Is Crime Contagious?
An Examination Of The Network Effects of Crime

Arturo Chavez, Robert Burklund, Dan Rigobon & Henry Aspegren

I. Background

In order to control and prevent crime, it is first necessary to understand it. Many studies have focused on the social and economic factors that are correlated with crime. However, comparatively few have focused on the underlying strategic dynamics of criminal activity itself. As Sean Connery explains in the classic film *The Untouchables*, "If they pull a knife, you pull a gun. He sends one of yours to the hospital, you send one of his to the morgue." Historical data shows that high crime rates are localized and idiosyncratic rather than solely determined by socio-economic factors. We seek to explain this using network effects. Our work is based on prior study by Papachristos [6] who has treated gun violence in Chicago as a social contagion. In this paper, we apply network analysis techniques to Chicago's recent crime history, extending beyond gun violence. As Papachristos mentioned in his paper, 48 Years of Crime in Chicago, "Chicago is, of course, a city of neighborhoods. And, as has been well documented elsewhere, crime rates are by no means equal across neighborhoods"[7]. The variation in crime across communities is the basis for our analysis. By quantifying network effects by crime category, we can characterize how different crime types propagate across the city. Understanding network effects in this context can help inform new policy and criminal enforcement strategies.

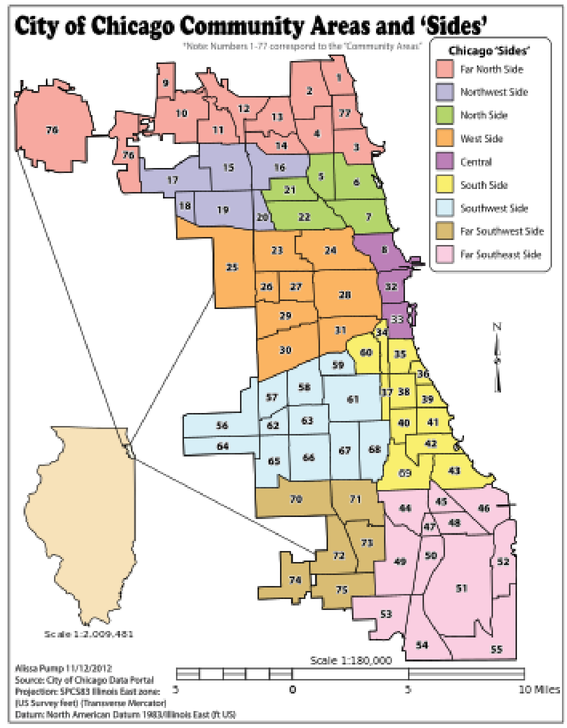

Fig. 1. Community Areas of Chicago

II. Methods

A. Model

To examine how an individual's choice to commit a crime changes the behavior of those around him we would ideally use the entire social network graph for the City of Chicago. This method of analysis is used throughout the literature to study social networks at smaller scale for which it is feasible to obtain the complete network structure (for example teenage delinquency in high schools [citation needed]). However no such dataset with the required scope is publicly available for the entire city. Instead we use a first order approximation and consider groups of similar people interacting with other groups of similar people. To model this we break the City of Chicago into culturally similar neighborhoods called **Community Areas**.

These divisions were created by University of Chicago social scientists and have remained static over the last several decades. Community areas are a useful unit of aggregation because a wealth of census and socio-economic data is available for them. From the map of Community Areas we derive a network representation of Chicago where each Community Area is a node and edges exist only between geographically adjacent community areas. Here we are making the assumption that any network effects, if they exist, will be most pronounced between adjacent areas. This simplification is consistent with a model of social interaction in which the bulk of connections are between people located geographically close to one another. While this is not universally true it is a reasonable first order approximation - we would expect the aggregate level of connection between individuals in adjacent community areas to be higher.

In order to model network effects in crime we used the extended linear-in-means model analyzed in [1]. This model contains fixed effects and both endogenous and exogenous network effects of neighboring communities.

$$Y_i = \alpha + \beta G X_i + \gamma X_i + \delta G Y_i + \epsilon_i$$

Where Y_i is the level of crime in community area i, G is the adjacency matrix of the graph of community areas, X_i is a matrix of socio-economic and census controls. The vector of parameters γ captures the fixed effects of socio-economic differences on crimes (e.g. poor areas may be more predisposed to higher levels of theft). The single parameter δ

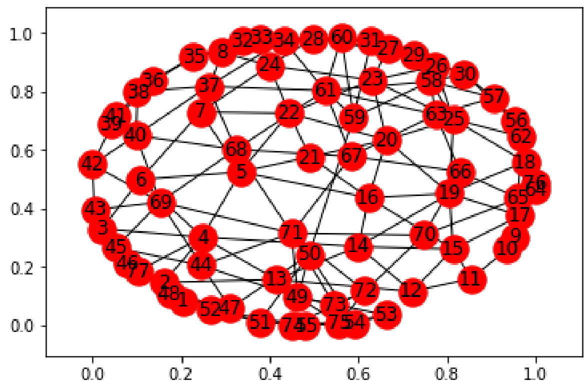

Fig. 2. Representation of the Community Areas of Chicago as a network

captures the *endogenous effects* of crime - that is the impact of crime in one area on the level of crime in the adjacent areas. The vector of parameters β captures the *endogenous or contextual effects* - that it the effect of the socio-economic factors of one district on the level of crime in its neighbors. Note that in this model each community area only affects the community areas adjacent to it.

B. Data

To get crime statistics we used a comprehensive dataset from the City of Chicago that contains all crimes committed between 2001 and 2017 [2]. This dataset contains the date, type, location and other information about every crime committed in Chicago. The location of the crime is accurate to the nearest block (the location on the block is randomized to provide a degree on anonymity). During this period there were a total of 1,456,714 crimes committed. To get the crime level Y_i for each community area we count the number of crimes committed in community area i over this four year period. We calculate these counts for each type of crime as classified under the Illinois Uniform Crime Reporting Codes [3]. This resolves problems of aggregation across different kinds of crimes.

The dataset that we used had several issues with data corruption. All of the years 2005-2011 were corrupted. Because of this we will be focusing mostly on the dataset from 2012-2017. With some work we were able to recover usable data from the years 2005-2007, however results obtained from this data should be treated as preliminary. Sadly, despite our best efforts we were unable to recover the data from 2008-2011.

For socio-economic and demographic data we used the closest census results from 2010 summarized to the levels of community areas [4] [5]. This includes population and demographic information as well as economic indicators such as education, and poverty levels. In total there are 74 different control variables (see appendix for list). Because of the number of parameters as well as their collinearity we cannot use all of the control variables. Instead we use two different approaches. In our first approach we look at which control variables that intuitively could affect the crime level in community areas, and then eliminated those that had the least statistical significance to use the top 10 most powerful predictors. In the second approach we run Principal Component Analysis (PCA) on X, the matrix of control variables. To avoid problems related to absolute magnitude we normalize each control variable to sum to 1. Then we use the top 10 principal components which explain the 96.809% of the variance of the control variables. This allows us to capture more of the variance of the control data without having to expand the number of parameters used in estimation and eliminates any potential collinearity. The drawback to this approach is that we lose the ability to interpret the meaning of the control variables.

C. Estimation

To obtain an unbiased estimate for the parameters in the model we use an Instrumental Variables (IV) approach with G^2X and G^2Y as instruments. A good discussion of the value of IV regression in situations such as ours is contained in [8]. This approach correctly identifies the network effects only under the exclusion restriction that the only way a community area affects a community area located at a distance of two edges away is through its affects on the intermediate community area that is a distance of one edge away. In other words communities can only affect one another via their affects on adjacent communities. As argued previously we believe this is a reasonable first order approximation to make. If this exclusion restriction holds then the estimate will be correct [1]. For our estimation to fail there would need to be uncontrolled effects between non-adjacent community areas that were greater or equal in magnitude to those between directly adjacent areas. If these correlations are not systematic across the City then this would simply introduce noise into our estimate and reduce our statistical power. Furthermore to use G^2X and G^2Y as instruments we have the additional requirement that G^2 and G are not co-linear. This clearly holds for our graph of Chicago since there are many connected triples of community areas that are not fully connected. The code for this project is available on Github [9].

III. RESULTS

After analyzing the dataset we found that ten of the thirty categories of crimes have statistically significant $(p < 0.05)$ network effects. The full table of GY coeffecients with PCA fixed effects is in Table V and the GY coefficients with standard fixed effects is in Table VI. Due to the much higher level of variance explained by the PCA fixed effects we chose to focus more on the results obtained from this analysis. As compared to the standard fixed effects we observed generally smaller, but still significant, network effects. There were howerver exceptions to this rule.

While our initial motivation for exploring this dataset was centered on analyzing gun violence in Chicago, we in fact found the most robust network effects in narcotics crime. However, preliminary results show that time series data would be useful in analyzing homicide.

A. Narcotics

The GY coefficient resulting from network effects of Narcotics offenses proved to be quite significant and robust. Choosing either PCA components or Community Area information rendered almost identical values for the network effect - with equal significance.

We choose to focus on the output of the regression with PCA fixed effects as the magnitude of the coefficients is implicitly more significant when more of the variance is fixed by controls. Looking at this coefficient across datasets pertaining to different time periods led to very statistically significant values $(p < 0.01)$ that were similar in magnitude. Table I contains the regression coefficients and relevant statistics for Narcotics offenses across time periods for which data was avaialable.

TABLE I
NETWORK EFFECTS IN NARCOTICS FOR VARIOUS TIME PERIODS (PCA CONTROLS)

Time Period	Coefficient (GY)	$\Pr(>\|t\|)$
2001-2004	0.20852	$9.33*10^{-7}$
2005-2007	0.18973	$2.11*10^{-5}$
2012-2017	0.26813	$1.46*10^{-12}$

B. Obscenity

Combining the results from our two sets of control variables, we find several crimes showing insignificant network effects. Examples of such crime types include: arson, domestic homicide, deceptive practice, intimidation, kidnapping, obscenity, offenses involving children, public peace violation, sex offense, and weapons violation. These crime types all had low GY coefficients that were also statistically insignificant.

One crime type that showed consistently low network effects was obscenity. Obcentity crime rate was not identified to be driven by network effects, under both sets of control variables. The crime types previously mentioned show comparable results.

TABLE II
OBSCENITY ENDOGENOUS NETWORK EFFECTS 2012-2017

Fixed Effects	Coefficient (GY)	$\Pr(>\|t\|)$
Standard	-0.03031	0.6754
PCA	-0.0555	0.238

C. Homicide

For Homicides there are two distinct types - domestic and non-domestic. During the period from 2012-2017 there were a total of 2649 homicides, of which 172 were classified as domestic. We observe that domestic homicides don't show any network effect but non-domestic homicides show a positive but weakly significant network effect. The effect is significant with standard controls but is not significant with PCA controls. This is due in part to the small number of homicides relative to the large number of Community Areas.

TABLE III
HOMICIDE ENDOGENOUS NETWORK EFFECTS 2012-2017

Type	Fixed Effects	Coefficient (GY)	$\Pr(>\|t\|)$
Domestic	Standard	-0.07549	0.2363
Domestic	PCA	-0.0940435	0.100278
Non-Domestic	Standard	0.1368	0.035**
Non-Domestic	PCA	0.04450	0.225740

We also observed significant contextual effects for non-domestic homicides even with PCA controls.

TABLE IV
HOMICIDE EXOGENOUS NETWORK EFFECTS 2012-2017

Control	Coefficient (GX)	$\Pr(>\|t\|)$
PC1	-0.09415	0.640959
PC2	0.28174	0.540146
PC3	-0.21805	0.619780
PC4	0.58711	0.493960
PC5	2.35211	0.008971**
PC6	-0.15850	0.825557
PC7	-2.63338	0.004274**
PC8	0.55648	0.746466
PC9	-0.02956	0.985275
PC10	0.37903	0.790497

IV. DISCUSSION

Overall our results are surprising but easy to interpret. For most types of crime there is a large and statistically significant network effect - even after controlling for a community's underlying predisposition to commit crime. Unsurprisingly most endogenous network effects are positive in sign indicating that those crime types are strategic complements. In particular Gambling, Motor Vehicle Theft, Narcotics and Prostitution all have very large $(> 10\%)$ and positive network effects. The organized and inherently social nature of these kinds of crimes could be producing this result. In the context of the rise of gang-related activity during this period this explanation seems reasonable. The fact that Narcotics, something with an already well documented network effect had the largest absolute effect further validates our findings. In particular, the network effects of over 20% and p-value of approximately 10^{-12} for narcotics surprised the authors. This suggests that a crime prevention strategy tailored to preventing network effects would be helpful in fighting narcotics crime.

Also significant is the conspicuous absence of network effects for certain types of crime. In particular Obscenity, Arson and Sex Offenses exhibit no network effects with any controls. This is an intuitive result when we consider the nature of these types of crimes. Ex-ante there is no reason why any of these offenses should have strategic effects - a person's choice to act obscene is not obviously related to anyone else's choice. Furthermore there is no gang structure or social organization that could facilitate network effects. These crimes are also not well predicted by any of our socio-economic controls. Instead this type of crime is consistent with a model without network effects in which a certain percentage of the population will commit these crimes regardless of the actions of others.

The homicide result is particularly interesting. While domestic homicides don't have a network effect the non-domestic homicides exhibit both exogenous and endogenous network effects. The endogenous effect is positive indicating that murder is a strategic complement. This result is not significant under PCA controls but is also positive in sign. Overall this suggests that there is some positive network effect for non-domestic murders. This makes sense given more context. Murders in Chicago overwhelmingly involve young black men killing other young black men. In most cases at least one and frequently both of the parties have a previous arrest. It is reasonable to conservatively assume that at least half of all murders in Chicago are due to gang violence [10]. This figure could be as high as 80% or 90%. There may be a large strategic component to gang violence. As Sean Connery's quote suggests - a gang killing can plausibly prompt direct retribution. We have demonstrated the existence of such an effect in Chicago. However our statistical power is limited, in part because murder is localized at more granular level than the Community Area.

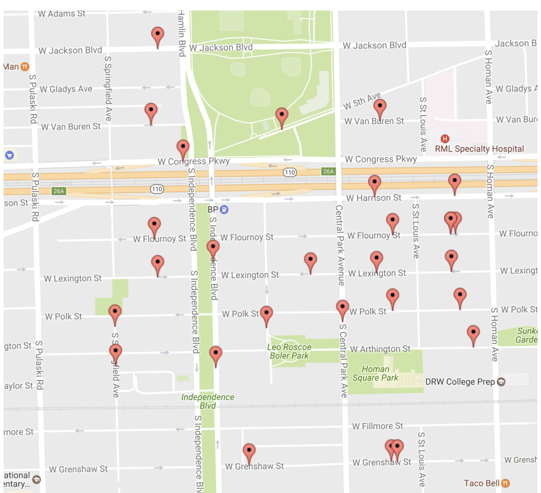

Fig. 3. Murders in Police Beat 1133 From 2012-2017

A smaller unit of analysis or a larger dataset may resolve this problem.

V. CONCLUSIONS & NEXT STEPS

These quantitative results have interesting implications for criminal enforcement. In particular we note that enforcement strategies should differ for crime types with large network effects. For these types of crimes strong enforcement in one Community Area exerts a positive externality on all the other Community Areas. In other words reducing crime in one area has a spillover effect to the rest of the city. While the social causes of crime may be poorly understood we have shown that crime itself is a cause of more crime. This means that a targeted intervention (in any form) in one area, has the potential to reduce crime across the entire city. Conversely, for crimes with no network effects, investment in strong enforcement in one area will only yields local results.

Due to issues with data corruption we were unable to do any more sophisticated analysis involving time series data, but with better data there are several interesting analyses that could be performed. Currently our model looks at time-independent network effects based on physical location. A modification of this model that would be interesting to explore is one in which each node in the graph is a pair (Community Area, year). Then the graph could be directed with each node effecting itself and its neighbors in the future as well as its neighbors now. Furthermore, we only used the adjacency matrix for the graph of Community Areas, whereas a more in depth analysis could first use some other dataset to refine this to a weighted graph of how closely the Community Areas are connected.

Another possible extension of this work would be a similar analysis on another city, besides Chicago. Our central assumptions regarding the cohesiveness of community districts and their correlation to adjacent districts are generalizable to other cities. Localization of crime in specific areas is also common throughout the US. Cities like Los Angeles, Detroit, or Philadelphia appear to be good candidates.

APPENDIX

A. Full Results

TABLE V

NETWORK EFFECTS OF CRIME TYPES IN ADJACENT COMMUNITY AREAS

	IV Regression Coefficients - With Principal Component Controls			
Crime Type	GY Coefficient	Std. Error	t-Value	$\Pr(>\|t\|)$
Arson	-0.0078	0.0327	-0.239	0.812
Assault	0.0419	0.0235	1.784	0.0799 *
Battery	0.0455	0.0259	1.75	0.0866 *
Burglary	0.0758	0.0328	2.307	0.0248 **
Concealed Carry Violation	-0.2017	0.0793	-2.542	0.0139 **
Sexual Assault	0.0263	0.0241	1.091	0.280
Criminal Damage	0.0406	0.0186	2.183	0.0334 **
Criminal Trespass	0.0629	0.0303	2.080	0.0422 **
Deceptive Practice	0.009781	0.0413	0.237	0.814
Gambling	0.112	0.0478	2.344	0.0227 **
Homicide (Non-Domestic)	0.04450	0.03632	1.225	0.225740
Homicide (Domestic)	-0.0940435	0.0562585	-1.672	0.100278
Human Trafficking	-0.155	0.0574	-2.703	0.0091 ***
Interference with Public Officer	0.0854	0.0449	1.901	0.0625 *
Intimidation	-0.0272	0.053	-0.514	0.609
Kidnapping	-0.0109	0.0329	-0.332	0.741
Liquor Law Violation	0.0591	0.0469	1.259	0.2133
Motor Vehicle Theft	0.08358	0.0221	3.784	0.00038 ***
Narcotics	0.268	0.0295	9.101	$1.46*10^{-12}$ ***
Non-Criminal	-0.191	0.0769	-2.484	0.0161 **
Obscenity	-0.0555	0.0465	-1.193	0.238
Offenses Involving Children	0.0398	0.0284	1.398	0.1678
Prostitution	0.1565	0.0557	2.808	0.00689 ***
Public Indecency	-0.12026	0.07298	-1.648	0.1051
Public Peace Violation	0.03161	0.03824	0.827	0.4121
Robbery	0.05596	0.02895	1.933	0.0584 *
Sex Offense	0.007983	0.02205	0.362	0.7187
Stalking	0.08415	0.0329	2.558	0.0133 **
Theft	0.0416	0.0315	1.319	0.19267
Weapons Violation	0.03037	0.03237	0.938	0.352

TABLE VI

NETWORK EFFECTS OF CRIME TYPES IN ADJACENT COMMUNITY AREAS

	IV Regression Coefficients - With Standard Fixed Effects			
Crime Type	GY Coefficient	Std. Error	t-Value	$\Pr(>\|t\|)$
Arson	0.04624	0.06815	0.678	0.5003
Assault	0.2494	0.0544	4.585	$2.66*10^{-5}$ ***
Battery	0.2221	0.04807	4.62	$2.36*10^{-5}$ ***
Burglary	0.1406	0.05783	2.431	0.0183 *
Concealed Carry Violation	-0.131	0.07901	-1.658	0.103
Sexual Assault	0.1827	0.05802	3.148	0.00265 ***
Criminal Damage	0.1764	0.05384	3.276	0.001825 ***
Criminal Trespass	0.2436	0.06433	3.786	0.00038 ***
Deceptive Practice	0.1238	0.00788	1.571	0.1219
Gambling	0.0817	0.0494	1.655	0.1036
Homicide (Non-Domestic)	0.1368	0.06326	2.162	0.035 **
Homicide (Domestic)	-0.07549	0.06305	-1.197	0.2363
Human Trafficking	-0.3364	0.04855	-6.93	$4.92*10^{-9}$ ***
Interference with Public Officer	0.2362	0.0571	4.137	0.000122 ***
Intimidation	-0.07878	0.08024	-0.982	0.3305
Kidnapping	-0.09647	0.07721	-1.249	0.2168
Liquor Law Violation	0.1373	0.06563	2.092	0.04107 **
Motor Vehicle Theft	0.1662	0.04273	3.889	0.000273 ***
Narcotics	0.2736	0.02603	10.51	$9.17*10^{-15}$ ***
Non-Criminal	-0.2029	0.08308	-2.442	0.01783 **
Obscenity	-0.03031	-0.07201	-0.421	0.6754
Offenses Involving Children	0.05227	0.06689	0.781	0.4379
Prostitution	0.1566	0.05801	2.7	0.0092 ***
Public Indecency	-0.2429	0.08519	-2.851	0.00613 ***
Public Peace Violation	0.07576	0.06767	1.119	0.2678
Robbery	0.1339	0.05364	2.496	0.0156 **
Sex Offense	0.071	0.0652	1.089	0.281
Stalking	0.1378	0.06044	2.28	0.0265 **
Theft	0.1276	0.06552	1.947	0.0566 *
Weapons Violation	0.04081	0.06137	0.665	0.5089

VI. Fixed Effects Full List

A. All Fixed Effects

1) Total Population.
2) Population by Race: White alone, Black or African American alone, American Indian and Alaska Native alone, Asian alone, Native Hawaiian and Other Pacific Islander alone, Some Other Race alone, Two or More Races, Hispanic or Latino.
3) Population by Gender and Age (with the following bins): Under 5 years, 5 to 9 years, 10 to 14 years, 15 to 17 years, 18 and 19 years, 20 years, 21 years, 22 to 24 years, 25 to 29 years, 30 to 34 years, 35 to 39 years, 40 to 44 years, 45 to 49 years, 50 to 54 years, 55 to 59 years, 60 and 61 years, 62 to 64 years, 65 and 66 years, 67 to 69 years, 70 to 74 years, 75 to 79 years, 80 to 84 years, 85 years and over.
4) Median Age, Percent Aged 16+ Unemployed, Percent Aged 25+ Without High School Dimploma, Percent Aged Under 18 or Over 64.
5) Total Households, Average Household Size, Total Housing Units, Occupied Housing Units, Vacant Housing Units, Occupied Housing Units Owned with a mortgage or a loan, Owned free and clear, Renter occupied, Percent of Housing Crowded, Percent of Households Below Poverty Line.
6) Per Capita Income, Hardship Index.

B. Standard Fixed Effects (Used For Regression)

White alone, Asian alone, Black or African American, Two or More Races, Hispanic or Latino, Median Age, Total Households, Percent Household Below Poverty Line, Percent Aged 25+ Without High School Dimploma and Per Capita Income.

References

[1] Yann Bramoull, Habiba Djebbari, Bernard Fortin. *Identification of peer effects through social networks*. Journal of Econometrics, Volume 150, Issue 1, May 2009, Pages 41–55, ISSN 0304-4076, https://doi.org/10.1016/j.jeconom.2008.12.021. (http://www.sciencedirect.com/science/article/pii/S0304407609000335) Keywords: Social network; Peer effects; Identification; Add Health dataset

[2] Chicago Police Department. *Crimes 2001 to present*. https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2

[3] Chicago Police Department. *Illinois Uniform Crime Reporting Codes*. https://data.cityofchicago.org/Public-Safety/Chicago-Police-Department-Illinois-Uniform-Crime-R/c7ck-438e/data

[4] Chicago Public Health. *Census Data - Selected socioeconomic indicators in Chicago, 2008 2012*. https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2

[5] *Spreadsheet: 2010 Census data*. https://datahub.cmap.illinois.gov/dataset/2010-census-data-summarized-to-chicago-community-areas/resource/b30b47bf-bb0d-46b6-853b-47270fb7f626

[6] Green B, Horel T, Papachristos AV. *Modeling Contagion Through Social Networks to Explain and Predict Gunshot Violence in Chicago*, 2006 to 2014.JAMA Intern Med.2017;177(3):326-333. doi:10.1001/jamainternmed.2016.8245

[7] Andrew V. Papachristos. (2013). *48 Years of Crime in Chicago: A Descriptive Analysis of Serious Crime Trends from 1965 to 2013*, ISPS Working Paper, ISPS13-023.

[8] Joshua D. Angrist, Alan B. Krueger. (2001). *Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments*. Journal of Economic Perspectives, Vol. 15, Number 4, 69–85.

[9] https://github.com/henryaspegren/6.207-project

[10] University of Chicago Crime Lab. (January 2016). *Gun Violence in Chicago, 2016*.