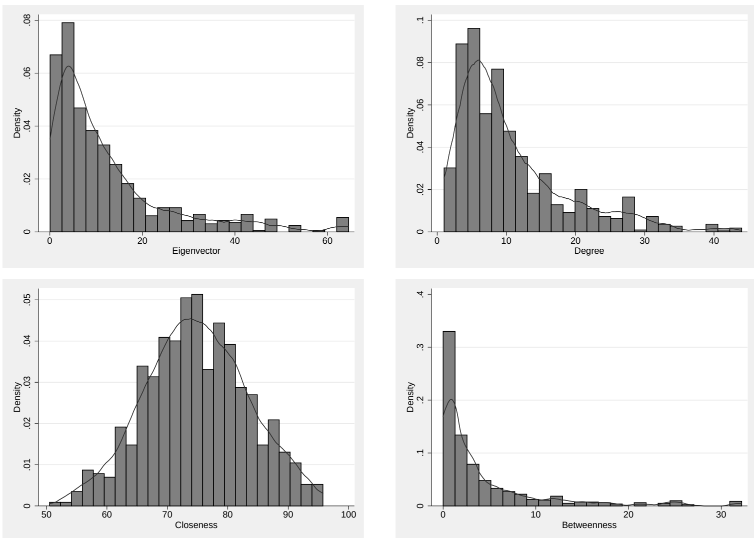


Fig. 2: *Densities of Centrality Indices*

Notes: The density shown excludes the top 1 percent of the distribution.

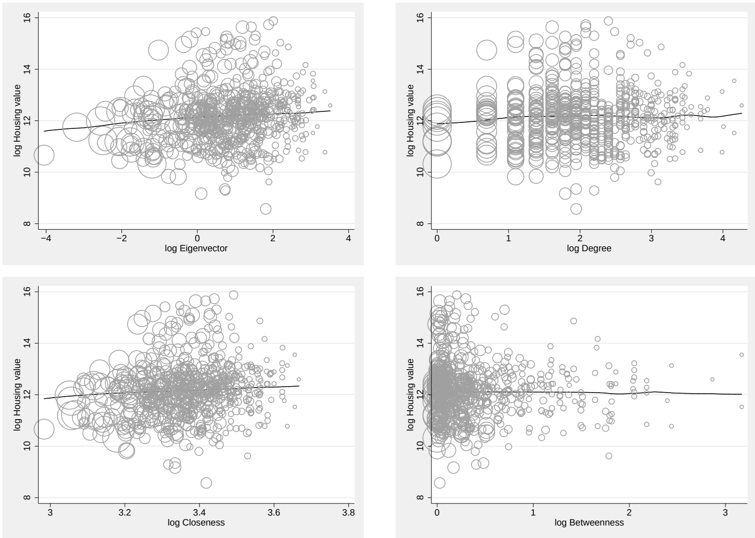


Fig. 3: *Housing values and Network Centrality (in logs)*

Notes: The size of the circles is proportional to the sampling weight. The solid line corresponds to the local polynomial smoothing regression line (using the Epanechnikov kernel).

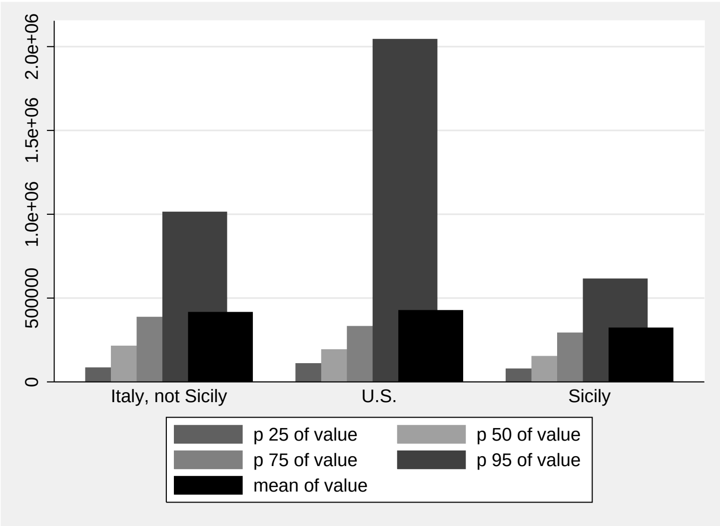


Fig. 4: *Distribution of Housing Values by Region of Birth*

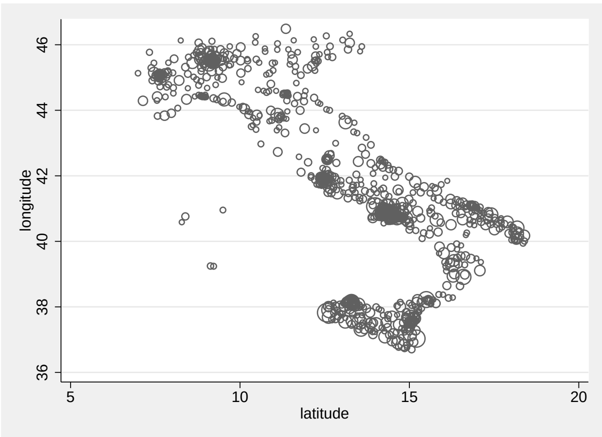


Fig. 5: *Geographical Distribution of Mafia Surnames*

Notes: Each circle represents a zip code. The size of the circles is proportional to the number of US Mafia members' surnames found in today's Italian phone directory.

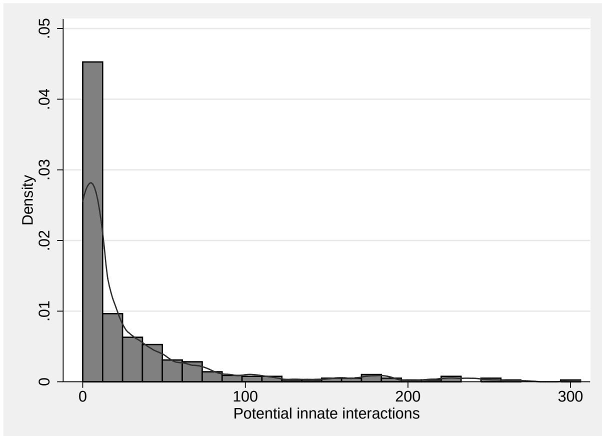


Fig. 6: *Density of Potential Innate Interactions*

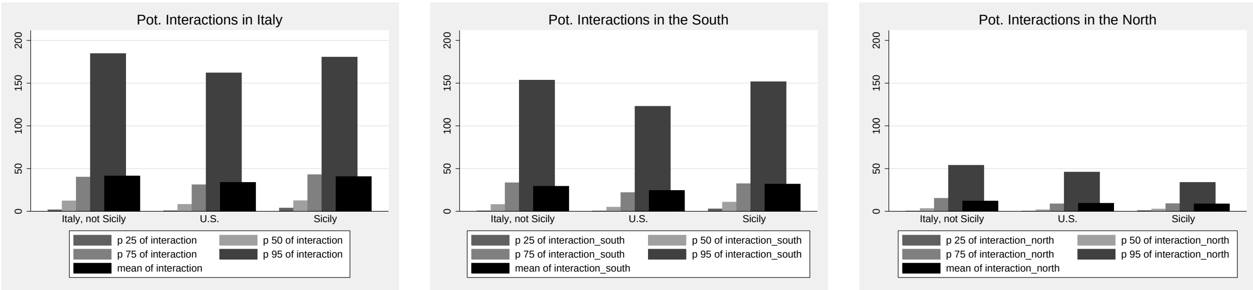


Fig. 7: *Distribution of the Potential Innate Interactions Index by Region of Birth*

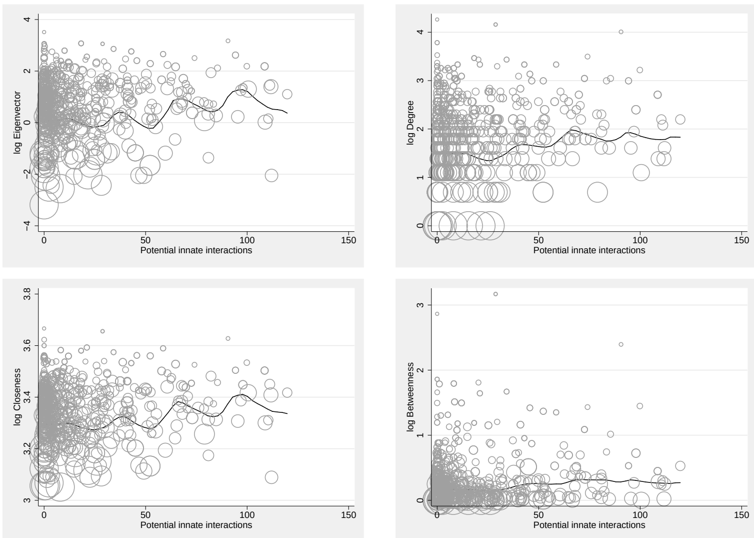


Fig. 8: *Log-Network Centrality and Potential Innate Interactions Index*

Notes: The size of the circles is proportional to the sampling weight. The solid line corresponds to the local polynomial smoothing regression line (using the Epanechnikov kernel).

Table 1: Summary Statistics

	Unweighted		Weighted			
	Mean	St.Dev.	Mean	St.Dev.	Min	Max
Individual characteristics						
Housing Value 1960	397,678	805,186	378,196	770,101	5,286	7,866,203
Potential innate interaction	36.18	76.79	26.10	51.59	0	654.02
Housing built after 1960	0.10	0.30	0.14	0.35	0	1
Year built unknown	0.12	0.33	0.12	0.33	0	1
Extended family members	1.58	1.01	1.30	0.73	1	6
Born outside Italy (mainly U.S.)	0.67	0.47	0.71	0.46	0	1
Born in Sicily	0.24	0.43	0.20	0.40	0	1
Age	48.63	7.59	48.20	7.57	24	72
Year of birth unknown	0.21	0.41	0.19	0.39	0	1
Height in feet	5.61	0.20	5.60	0.19	5	6.17
Weight in pounds	177.66	26.87	176.63	29.12	95	365
Age at first arrest	23.03	7.32	23.09	7.11	8	57
Never arrested	0.14	0.34	0.14	0.35	0	1
Connected wife	0.19	0.39	0.15	0.36	0	1
Married	0.80	0.40	0.76	0.43	0	1
Divorced	0.07	0.26	0.10	0.30	0	1
Number of children	1.07	1.47	1.06	1.44	0	8
Siblings	2.10	2.14	2.02	2.08	0	11
Types of crime committed	2.64	1.68	2.71	1.75	0	9
Number of businesses	1.08	1.01	1.00	0.94	0	5
Involved in drug dealing	0.56	0.50	0.59	0.49	0	1
Network characteristics						
Degree	11.10	9.28	6.21	5.52	1	71
Eigenvector (std.)	12.77	14.12	6.59	8.31	0.05	100
Betweenness (std.)	4.94	9.33	1.90	4.21	0	100
Closeness (std.)	75.12	8.76	69.78	8.27	50.51	100

Table 2: *Housing Values and Centralities by Leadership Status*

	Mafia leader (weighted)		Mafia leader (unweighted)	
	No (91%)	Yes (9%)	No (87%)	Yes (13%)
Housing value	380,453	362,479	399,654	387,587
Involved in drug dealing	0.62	0.39	0.58	0.45
Betweenness	1.76	2.86	4.26	8.46
Closeness	69.63	70.85	74.51	78.23
Eigenvector	6.21	9.26	11.54	19.03
Degree	5.94	8.10	10.25	15.46

Notes: About 13 percent of gangsters are described as either "leader" or "boss" in the FBN records.

Table 3: *Correlation Table*

	log Clo.	log Bet.	log Eig.	log Deg.	Pot. Int.	Pot. Int. South
log Betweenness	0.533					
log Eigenvector	0.968	0.460				
log Degree	0.801	0.646	0.806			
Potential innate interactions	0.159	0.309	0.121	0.165		
Pot. Int. in the South	0.185	0.345	0.148	0.191	0.959	
Pot. Int. in the North	0.060	0.150	0.034	0.066	0.811	0.610

Notes: Clustered (by surname) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: *Housing Value Regressions and Closeness Centrality*

	(1)	(2)	(3)	(4)	(5)
			log Housing value		
log Closeness	2.069***	2.524***	1.418***		
	(0.406)	(0.452)	(0.497)		
Closeness				0.036***	0.020***
				(0.007)	(0.007)
Housing built after 1960	0.100	0.054	0.048	0.051	0.045
	(0.156)	(0.159)	(0.157)	(0.160)	(0.158)
Year built unknown	-0.346***	-0.277**	-0.298**	-0.276**	-0.298**
	(0.121)	(0.122)	(0.140)	(0.122)	(0.140)
Extended family members		-0.216***	-0.067	-0.224***	-0.069
		(0.073)	(0.056)	(0.074)	(0.056)
Born outside Italy (mainly U.S.)		-0.089	-0.038	-0.085	-0.035
		(0.173)	(0.160)	(0.172)	(0.160)
Born in Sicily		-0.273	-0.173	-0.269	-0.169
		(0.188)	(0.161)	(0.188)	(0.161)
Age		-0.065	-0.076	-0.063	-0.075
		(0.074)	(0.075)	(0.074)	(0.075)
Age squared/100		0.067	0.082	0.065	0.081
		(0.081)	(0.080)	(0.081)	(0.080)
Year of birth unknown		-1.750	-1.935	-1.729	-1.920
		(1.602)	(1.663)	(1.610)	(1.664)
Height in feet		0.530*	0.224	0.536*	0.224
		(0.297)	(0.279)	(0.299)	(0.280)
Weight in pounds		-0.001	-0.000	-0.001	-0.000
		(0.002)	(0.002)	(0.002)	(0.002)
Age at first arrest		-0.013*	-0.016***	-0.013*	-0.016***
		(0.007)	(0.006)	(0.007)	(0.006)
Never arrested		-0.127	-0.181	-0.141	-0.188
		(0.204)	(0.183)	(0.205)	(0.183)
Married		-0.080	-0.000	-0.083	-0.001
		(0.151)	(0.138)	(0.151)	(0.139)
Divorced		-0.042	-0.005	-0.052	-0.010
		(0.222)	(0.184)	(0.222)	(0.184)
Number of children		-0.030	-0.062**	-0.031	-0.064**
		(0.030)	(0.028)	(0.030)	(0.028)
Siblings		0.016	-0.010	0.016	-0.010
		(0.025)	(0.025)	(0.026)	(0.025)
Types of crime committed		0.021	-0.002	0.019	-0.003
		(0.038)	(0.034)	(0.038)	(0.034)
Number of businesses		0.043	0.092**	0.043	0.092**
		(0.051)	(0.043)	(0.052)	(0.043)
Involved in drug dealing		0.346***	0.003	0.344***	-0.000
		(0.099)	(0.091)	(0.099)	(0.091)
State of residence fixed effects			X		X
Observations	641	641	637	641	637
R-squared	0.066	0.144	0.349	0.141	0.348

Notes: Clustered (by 530 surnames) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: *Housing Value Regressions With Other Centrality Measures*

	(1)	(2)	(3)	(4)	(5)	(6)
	log Housing Value					
log Eigenvector	0.184***	0.218***				
	(0.040)	(0.043)				
log Degree			0.148**	0.220***		
			(0.068)	(0.083)		
log Betweenness					-0.085	0.036
					(0.115)	(0.158)
Other Xs		X		X		X
Observations	641	641	641	641	641	641
R-squared	0.062	0.136	0.025	0.098	0.014	0.076

Notes: All regressions control for the housing variables. The additional regressors ("Other Xs") are the same as in Table 4. Clustered (by 530 surnames) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: *Two Stage Least Squares Regressions*

	(1)	(2)	(3)	(4)
Panel A: First stage regression	**log Closeness**	**log Eigenvector**	**log Degree**	**log Betweennes**
Potential innate interactions (in %)	0.033**	0.268*	0.116	0.138***
	(0.015)	(0.158)	(0.094)	(0.053)
Zero potential innate interactions	-0.019	-0.308	-0.236	0.011
	(0.020)	(0.234)	(0.153)	(0.032)
R-squared	0.160	0.146	0.205	0.326
Kleibergen-Paap rk Wald F statistic	3.479	2.663	2.343	12.60
F-stat on potential innate interactions	5.208	2.887	1.523	6.726
Panel B: Main regression		**log Housing value**		
log Closeness	6.350**			
	(3.196)			
log Eigenvector		0.627*		
		(0.329)		
log Degree			0.977*	
			(0.537)	
log Betweenness				1.201
				(1.054)
Observations	641	641	641	641
Endogeneity p-value	0.305	0.214	0.146	0.244
Kleibergen-Paap rk Wald F statistic	3.479	2.663	2.343	3.399
Anderson and Rubin p-value		0.212		

Notes: Panel A shows the first stage regressions, Panel B the main log Housing value 2SLS regressions. Additional housing variables and the additional regressors ("Other Xs") are the same as in Table 4. The Potential Innate Interactions are based on Southern regions.
The Stock-Yogo weak ID test critical values for F tests on single endogenous regressor are: 16.38 (10% maximal IV size), 8.96 (15%), 6.66 (20%) and 5.53 (25%). Clustered (by surname) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 7: *Two-stage Regressions With Stronger First Stages*

	(1)	(2)	(3)	(4)
Arrested gangsters only		$\surd$		
U.S. born gangsters only				$\surd$
Panel A: First stage regression		**log Closeness**		
Potential innate interactions (in %)	0.033**	0.036**		0.027
	(0.015)	(0.014)		(0.018)
Zero potential innate interactions	-0.019	-0.028		-0.024
	(0.020)	(0.022)		(0.020)
PII using whole Italy			0.023**	
			(0.011)	
Zero PII in whole Italy			0.000	
			(0.000)	
R-squared	0.160	0.163	0.155	0.177
Kleibergen-Paap rk Wald F statistic	3.479	4.328	2.670	2.234
F-stat on potential innate interactions	5.208	6.540	4.899	2.394
Panel B: Main regression		**log Housing value**		
log Closeness	6.350**	7.103**	8.343**	8.784**
	(3.196)	(3.223)	(4.144)	(4.344)
Observations	641	554	641	429
Endogeneity p-value	0.212	0.0838	0.168	0.115
p value on PII	0.257	0.0657	0.111	0.0908
Anderson and Rubin p-value	0.212	0.0838	0.168	0.115

Notes: All regressions control for the additional housing variables and the additional regressors as in Table 4. The Southern regions are *Puglia*, *Campania*, *Calabria*, *Molise*, and *Sicilia*. The Potential Innate Interactions are based on Southern regions. Clustered (by surname) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 8: *Robustness Regressions*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	OLS	2SLS
Panel A: First stage regression				**log Closeness**				
Potential innate interactions (in %)	0.032***	0.033**	0.036**	0.032***	0.027**	0.040		0.034**
	(0.012)	(0.015)	(0.015)	(0.012)	(0.013)	(0.039)		(0.015)
Zero potential innate interactions	-0.019	-0.016	-0.012	-0.011	-0.020			-0.017
	(0.019)	(0.020)	(0.019)	(0.019)	(0.020)			(0.020)
PII squared						-0.007		
						(0.036)		
PII cube						0.001		
						(0.005)		
R-squared	0.224	0.178	0.238	0.228	0.306	0.155		0.165
F-stat on potential innate interactions	7.531	4.950	6.084	6.693	4.165	1.009		5.413
Panel B: Main regression				**log Housing value**				
log Closeness	6.370**	6.175*	6.004*	7.294**	6.893*	5.085*	2.486***	6.274**
	(3.231)	(3.326)	(3.108)	(3.695)	(3.701)	(3.010)	(0.453)	(3.174)
Fraction of Sicilian associates			-0.634					
			(0.393)					
Fraction of Italian associates			-0.049					
			(0.443)					
Connected wife							0.139	-0.022
							(0.143)	(0.228)
Average log housing value (MSA)	√							
Without controlling for legal businesses		√						
13 job dummies					√			
5 modus operandi dummies				√				
12 crime convictions dummies				√				
Kleibergen-Paap rk Wald F statistic	4.685	3.118	3.664	3.910	2.829	8.331		3.527
Endogeneity p-value	0.205	0.355	0.405	0.255	0.309	0.196		0.302
Anderson and Rubin p-value	0.149	0.257	0.196	0.151	0.227	0.0156		0.212
p-value on additional dummies			0.229	0.0322	0.590			

Notes: The additional controls are the same as in the last columns of Table 4. Column 4 is based on the subset of mobsters born in the United States, all the other regressions use the entire sample. Additional coefficients are shown in appendix Tables **??** and **??**. The Potential Innate Interactions are based on Southern regions. Clustered (by 530 surnames) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

A. Appendix: A Short History of the Mafia and The American Mafia

This section is a brief history of the so-called "made" men of came to the United States, and a description of how the Mafia operated in the 1960s when the FBN was filing the 800 records.

There is evidence of Mafia style organisations (*Camorra* in Campania, *'Ndrangheta* in Calabria, and *Mafia* in Sicily) operating on Italian post-unification (1861) territory.[59] The collapse of Southern feudal institutions when the Northern troops conquered their territories led to a sudden lack of law enforcement, and with a rising demand for protection local bandits started offering such services (Gambetta, 1996). The additional ingredient that might have contributed to the emergence of the Mafia is what Banfield (1967) described as a self-interested, family centric society which sacrificed the public good, or social capital. Lack of social capital has probably prevented an early eradication of the Mafia in Southern Italy.[60]

This means that the land that Southern Italian migrants were leaving behind some 50 years later was already in the hands of the Mafia. Historians define two major waves of immigration from Sicily, before and after World War I (WWI). Before WWI immigrants were mainly driven by economic needs. Several Mafia bosses, like Lucky Luciano, Tommaso Lucchese, Vito Genovese, Frank Costello, etc, were children of these early immigrants. Even though between 1901 and 1913 almost a quarter of Sicily's population departed for America, many of the early immigrant families were not from Sicily. In that period around two million Italians, mainly from the South emigrated to the US

[59]See Buonanno *et al.* (forthcoming).

[60]Putnam *et al.* (1994) documents large differences in Social Capital between Southern and Northern Italy and Pinotti (forthcoming) shows not only that the Mafia is expanding to other Southern regions, but also that the cost of such expansions are large in terms of lost economic growth.

(Critchley, 2009). These baby immigrants later became street gang members in the slums; they spoke little Italian, and worked side by side with criminals from other ethnicities, mainly Jewish and Irish (Lupo, 2009). Lured by the criminal successes of the first wave of immigrants, and (paradoxically) facilitated by prohibitionism, the second wave of immigrants that went on to become Mafia bosses were already criminals by the time they entered the United States. Charles Gambino, Joe Profaci, Joe Bonanno, and others were in their 20s and 30s when they first entered the US (the average age in 1960 of Italian born gangsters is indeed 6 years higher (54 years) than for American born ones), and they all came from Sicily. Another reason for this selection of immigrants was the fascist crack-down of the Mafia, which forced some of these criminals to leave Sicily. After the second wave of immigration the Mafia became more closely linked to the Sicilian Mafia and started adopting its code of honour and tradition.[61]

In 1930 and 1931 these new arrivals led to a Mafia war, called the Castellamare war, named after a small city in Sicily where many of the new Mafia bosses came from. The war lasted until Maranzano, who was trying to become the "Boss of the Bosses," was killed, probably by Lucky Luciano who had joined the Masseria Family.[62] This war put Lucky Luciano at the top of the Mafia organisation but also led to a reaction by the media and the prosecutors.[63] Between 1950 and 1951, the Kefauver Committee, officially the Senate Special Committee to Investigate Crime in Interstate Commerce, had a profound impact on the American public. It was the first committee set up to gain a better understanding of how to fight organized crime, and the main source of information was a list of 800 suspected criminals submitted by FBN's Commissioner Anslinger, most likely an early

[61]See Gosch and Hammer (1975).

[62]Before this event, in order to end the power-struggle between Masseria and Maranzano, Lucky Luciano had offered to eliminate Joe "the Boss" Masseria, which he did at an Italian restaurant by poisoning Masseria's food with lead.

[63]In 1936 Thomas E. Dewey, appointed as New York City special prosecutor to crack down on the rackets, managed to obtain Luciano's conviction with charges on multiple counts of forced prostitution. Luciano served only 10 of the 30 to 50 years sentenced. In 1946 thanks to an alleged involvement in the Allied troops' landing in Sicily he was deported to Italy, from where he tried to keep organising "the organisation."

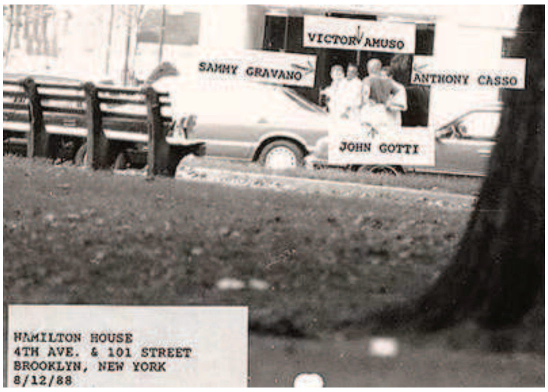
version of the records used in this paper (McWilliams, 1990, pg. 141).[64]

Throughout the 1950s the FBN continued to investigate the Mafia, and in 1957, an unexpected raid of an American Mafia summit, the "Apalachin meeting," captured considerable media attention. Police detained over 60 underworld bosses from the raid. After that meeting everyone had to agree with the FBN's view that there was one large and well organized Mafia society. Robert Kennedy, attorney general of the United States, and J. Edgar Hoover, head of the Federal Bureau of Investigations, joined Harry J. Anslinger, the US Commissioner of Narcotics, in his war against the mob. The same years a permanent Senate Select Committee was formed – the McClellan commission. Anslinger's FBN conducted the investigative work and coordinated nationwide arrests of Apalachin defendants. Lucky Luciano died of a heart attack at the airport of Naples in 1962.

[64]The Committee could not prove the existence of a Mafia and after Luciano's expatriation several other Families headed the organisation: Costello, Profaci, Bonanno, and Gambino. Family ties were of utmost importance. According to Bonanno's autobiography (Bonanno, 1983), he became the Boss of the Bosses in part by organising the marriage between his son Bill and the daughter of Profaci, Rosalia in 1956. In 1957 Gambino took over the leadership.

B. Online Appendix Figures

Fig. 9: *Surveillance Pictures of Nick Giso in 1980, and Sammy Gravano in 1988*



1

NAME	: Joseph BONANNO
ALIASES	: Joe Bananas, Joe Bononno, Joe Bonnano, Joe Bouventre
DESCRIPTION	: Born 1-18-1905 Castellammare, Sicily, 5'9", 190 lbs, brown eyes, brown-grey hair, naturalized 5-17-45, Brooklyn, NY.
LOCALITIES FREQUENTED	: Resides 1847 East Elm Street, Tucson, Arizona. Travels extensively about U.S. & makes frequent trips to Italy.
FAMILY BACKGROUND	: Married Filippa LaBruzzo; daughter: Catherine; sons: Salvatore (married to Rosalie Profaci, niece of Giuseppe Profaci) and Joseph: father: Salvatore; mother: Catherine Bouventre; both parents deceased.
CRIMINAL ASSOCIATES	: Lucky Luciano, Francisco Costiglia, Giuseppe Profaci, Anthony Corallo, Thomas Lucchese, Carmine Galante.
CRIMINAL HISTORY	: FBI #2534540 NYCPD #B-85172 I&NS #C-6602167 Record dating from 1930 includes arrests for grand larceny, possession of gun, transportation of machine guns, obstruction of justice.
BUSINESS	: Has interests in Grande Cheese Co., Fond du Lac, Wis.; Alliance Realty & Insurance, Tucson, Arizona; and Brunswick Laundry Service, Brooklyn, N.Y.
MODUS OPERANDI	: Attended 1957 Apalachin Mafia meeting and Binghamton, NY, meeting 1956. One of the most important Mafia leaders in U.S. and attends all top-level Mafia meetings. Makes trips to Italy to confer with Mafia leaders there and to negotiate for international narcotic trafficking.

Fig. 10: *Record Number One: Joe Bonanno*

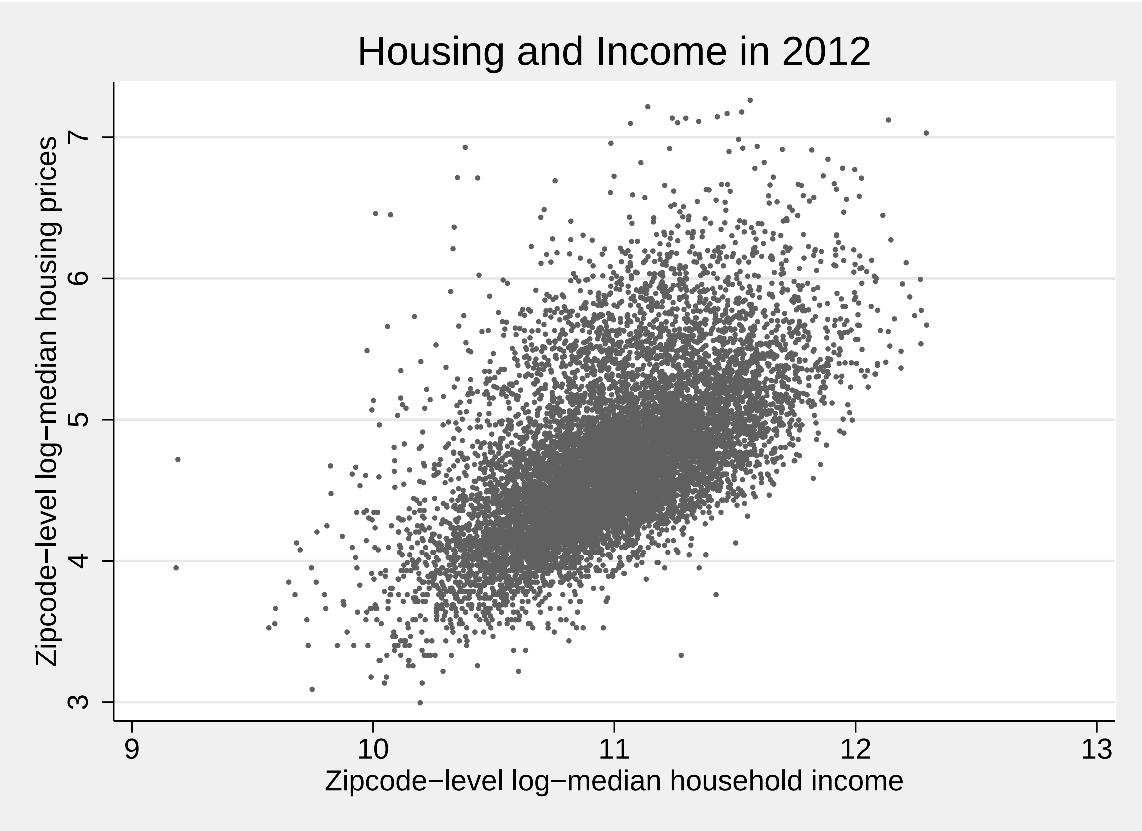


Fig. 11: *Correlation at the Zip Code-level Between Log-Median Housing Prices and Log-Median Income*

Notes: The Zip code-level data on housing and income are available at http://www.zillow.com/blog/research/data/, and http://factfinder2.census.gov.

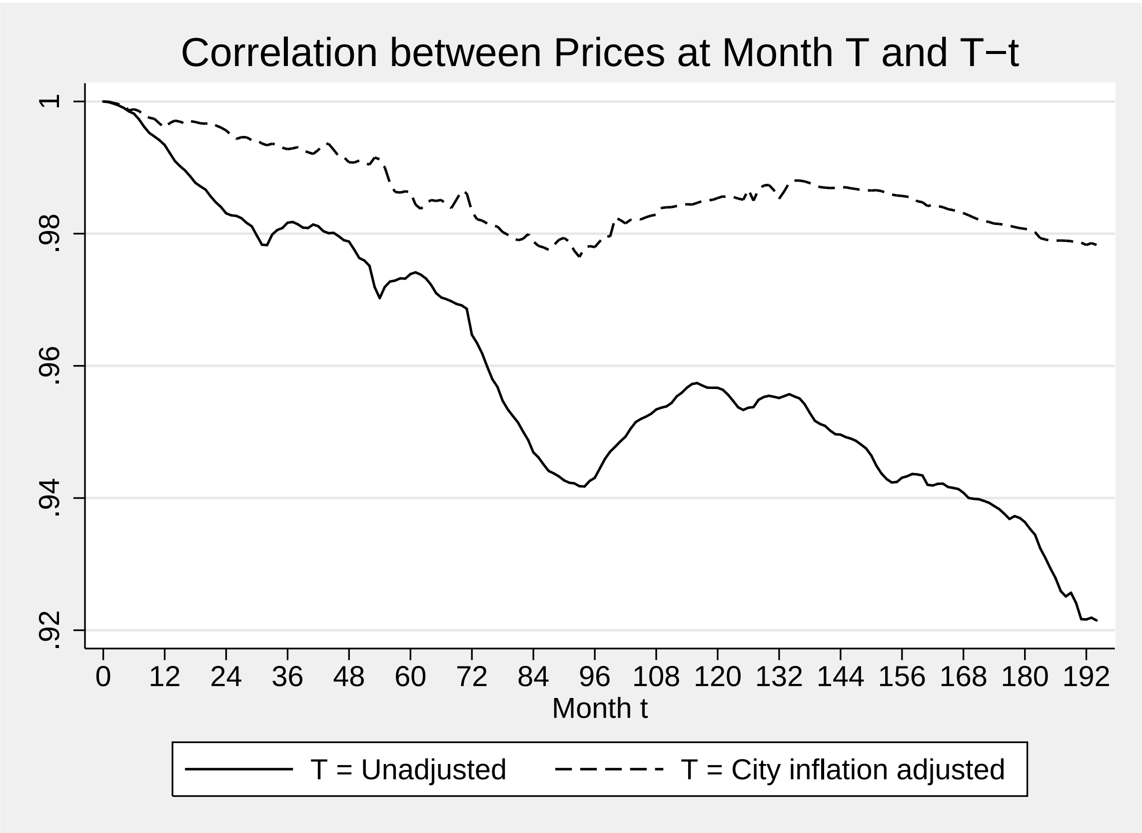


Fig. 12: *Auto-correlation of Zip-level Median Prices between 1997 and 2013*

Notes: The zip-level data are available here http://www.zillow.com/blog/research/data/.

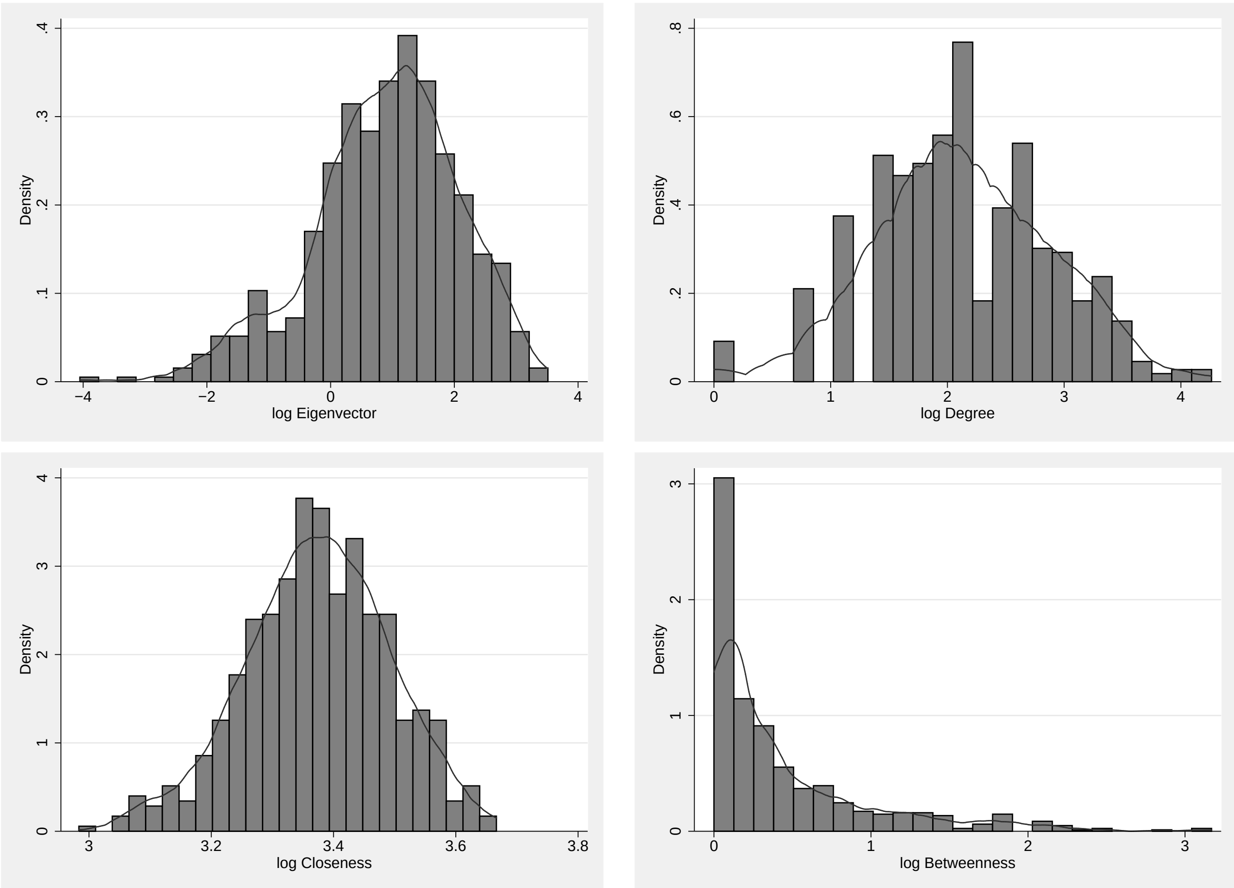


Fig. 13: *Densities of Log Centrality Indices*

Notes: The density shows the entire distribution.

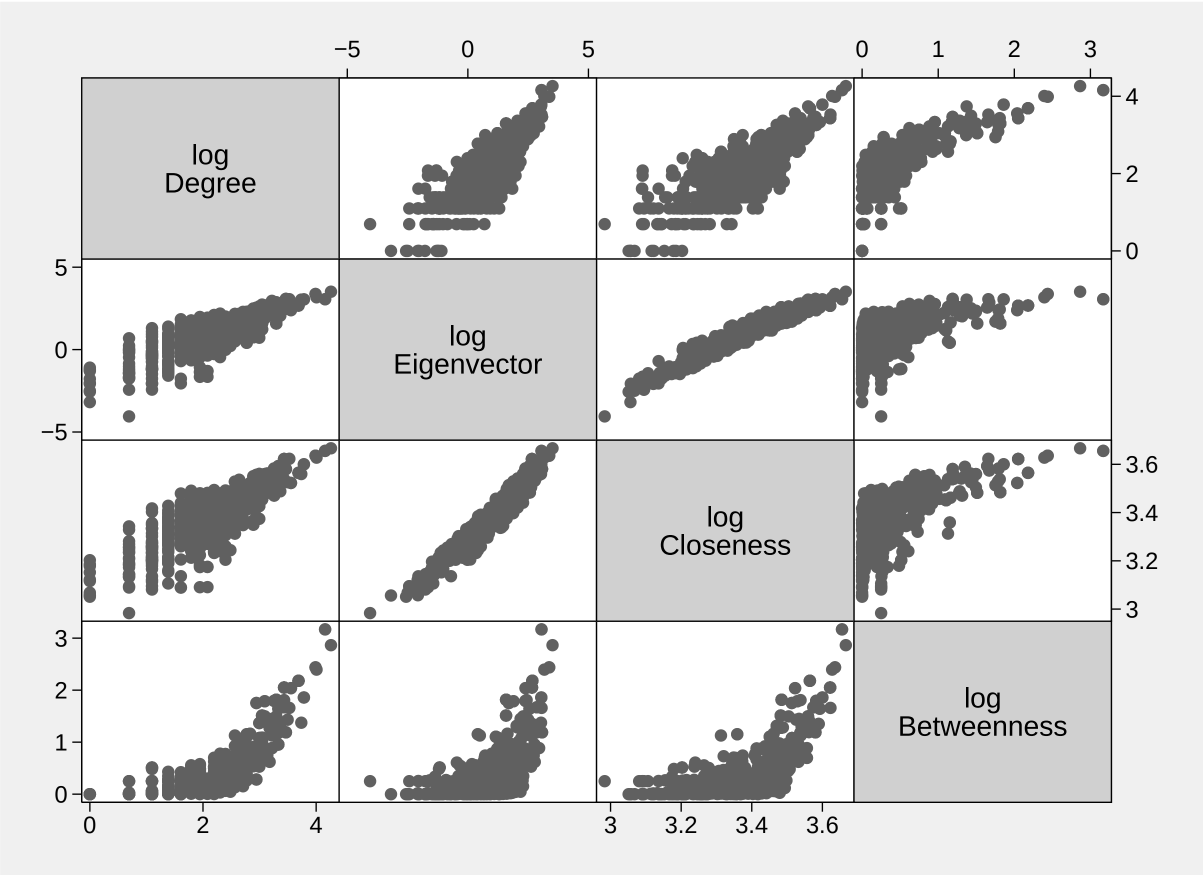


Fig. 14: *Relationship Between Log-centrality Indices*

Notes: The density shown excludes the top 1 percent of the distribution.

C. Online Appendix Tables

Table 9: *Testing for Heteroscedasticity: Squared Residual Regressions*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Squared residuals (× 100)							
	Closeness		Eigenvector		Degree		Betweenness	
	OLS	NLLS	OLS	NLLS	OLS	NLLS	OLS	NLLS
Average distance	-1.10%	-0.01%	-0.78%	-0.01%	-0.06%	0.00%	0.35%	0.00%
	(0.023)	(0.000)	(0.023)	(0.000)	(0.024)	(0.000)	(0.024)	(0.000)
Constant	132.1*	4.870***	123.7*	4.801***	108.3	4.683***	97.80	4.596***
	(62.569)	(0.618)	(63.007)	(0.618)	(65.183)	(0.611)	(65.587)	(0.609)
Observations	641	641	641	641	641	641	641	641

Notes: The residuals are based on Column 3 in Table 4 and Columns 2, 4, and 6 in Table 5. The non-linear least squares regressions (NLLS) model the squared residual as an exponential function of distance. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 10: *Place of Birth Heterogeneity*

	(1)	(2)
	log Housing value	
log Closeness	0.683	1.375
	(0.880)	(0.894)
× born outside Italy	1.079	0.872
	(0.985)	(0.975)
× born in Italy	1.361	1.500
	(1.094)	(1.126)
Other regressors		X
Observations	641	641
R-squared	0.070	0.134

Notes: All regressions control for the housing variables. The additional regressors ("Other Xs") are the same as in Table 4. Clustered (by 530 surnames) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 11: *Housing Value Regressions Without Weighting for the Sampling Design*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				log Housing value				
log Closeness	1.504***	2.299***						
	(0.328)	(0.403)						
log Eigenvector			0.152***	0.212***				
			(0.034)	(0.040)				
log Degree					0.042	0.136**		
					(0.054)	(0.067)		
log Betweenness							-0.049	0.039
							(0.070)	(0.092)
Other regressors		X		X		X		X
Observations	641	641	641	641	641	641	641	641
R-squared	0.046	0.132	0.046	0.128	0.020	0.093	0.020	0.087

Notes: All regressions control for the housing variables. The additional regressors ("Other Xs") are the same as in Table 4. Clustered (by 530 surnames) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 12: *Full First Stage Regressions: Centrality Measures and the Potential Innate Interactions Based on Surnames*

	(1) log Closeness	(2) log Eigenvector	(3) log Degree	(4) log Betweennes
Potential innate interactions (in %)	0.033**	0.268*	0.116	0.138***
	(0.015)	(0.158)	(0.094)	(0.053)
Zero potential innate interactions	-0.019	-0.308	-0.236	0.011
	(0.020)	(0.234)	(0.153)	(0.032)
Housing built after 1960	-0.033*	-0.416**	-0.279**	-0.056**
	(0.019)	(0.202)	(0.123)	(0.025)
Year built unknown	-0.023	-0.162	-0.074	-0.032
	(0.022)	(0.232)	(0.146)	(0.032)
Extended family members	0.031**	0.256*	0.292***	0.197***
	(0.014)	(0.148)	(0.087)	(0.032)
Born outside Italy (mainly U.S.)	-0.004	-0.032	0.013	-0.018
	(0.032)	(0.359)	(0.192)	(0.036)
Born in Sicily	0.019	0.244	0.079	0.041
	(0.031)	(0.346)	(0.199)	(0.039)
Age	0.006	0.043	0.004	0.013
	(0.009)	(0.103)	(0.059)	(0.017)
Age squared/100	-0.005	-0.033	0.003	-0.011
	(0.010)	(0.108)	(0.062)	(0.018)
Year of birth unknown	0.149	0.793	0.331	0.390
	(0.210)	(2.374)	(1.358)	(0.386)
Height in feet	0.070*	0.830*	0.275	0.042
	(0.037)	(0.425)	(0.214)	(0.058)
Weight in pounds	-0.000	-0.003	0.001	0.000
	(0.000)	(0.005)	(0.002)	(0.000)
Age at first arrest	-0.001	-0.016	0.000	0.000
	(0.001)	(0.011)	(0.006)	(0.001)
Never arrested	-0.051	-0.629*	-0.080	0.020
	(0.031)	(0.350)	(0.175)	(0.043)
Married	0.014	0.166	0.181	0.006
	(0.018)	(0.201)	(0.128)	(0.029)
Divorced	-0.027	-0.374	-0.180	0.007
	(0.029)	(0.338)	(0.217)	(0.046)
Number of children	0.001	-0.008	-0.041	0.004
	(0.006)	(0.061)	(0.033)	(0.007)
Siblings	0.005	0.044	0.030	0.009*
	(0.003)	(0.034)	(0.020)	(0.005)
Types of crime committed	-0.010*	-0.092	-0.029	-0.003
	(0.006)	(0.062)	(0.037)	(0.007)
Number of businesses	0.010*	0.118*	0.061*	0.024**
	(0.006)	(0.061)	(0.033)	(0.010)
Involved in drug dealing	-0.003	-0.121	-0.075	0.033
	(0.015)	(0.167)	(0.105)	(0.024)
Housing variables				
Observations	641	641	641	641
R-squared	0.160	0.146	0.205	0.326
Kleibergen-Paap rk Wald F statistic	3.479	2.663	2.343	12.60
F-stat on potential innate interactions	5.208	2.887	1.523	6.726

Notes: All regressions control for the housing variables. The additional regressors ("Other Xs") are the same as in Table 4. Clustered (by 530 surnames) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 13: *Full IV Regressions: Centrality Measures and the Potential Innate Interactions Based on Surnames*

	(1)	(2)	(3)	(4)
		log Housing value		
log Closeness	6.350**			
	(3.196)			
log Eigenvector		0.627*		
		(0.329)		
log Degree			0.977*	
			(0.537)	
log Betweenness				1.201
				(1.054)
Housing built after 1960	0.178	0.227	0.238	0.036
	(0.193)	(0.219)	(0.232)	(0.179)
Year built unknown	-0.193	-0.243	-0.276*	-0.292**
	(0.174)	(0.171)	(0.162)	(0.136)
Extended family members	-0.366**	-0.323***	-0.439**	-0.387
	(0.144)	(0.124)	(0.189)	(0.277)
Born outside Italy (mainly U.S.)	-0.056	-0.050	-0.077	-0.085
	(0.225)	(0.239)	(0.222)	(0.184)
Born in Sicily	-0.349	-0.381	-0.305	-0.276
	(0.244)	(0.264)	(0.246)	(0.200)
Age	-0.089	-0.077	-0.054	-0.063
	(0.079)	(0.079)	(0.083)	(0.082)
Age squared/100	0.086	0.076	0.054	0.066
	(0.086)	(0.088)	(0.091)	(0.088)
Year of birth unknown	-2.327	-1.900	-1.739	-1.822
	(1.725)	(1.685)	(1.822)	(1.829)
Height in feet	0.265	0.203	0.462	0.646**
	(0.388)	(0.445)	(0.383)	(0.320)
Weight in pounds	-0.001	-0.000	-0.003	-0.002
	(0.002)	(0.002)	(0.002)	(0.003)
Age at first arrest	-0.008	-0.007	-0.017**	-0.016**
	(0.008)	(0.009)	(0.008)	(0.008)
Never arrested	0.051	0.125	-0.187	-0.285
	(0.254)	(0.278)	(0.230)	(0.232)
Married	-0.111	-0.131	-0.210	-0.046
	(0.172)	(0.183)	(0.208)	(0.152)
Divorced	0.101	0.172	0.113	-0.119
	(0.287)	(0.325)	(0.347)	(0.233)
Number of children	-0.032	-0.023	0.012	-0.032
	(0.038)	(0.040)	(0.042)	(0.032)
Siblings	-0.002	0.002	0.000	0.017
	(0.033)	(0.034)	(0.037)	(0.028)
Types of crime committed	0.057	0.048	0.020	-0.002
	(0.052)	(0.049)	(0.046)	(0.040)
Number of businesses	0.009	0.001	0.016	0.038
	(0.063)	(0.066)	(0.065)	(0.062)
Involved in drug dealing	0.361***	0.421***	0.419***	0.297***
	(0.110)	(0.118)	(0.127)	(0.110)
Observations	641	641	641	641

Notes: The additional housing variables are the same as in Table 4. Clustered (by 530 surnames) standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.