Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 108C (2017) 2537–2541

Procedia Computer Science

International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland

Behavioral Characterization of Criminality Spread in Cities

Gabriel Spadon*, Lucas C. Scabora, Paulo H. Oliveira, Marcus V. S. Araujo, Bruno B. Machado, Elaine P. M. Sousa, Caetano Traina-Jr, and Jose F. Rodrigues-Jr†

University of Sao Paulo, Brazil

Abstract

Complex networks are commonly used to model urban street networks, which allows aiding the analysis of criminal activities in cities. Despite several works focusing on such application, there is a lack of a clear methodology focused in the analysis of crime behavior. In this sense, we propose a methodology for employing complex networks in the analysis of criminality spread within criminal areas of a city. Here, we evaluate synthetic cases of crime propagation concerning real criminal data from the North American city of San Francisco — CA. Our results confirm the effectiveness of our methodology in analyzing the crime behavior by means of criminality spread. Hence, this paper renders further development and planning on public safety in cities.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Science

Keywords: Complex Networks, Computer Aided Criminality Analysis, Crime Spread, Public Safety.

1 Introduction

Complex networks are proficient at representing complex systems, from neural networks to subway systems [1]. They can be employed, for instance, to model social behaviors, analyzing urban structures and evaluating the spread of epidemics. The arrangement of elements in a city impacts its citizens' behaviors. Violence and social disorder may be part of such behaviors and are usually influenced by the urban structure. This is the case of low-flow areas, which tend to be propitious for crime events since they end up being less surveilled [2]. In this context, it is helpful to identify and analyze crime dynamics through computational techniques. Similar topics were investigated in studies that focused on mapping [3, 4] and analysing [5, 6] crimes.

Our work provides a methodology that aids in the analysis of criminality. Its purpose is to answer how crimes behave in various scenarios of criminality spread. To do so, we employ notions of epidemic propagation, so that an epidemic represents a high level of criminality dispersion within a criminal region of a city. Accordingly, the hypothesis of this work is that *epidemic propagation models, allied with network mapping techniques and community-related measures, can aid in the characterization of criminal behavior and dispersion in a city.*

* Corresponding author: spadon@usp.br.
† The authors are grateful to Capes, CNPq, and FAPESP for their financial support.

1877-0509
Peer-review under responsibility of the scientific committee of the International Conference on Computational Science
10.1016/j.procs.2017.05.118

Considering that different crime types are related by being close to each other [7], we intend to identify how they behave in various scenarios of criminality spread. Accordingly, our proposal provides knowledge about criminality in the context of urban structures, working on issues related to: **(A)** the identification of criminal areas; **(B)** the analysis of the homogeneity of crimes' distribution; **(C)** the identification of behaviors regarding crimes' sparsity; and **(D)** the description of synthetic cases in which the spread of crimes behaves differently.

2 Proposed Methodology

Our methodology is divided into four phases: *Mapping of Urban Crimes*; *Criminal Community Identification*; *Community Distribution Analysis*; and *Behavioral Analysis of Crime Propagation*. Or dataset consists of an electronic map of San Francisco — CA[1] and a list of 1,916,911 georeferenced criminal incidents from Jan 1st, 2003 to May 5th, 2016[2]. To validate our methodology, we employed Homogeneity ($\mathcal{H}$) and Completeness ($\mathcal{C}$) scores, derived from an entropy-based cluster evaluation [8]. The early three phases have already been reported in Spadon *et al.* [7], and, therefore, due to space limitations, we only detail the main contribution of this paper, which focuses on the analyses of crime propagation in urban streets.

Crime Propagation Behavior Analysis. Crimes tend to be related to the dynamics of the urban streets. These dynamics can be linked to the structural features of the network, according to models of epidemic propagation. In this sense, we derived a model, based on notions of epidemic propagation, to portray the crime dispersion in a city region. Such model relies on basic principles of epidemic dynamics. According to Pastor-Satorras *et al.* [9], the **S**usceptible-**I**nfected-**S**usceptible (`SIS`) model is compartmentalized into two possible states (susceptible and infected), thus allowing two possible transitions. The former represents the interaction of a susceptible node with an infected one $\mathtt{S} \rightarrow \mathtt{I}$, becoming infected as well. The latter represents the opposite, when an infected node recovers and becomes susceptible $\mathtt{I} \rightarrow \mathtt{S}$.

An epidemic is defined in a time interval that holds t steps. Such steps can be described in any time unit, which will correspond to an iteration of a loop in an algorithm. Moreover, in a given epidemic, $\mathtt{N} = \mathtt{S}(t) + \mathtt{I}(t)$ holds true for all values of t, where $\mathtt{N}$ is the number of nodes, $\mathtt{S}(t)$ is the number of susceptible individuals, and $\mathtt{I}(t)$ is the number of infected nodes. Given this notion, two deterministic equations can be derived. First, $\frac{d\mathtt{P_I}}{dt} = +(p\mathtt{P_I})\mathtt{P_S} - q\mathtt{P_I}$ describes the infected nodes per time, whereas $\frac{d\mathtt{P_S}}{dt} = -(p\mathtt{P_I})\mathtt{P_S} + q\mathtt{P_I}$ describes the susceptible nodes; where $\mathtt{P_c}$, $\mathtt{c} \in \{\mathtt{I}, \mathtt{S}\}$, represents the percentage of nodes that are infected or susceptible, $p\mathtt{P_I}$ is the force of infection, and p and q are the infection and recovering probabilities, respectively.

Our model is based on the `SIS` model (see Algorithm 1) driven by the fact that human actions are able to prevent crime events. It is a reinterpretation of the algebraic nuances of the `SIS` model in the context of crime propagation. The spread of crimes across nodes indicates that particular nodes are becoming more dangerous. The degree of danger in a region turns it into a propitious spot for crimes; hence, criminal regions are being expanded.

The key assumption of our methodology is that if a crime occurs in a given spot (node) and nothing is done to prevent new crimes in its surroundings, then criminality tends to raise in neighboring nodes with a certain probability. Contrarily, when prevention policies are employed to counter crimes, the infection probability decreases with the distance from the initial crime spot. That is, the georeferenced characteristic of the network interferes in the way that

[1] Provided by www.openstreetmap.org.
[2] Available at data.sfgov.org.

Algorithm 1: Dynamic Propagation of Urban Criminality

```
Data: G = {V, E}: undirected urban network; C: set of urban crimes; l: percentage of human interference
I ← { }; S ← { }
for v ∈ G(V) do
    if |C_v| > 0 then
        I ← I ∪ {v}                                   // The seeds are marked as infected
        v_p ← 0.5 ; v_q ← 0        // v_p is the infection probability, and v_q is the recovering one
    else
        S ← S ∪ {v}                                   // The others are marked as susceptible
        v_p ← ∅ ; v_q ← ∅                             // Notice that such nodes are inert
for i ∈ I do
    for n ∈ neighborhood(i) do
        if (n ∉ I) and (rand([0, 1]) < i_p) then
            I ← I ∪ {v}; S ← S − {v} // The node is marked as infected and it is unmarked as susceptible
            n_p ← i_p × (1 − l)                       // It inherits the probability of its infector
            n_q ← 1 − n_p                             //  by considering the infection loss l
    if rand([0, 1]) < i_q then
        I ← I − {v}; S ← S ∪ {v}                      // The node recovers itself becoming susceptible
        n_p ← ∅ ; n_q ← ∅                             // Then, it becomes inert
return I                                              // The algorithm returns a set of infected nodes.
```

criminality spreads. This implies that crimes are more intense in the surroundings of where a real crime occurs, and are less intense or even nonexistent in farther regions. Shortly, our model has the following assumptions: **(I)** crimes propagate more easily to closer nodes than to farther ones; **(II)** the closer a newly-infected node is to a criminal node, the less likely it is to recover; **(III)** human interference in the system changes the way that criminality propagates.

The proposed model relies on two node states: criminal (nodes with at least one crime) and non-criminal (nodes with none) — for details of mapping and quantifying crimes as attributes of complex networks see Spadon *et al.* [7]. The contagious process of our model occurs reactively. An infected node propagates its criminality to all the nodes connected to it. In addition, one node at a time can be recovered, whereas many susceptible nodes can be infected simultaneously.

An epidemic propagates from an infected node to the directly connected nodes according to a probability p, whereas there is a probability q that an infected node will recover and become susceptible again. Moreover, $p + q = 1$ holds true in every stage of the epidemic propagation. Considering a system that satisfies this model, the nodes with registered crimes are the seeds; that is, based on our dataset, they are the early disseminators of criminality. Accordingly, as time goes by, the crimes that occurred in the seeds continuously impact the directly connected nodes. For this reason, we assign to the seeds a high probability of infection, *i.e.* $p = 0.5$. We rely on a high probability of infection for the seeds because the seeds have been acting as crime spots over the years. Intuitively, the seeds have also a probability $q = 0.5$ of being recovered.

The probability of farther nodes becoming infected is influenced by their distance from the seeds, *i.e.* the propagation is impacted by the network topology. Conversely, the epidemic tends to lose its strength as it goes farther from a given criminal node. Such distance impacts the infection probability p and is referred to as the infection loss l, $0 \leq l \leq 1$, where 0 indicates no loss and 1 indicates the opposite. Hence, when a susceptible node becomes infected, it inherits a percentage p_{new} of the infection probability p_{old} from its infector, defined as $p_{new} = p_{old} \times (1 - l)$; symmetrically, the new infected node acquires a higher probability of recovering, which is defined as $q_{new} = 1 - p_{new}$. The value of l can be understood as the human interference; for instance, when the police act in a crime region and, due to the effectiveness of its actions, the criminality tends to propagate slower than in a system where there is no interference.

Based on this model and on complex network analysis techniques — data mapping, community detection, similarity assess, cluster-based behavioral analysis, our work presents how to analyze the impact of the intervention on criminal spots, allowing to prevent epidemic outbreaks in criminal areas. We show how the crimes are impacted in different situations of human intervention by measuring their score of Homogeneity along with their Completeness scores.

3 Results and Discussions

Our results start by labeling nodes into two categories, criminal and non-criminal. The criminal nodes served as seeds, whereas non-criminal nodes were marked as susceptible. In this context, we heuristically evaluated three specific scenarios. Each one has a different semantic perspective that explains their criminality spread across the nodes (Section 2). The scenarios are described as follows: **(1)** the first one outlines the chaos ($l = 0.00$), in which there is no interest in acting against crimes; **(2)** the second one represents the desire to change ($l = 0.25$), in which there are moderate efforts to revert the criminal scenario, but they are still not enough; **(3)** the third one describes the decay of criminality ($l = 0.50$), which ensures peace to all subregions of a city and also portrays a more effective crime prevention policy. For other values of l, the target scenario should behave linearly regarding the aforementioned ones. We evaluated each scenario by means of criminality propagation, adapting the SIS epidemic model. Such evaluations indicate that by adopting crime prevention policies, the number of criminal and non-criminal nodes differ considerably. The results are depicted in Figure 1 using $p = 0.5$.

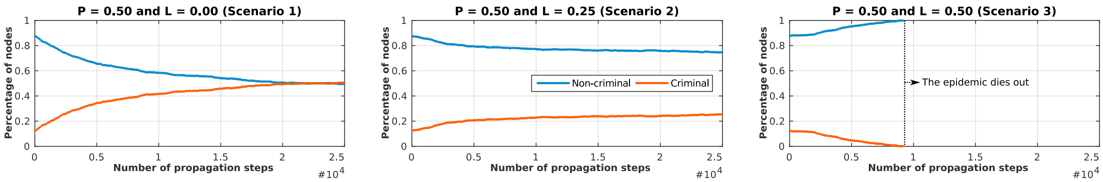

Figure 1: Results from crime propagation using a model derived from the SIS epidemics model. We show two cases where the values of p vary to depict how the model behaves. Each case is evaluated by altering the constant that denotes human intervention in a city subregion.

In *Scenario 1*, it is possible to see an increase of criminality by crimes not being properly prevented. Consequently, the number of criminal nodes, which is originally less than 10% of all nodes, comes to be almost equal to the number of non-criminal ones in almost $20,000$ propagation steps. However, in *Scenario 2*, a decay in the criminality spread is seen, since the number of criminal nodes does not increase as in the former scenario. In this case, they stay lower than the non-criminal nodes, denoting a resistance to the spread. Besides, criminal nodes represent about 20% of all nodes at the end of the simulation, which indicates that criminality was contained compared to the previous scenario, but nonetheless human intervention was not fully effective. Subsequently, in *Scenario 3*, the number of criminal nodes was unable to become higher than the number of non-criminal ones. Instead, close to $9,000$ propagation steps, the criminal nodes were powerless to further propagate criminality. Furthermore, they were capable of recovering themselves, and thus the epidemic died out. Hence, the number of infected nodes converged to zero, which denotes that the criminality was contained. At last, we evaluated how intrinsic is the criminality in the studied region of the city using cluster evaluation measures.

Along these lines, in *Scenario 1* ($\mathcal{H} = 0.00111$; $\mathcal{C} = 0.00047$), we have an outbreak case, in which, considering the Homogeneity score, the criminality is scattered across all the studied subregion, and it is common to see criminal and non-criminal nodes equally distributed. Also, both node types are scattered and recurrent, represented by the Completeness score close to zero. In *Scenario 2* ($\mathcal{H} = 0.00665$; $\mathcal{C} = 0.00266$), the majority of nodes are non-criminal when compared to *Scenario 1*, and the region is more unbalanced in terms of criminal and non-criminal nodes. This is a consequence of the human intervention in the system. Furthermore, it is possible to see that the values for both scores contrast with the ones of the other scenarios — the chaos of *Scenario 1* and the peace of *Scenario 3*. This is evidence that criminality is being changed due to a stronger human interference. Specifically, in *Scenario 3* ($\mathcal{H} = 1.00000$; $\mathcal{C} = 0.00000$), crimes were eradicated, represented by the Homogeneity equal to 1, indicating that all communities have only non-criminal nodes. Also, Completeness equals to 0, pointing out that only one of the classes has prevailed while the other has been eradicated. Hence, *Scenario 3* exhibits a behavior that tends to be opposite to the one seen in *Scenario 1*.

The results demonstrate that our methodology provides a view of the crime scenario of a city, taking into account different degrees of human intervention in criminal activities of a given city subregion. Therefore, it enables the analysis of crime outbreaks and the impact on criminal regions by containing or disseminating criminality to the entire network.

4 Conclusion

In this paper, we propose a technique for assessing crime behavior in street networks considering the perspective of the criminality spread. Our results are based on the analysis of real crime data of the city of San Francisco — CA which was represented by means of complex networks through electronic maps. Our results indicate that epidemic propagation models, allied to network mapping techniques and community-related measures, can aid the characterization of crime behavior and its dispersion across a city. Besides, criminal regions are the most propitious to propagate criminal activity through the network as a whole, requiring a higher human intervention. Lastly, as future work, we will expand our methodology by including time series analyses in the context of crime behavior.

References

[1] S Boccaletti, V Latora, Y Moreno, M Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Phys Rep*, 424:1–133, 2006.

[2] R Deryol, P Wilcox, M Logan, and J Wooldredge. Crime places in context: An illustration of the multilevel nature of hot spot development. *J Quant Criminol*, 32:1–20, 2016.

[3] V Spicer, J Song, P Brantingham, A Park, and M A Andresen. Street profile analysis: A new method for mapping crime on major roadways. *Appl Geogr*, 69:1–9, 2016.

[4] S Shiode and N Shiode. Network-based space-time search-window technique for hotspot detection of street level crime incidents. *Int J Geogr Inf Sci*, 27:1–16, 2013.

[5] S White, T Yehle, H Serrano, M Oliveira, and R Menezes. The Spatial Structure of Crime in Urban Environments. *Int Conf on Social Info*, 9471:1–9, 2015.

[6] S J Rey, E A Mack, and J Koschinsky. Exploratory Space-Time Analysis of Burglary Patterns. *J Quant Criminol*, 1, 2012.

[7] G Spadon, L C Scabora, M V S Araujo, P H Oliveira, E P M Sousa, C Traina-Jr, B B Machado, and J F Rodrigues-Jr. Complex network tools to understand the behavior of criminality in urban areas. In *Information Technology: New Generations*. Springer, 2017.

[8] A Rosenberg and J Hirschberg. V-measure: A conditional entropy-based external cluster evaluation measure. *EMNLP-CoNLL*, 7:1–11, 2007.

[9] R Pastor-Satorras, C Castellano, P Van Mieghem, and A Vespignani. Epidemic processes in complex networks. *Rev Mod Phys*, 87:1–54, 2015.