The Spatial Structure of Crime in Urban Environments

Sarah White[1], Tobin Yehle[2], Hugo Serrano[3], Marcos Oliveira[3], and Ronaldo Menezes[3]
sarahw14@live.unc.edu,tobin.yehle@utah.edu
hbarbosafilh2011@my.fit.edu,moliveirajun2013@my.fit.edu
rmenezes@cs.fit.edu

[1] College of Arts and Sciences, University of North Carolina at Chapel Hill, USA
[2]
[3] BioComplex Laboratory, Florida Institute of Technology, USA

Abstract. It is undoubtedly *cliché* to say that we are in the Age of Big Data Analytics or Data Science; every computing and IT publication you find talks about Big Data and companies no longer are interested in software engineers and analysts but instead they are looking for Data Scientists! In spite of the excessive use of the term, the truth of the matter is that data has never been more available and the increase in computation power allows for more sophisticated tools to identify patterns in the data and on the networks that governs these systems (complex networks). Crime is not different, the open data phenomena has spread to thousand of cities in the world, which are making data about crime activity available for any citizen to look at. Furthermore, new criminology studies argue that criminals typically commit crimes in areas in which they are familiar, usually close to home. Using this information we propose a new model based on networks to build links between crimes in close physical proximity. We show that the structure of the criminal activity can be partially represented by this spatial network of sites. In this paper we describe this process and the analysis of the networks we have constructed to find patterns in the underlying structure of criminal activity.

Keywords: network science, crime mapping, social disorganization, routine activity

1 Introduction

The understanding of crimes, the organizational process in criminal activities, and the emergence of crime spots in certain regions of a city can lead to the creation of better tools to enable more effective law enforcement, culminating in safe cities. The increasing amount of criminal data available nowadays may be used to guide this understanding of crime. For instance, information about crime is usually accompanied by location metadata indicating where the criminal activity occurred. This type of spatial data has been used to understand delinquent behavior back in the 1920s [1,2], and many theories have been proposed to explain the existence of areas of concentrated crime, referred as hotspots [3,4,5,6,7].

These hotspots are related to the fact that criminal activities do not occur uniformly across a region [8,9]. The theories of social disorganization, social control and collective efficacy explain this phenomenon by means of the residents of the area considered [6,7]. These studies recognize the active role of the environment on the criminal activities. For example, many aspects of the community lead to the inability of the neighborhood to publicly control the behavior of people, and thus to the increase of the likelihood of crime [10]. Moreover, prior work shows that crimes happen close to each other are carried out by people with familiar with similar geographical areas [11].

The hotspot analysis is an approach used to find regions with these aforementioned issues [8] and has been one of the main artifacts in the analyses of crime incidence [12]. In some sense, hotspot analysis is similar to a histogram because it allows to depict the crime frequency (or similar measure) in sub-regions of a certain region. Thus, any information about the relation between crimes is not present in this analysis—any criminal structure that could be underlying the criminal activity is not included.

In order to capture this structure, we propose to analyze crime activity using a network framework. The rationale here is that criminal activities rely heavily on different types of networks, such as a social network within gangs and the word-of-mouth flow of information regarding illegal market [13,14,15]. The presence of these relations suggests that network science can be a powerful tool for analyzing criminal activity.

In this paper, we define a crime network in which crimes are connected if they occur nearby. We used data of crime incidents from US police department records to generate complex networks of criminal activity in Los Angeles, CA, and Miami, FL and we found borders between communities in these structures using an approach adapted from Thiemann et al. [16]. Finally, in order to show that these networks capture real-world phenomena, we compared these borders to similar boundaries in demographic data. We found statistically significant variations that indicates the structure of the crime networks reflects real-world phenomena that cannot be accounted for by demographic differences.

2 Background

2.1 Criminology

There exist two theories in criminology that are of interest to us: the *routine activity* and *social disorganization* theories. The former argues that criminal activity occurs at the convergence of three things: a potential offender, a lack of guardianship or supervision, and a target [17]. The latter contends that criminal activity is the result of the social and physical environments of the neighborhood at hand [18]. Both theories seek to model crime phenomena using spatial and geographical context. For example, elevated crime rates are expected in a neighborhood that lacks a strong social community and access to resources.

The routine activity theory asserts that crime is a convergence between criminal opportunity and a potential offender, where this convergence serves as a point in time and space. The opportunistic nature of this theory means that criminal activity typically occurs in the sphere of familiarity of the criminal. Despite this sphere of familiarity being peculiar to the individual, areas of high traffic, such as downtown areas, lie within the sphere of familiarity of many individuals. This aspect is also related to the fact that criminals typically commit crimes within a short distance from their home [11].

Metropolitan areas are typically organized by regions of different land use through the natural development of urban centers and unnaturally through zoning laws. These different land uses include: residential use, commercial use, and industrial use. The presence of types of crimes differs between these land uses; neighborhoods with residential housing and no commercial businesses are perceived as *safe* and non-residential land uses are correlated with an increase in criminal activity [19]. Non-residential areas is typically found to have higher traffic in comparison to residential areas, consequently they witness to more crime. Also, the places within these areas, such as shopping centers or public parks, coincide with an increase in *foreign* or non-residential presence. The presence of such *strangers* negatively impacts a neighborhood's social structure.

Street networks influence human mobility and thus the potential offenders mobility. In fact, the convergence of a potential offender with a criminal opportunity is much more likely to occur and be exploited on a street that is relatively accessible and frequently traveled [17]. Street networks are not only correlated with an increase in crime incidence but additionally have a relationship with the typical *journey-to-crime* length of an offender [11]. Roadways and public transportation link together different areas of a criminal's sphere of familiarity and facilitate travel outside of a criminal's immediate neighborhood. The type of crime can affect the *journey-to-crime* length. For example, violent crime trips are shorter in length than property crime trips [11].

Both of aforementioned theories serve to explain crimes in terms of the spatial context of the neighborhood. By the social disorganization theory, a break down in the social structure of a neighborhood causes the elevated crimes rates. On the other hand, the routine activity theory implies that the higher traffic associated with non-residential land use is the cause of such crime rates.

2.2 Network Science

In order to find the areas of crimes, we used the algorithm *label propagation* for community detection applied to the network of crimes we created [20]. The stochastic output of the method is essential to our analysis of crime borders, because the borders have strength based on the frequency of the community pertinence [16]. Label propagation finds communities by assigning community labels to vertices matching the most common label of their neighbors. The procedure starts by assigning an unique label to each vertex. In each iteration the label of each vertex is assigned to be the most common label of its neighbors in the previous iteration, breaking ties randomly. If the edges have weights it assigns the label connected by the highest weight. The label propagation algorithm has complexity of $O(m)$, where m is the number of edges in the graph [20].

Therefore, this algorithm is both fast and stochastic which makes it a perfect choice for our chosen method of producing borders between communities.

3 Identifying Borders of Crime

3.1 The Dataset

In order to create the networks and their borders, a collection of police department records in the USA that spans the time frame 2007 through 2010 was gathered from *SpotCrime*[4]. Each crime event in the dataset is characterized by description, address, geotag, type, date, and time. The types of crimes are arrest, assault, vandalism, burglary, theft, robbery, shooting, and other. Using the geotag of each crime we derived the ZCTA (ZIP Code Tabulation Area) it belonged to. We used the data from Los Angeles, CA, and Miami, FL to construct our model because they are large metropolises, geographically different, culturally diverse, and have different demographics.

To compare the crime borders against socio-economic and demographic borders inherent to the metropolises considered, we clustered ZCTA of the respective areas based on features extracted from US Census data of the American Community Survey for 2007-2011 (the time period most similar to the time span of the crime records). Then, these clusters are compared with the crime borders. The features used to cluster were: percent of population with a high school degree, joblessness, poverty rate, median income, percent of population receiving public assistance, percent of households that have moved in the past year, percent of properties that are vacant, percent of renter-occupied households, and percent of female-headed households. These variables are based on the work of Willits et al., which shows that they accounted for over 70% of the variance between neighborhoods at the block level [18]. Finally, these clusters are reduced to borders in order to compare with the crime borders. Thus, a border of this kind exists between two ZCTA if they are geographically adjacent and are not in the same cluster.

3.2 Building Networks and the Borders

The networks analyzed in this paper have each vertex representing locations and these vertices are connected if the crimes associated with the vertices occurred within a certain distance. This method of linking crimes is based on previous findings that criminals generally act in a small area, and the crimes occurring near to other crimes are committed by the same, or similar people [11].

The choice of the location information that each vertex represents also defines the resolution of the network. For example, if the vertices represent the ZCTA of the location, the structure within this place is lost. However, this approach leads to reduction of computational cost of analysis, as well as to simplification of the results. In order to create this structure, a network where each vertex represents a

[4] `http://www.spotcrime.com`

single crime is built, then the crimes and their edges are collapsed into a network with fewer connections. Thus, the edges between crimes in different regions are represented by a single edge with weight equal to the number of original edges. On the other hand, the edges between crimes in a single region are treated as self edges and are removed.

The borders between vertices in the networks are analyzed by assigning to each vertex an associated map area. The vertices in the ZCTA-level networks contain all crimes in a single ZCTA, thus the shape of the ZCTA that each vertex represents is the physical region for the vertex. For the crime-level networks we used Voronoi maps. The region associated with each crime is the Voronoi cell surrounding that crime. Once the vertices in a network have an associated physical area, any communities in the network are a union of the areas of each node belonging to the community. Since each node belongs to a single community, the areas of all the communities completely cover the area of the network.

The analysis of the borders are not only concerned with their location, but the borders' strength is also a subject of study. The use of a stochastic community detection algorithm allows us to measure the probability of a border occurring as well as its physical position. The borders for many runs of a community detection algorithm are overlaid, and overlapping borders are combined. The weight w_{ij} of the border between two adjacent regions i and j is defined as:

$$w_{ij} := \frac{1}{R} \sum_{r=1}^{R} \delta(c_i^r, c_j^r) \tag{1}$$

where R is the number of runs of the community detection algorithm, c_i^r is the community of region i in the rth run of community detection, and $\delta(a, b)$ is 1 if a and b are different community labels, and 0 otherwise. This results in the weight of each border being the number of times the regions it divides appear in different communities normalized to a maximum value of 1.

3.3 Comparing Borders

The comparison between different sets of borders for the same region is made with the *absolute cross correlation* of a network representation of the borders [16]. A set of borders can be embedded in a network by representing the physical regions as vertices, and the borders between them as edges connecting adjacent regions. These edges have weights representing the strength of the borders between the regions. The absolute cross correlation is the normalized scalar product of the weights of edges in two networks. For two border networks b and b', the cross correlation is defined as:

$$c(b, b') := \frac{\sum_{e \in E} b(e) b'(e)}{\sqrt{\sum_{e \in E} b(e)^2} \sqrt{\sum_{e \in E} b'(e)^2}} \tag{2}$$

where $b(e)$ is the weight of edge e in b, and E is the set of edges in the two border networks. Equation 2 leads to high values when borders overlaps and borders not overlapping produce low values. However, different weight distributions of networks produce vastly different cross correlations, thus this evaluation by itself does not give meaningful results [16]. To compare the cross correlation between network, the comparison is made with a random null model in such way that a z-score is used to compare the cross correlations of different sets of networks.

However, the border network represents associations between adjacent regions, thus adding or removing edges from the graph in order to randomize it is not possible. Furthermore, cross correlation requires two networks with the same set of edges. For this reason, the weights of the borders are redistributed to find a random network using the iterative border redrawing method used by [16]. This method is a random process which iteratively redraws borders until a sufficiently random set of borders is achieved.

4 Results

The spatial distribution of different types of crimes is strongly influenced by factors such as education, income levels, family structure and organization of public spaces [18]. In this section we investigate whether the network structures that support the criminal activities also influence the spatial distribution of crimes and how criminals move in space. In other words, we compare the structure of the borders generated from crime networks with borders of areas clustered by their socio-demographic characteristics. The intuitive notion here is that the more influential networks used by criminals from different areas (e.g. transportation and social networks) the stronger the relationship between these areas will be in the crime network that we generate. Conversely, the smaller the influence of these networks, the weaker the connection between the network crime areas in and hence the stronger the borders between these areas.

This intuition is a plausible assumption once the distribution of distances traveled by offender from their places of residence to crime locations is characterized by the predominance of short trips with sporadic long trips. This distances vary with many factors such as the type of crime, gender and age of the offender, but the average trip length is approximately 1.6 miles with 84% of trips being shorter than 3.1 miles and only 7 % of the trips were longer than five miles [11].

In order to test whether the borders of crimes were capturing some underlying network structure we compared them against borders generated by demographic data. To ensure the comparison reflects the location, not the number of borders, we needed the demographic data and the crime data to have a similar number of communities. To get borders in the demographic data of roughly the same resolution as the borders in the crime networks we cut the dendrogram of demographic clusters at four heights resulting in a similar number of communities in both data sets. The three methods of hierarchical clustering and the four levels we cut each dendrogram at left us with 12 sets of demographic borders to compare each network to. The clusters for single linkage tended to be much more sparse than the other two methods of clustering so we dropped all but the most fine resolution of borders for single linkage. This left us with nine sets of borders for each metropolitan area we analyzed. Figure 1 shows the spatial representation of the borders structures compared to the socio-demographic characteristics for the Los Angeles Metropolitan Area. Figures 1(a), 1(b) and 1(c) depict respectively the borders for theft, burglary and assault, and a zoom in Santa Monica area is shown in figures 1(d), 1(e) and 1(f).

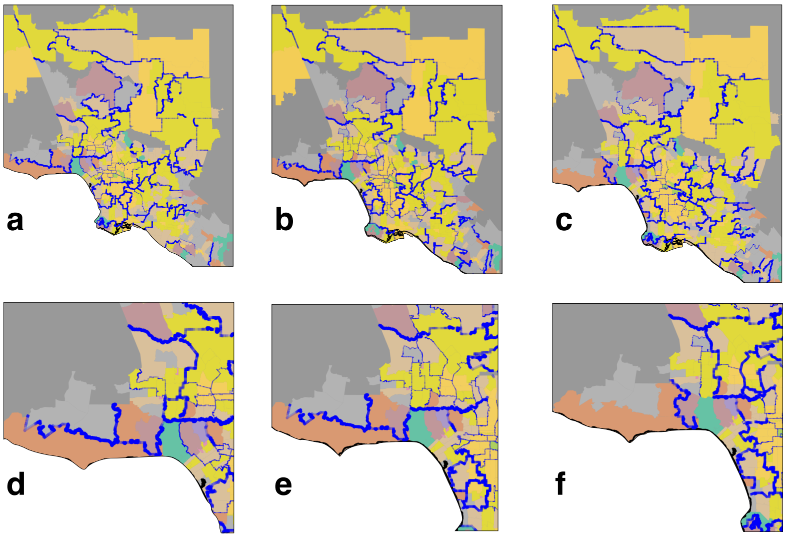

Fig. 1: Los Angeles Metropolitan Area – the structure of the borders (blue curves) for different crime types overlaying the ZCTAs clustered by their socio-demographic characteristics. (**a**) When we compare the borders of thefts with borders of socio-demographic clusters we can observe that in most cases the two sets do not coincide, suggesting that thefts tend to occur close to the socio-demographic borders. In fact, the theft borders partially correlate with the administrative limits of Los Angeles County regions such as Central L.A., Harbor, South L.A., Westside and San Gabriel Valley. (**b**) Similarly, burglary borders show a strong correlation with the county regions. Also, both theft and burglary borders set show a strong correlation. (**c**) The assault borders on the other hand seems to follow a different regime when compared to property crimes such as theft and burglary once their positions do not coincide. It is possible to see that assault borders tend to create subregions within socio-demographic clusters, particularly in dense areas such as Central L.A. Even though the borders of property crimes coincide in many areas of L.A., it is possible to observe some interesting dissimilarities such as a strong theft border across the center of the San Francisco Valley (**d**) that is nos present on the burglary borders set (**e**). Also, the strong borders of theft and burglary networks between Malibu and the Santa Monica Mountains show that the mountains represent a real topological barrier between Malibu and Calabasas and Agoura Hills. (**f**) Such borders, on the other hand, does not exist on the assault networks.

A lack of correspondence between the socio-demographic characteristics and the borders of crimes networks suggests that such structures are not just a proxy for the characteristics of the an area. But before we assume that the borders structures are representing some underlying network phenomena, we need to validate it against a null model. In order to carry this analysis out, we chose to generate random borders based on the demographic borders and reuse these for the comparison to each set of crime borders. For each of the 9 sets of demographic borders there were networks for each of four years (2007-2010), five distance parameters (0.1, 0.8, 1.6, 2.4, and 3.2 miles), and four types of crimes (all types, assault, burglary, and theft), for a total of 80 networks. This left us with 720 z-scores over a range of parameters for each city we were interested in.

Figure 2 highlights some of the different correlations between crime types, cities, and distance between crimes. For both of Miami and Los Angeles we saw more than a single standard deviation between other crime types for at least one of the five distances we analyzed. The fact that different crime types have statistically significant differences in correlation with the socio-economic boundaries in a city shows the structure of the crime networks are driven by different underlying phenomena. From this we concluded the structure of the

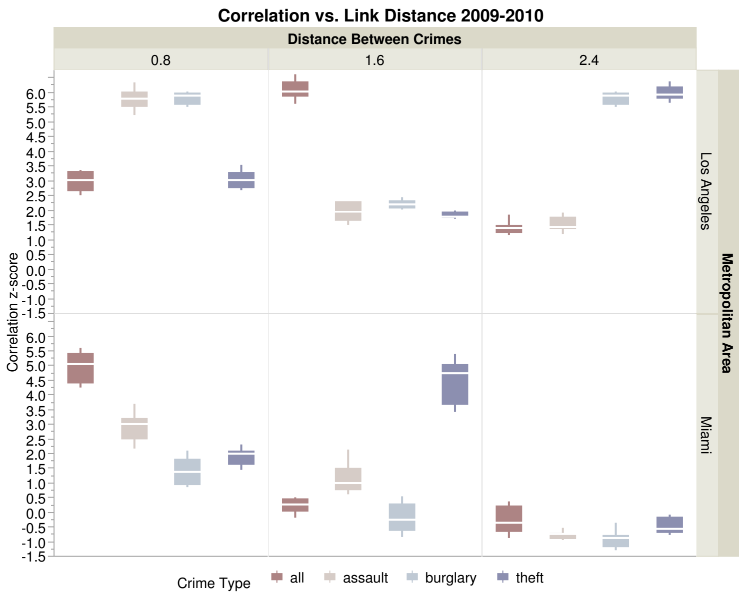

Fig. 2: As the distance between associated crimes changes the correlation with the demographic borders also changes. The varying patterns of this change for different crime types shows the crime networks have different structure for different types of crime.

crime networks could be used to find interesting patterns in the structure of criminal activity in a metropolitan area.

5 Conclusions and Future Works

The reliance of criminal activities on different types of networks makes network science an obvious choice for analyzing criminal activity. Using assumptions about the spatial distribution of crimes, we proposed a new model in which networks are built with links between crimes in close physical proximity. We showed that there exists an underlying structure of criminal activity that can be represented by the spatial distribution of the crimes. The model we propose allows the construction of criminal networks without the need for gathering extra data about individual criminals or patterns of crimes. No additional work is required to gain insights available from this new model because the data we used to construct the networks is already a part of law enforcement bookkeeping. The growth of computational resources available to law enforcement agencies allows for more complex analysis than the heatmaps that have been useful in the past.

For Los Angeles, CA and Miami, FL, we generated networks for multiple spatial distances, ranging from 0.1 miles to 3.2 miles, and various crime types. We showed that these networks capture real-world phenomena, by comparing borders between communities of crimes. We compared these borders to similar boundaries in demographic data and found statistically significant variations, which indicates the structure of the crime networks is reflecting real-world phenomena.

Acknowledgments

The authors acknowledge support from NSF grant No. 1263011 and NSF grant No. 1152306. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. Marcos Oliveira would like to thank the Science Without Borders program (CAPES, Brazil) for financial support under grant 1032/13-5.

References

1. Clifford R Shaw and Henry D McKay. *Juvenile delinquency and urban areas (Revision of 1942 ed.).* University of Chicago Press, 1969.
2. Clifford R Shaw, Frederick M Zorbaugh, Henry D McKay, and Leonard S Cottrell. *Delinquency areas.* University of Chicago Press, 1929.
3. Ralph B Taylor, Stephen D Gottfredson, and Sidney Brower. Block crime and fear: Defensible space, local social ties, and territorial functioning. *Journal of Research in crime and delinquency*, 21(4):303–331, 1984.
4. Robert J Sampson, Stephen W Raudenbush, and Felton Earls. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328):918–924, 1997.

5. Robert J Bursik Jr, Harold G Grasmick, et al. *Neighborhoods & Crime*. Lexington Books, 1999.
6. Jeffrey D Morenoff, Robert J Sampson, and Stephen W Raudenbush. Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence*. *Criminology*, 39(3):517–558, 2001.
7. Patrick C Jobes, Elaine Barclay, Herb Weinand, and Joseph F Donnermeyer. A structural analysis of social disorganisation and crime in rural communities in australia. *Australian & New Zealand Journal of Criminology*, 37(1):114–140, 2004.
8. John E. Eck, Spencer Chainey, James G. Cameron, Michael Leitner, and Ronald E. Wilson. *Mapping Crime: Understanding Hot Spots*. National Institute of Justice, 2005.
9. V Furtado, A Melo, ALV Coelho, and R Menezes. Simulating crime against properties using swarm intelligence and social networks. *Artificial crime analysis Systems: Using Computer Simulations and Geographic Information Systems*, pages 300–318, 2008.
10. Charis E. Kubrin and Ronald Weitzer. New directions in social disorganization theory. *Journal of Research in Crime and Delinquency*, 40(4):374–402, 2003.
11. Ned Levine and Patsy Lee. Journey-to-crime by gender and age group in manchester, england. In *Crime modeling and mapping using geospatial technologies*, pages 145–178. Springer, 2013.
12. Eric Jefferis. A multi-method exploration of crime hot spots: a summary of findings. *US Department of Justice, National Institute of Justice, Crime Mapping Research Center, Washington, DC*, 1999.
13. Malcolm K Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social networks*, 13(3):251–274, 1991.
14. Jeffrey Scott McIllwain. Organized crime: A social network approach. *Crime, Law and Social Change*, 32(4):301–323, 1999.
15. Peter Klerks. The network paradigm applied to criminal organisations. *Transnational organised crime: perspectives on global security*, page 97, 2003.
16. Christian Thiemann, Fabian Theis, Daniel Grady, Rafael Brune, and Dirk Brockmann. The structure of borders in a small world, 11 2010.
17. DJK Beavon, PL Brantingham, and PJ Brantingham. The Influence of Street Networks on the Patterning of Property Offenses. *Crime prevention studies*, 1994.
18. Dale Willits, Lisa Broidy, Ashley Gonzales, and Kristine Denman. Place and Neighborhood Crime: Examining the Relationship between Schools, Churches, and Alcohol Related Establishments and Crime. *Institute for Social Research, ...*, (March), 2011.
19. Sarah Foster, Lisa Wood, Hayley Christian, Matthew Knuiman, and Billie Giles-Corti. Planning safer suburbs: Do changes in the built environment influence residents' perceptions of crime risk? *Social Science & Medicine*, 97:87–94, 2013.
20. Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E*, 76:036106, Sep 2007.