UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Characterizing mobility patterns and describing collective behavior in georeferenced graphs through analytical processing of real-world complex networks

Gabriel Spadon de Souza

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-C²MC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Gabriel Spadon de Souza

Characterizing mobility patterns and describing collective
behavior in georeferenced graphs through analytical
processing of real-world complex networks

Master dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP,
in partial fulfillment of the requirements for the
degree of the Master Program in Computer Science
and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Prof. Dr. José Fernando Rodrigues Júnior

USP – São Carlos
August 2017

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

S719c Souza, Gabriel Spadon
Characterizing mobility patterns and describing
collective behavior in georeferenced graphs through
analytical processing of real-world complex
networks / Gabriel Spadon Souza; orientador José
Fernando Rodrigues Junior. -- São Carlos, 2017.
72 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2017.

1. Complex Networks. 2. Urban Systems. 3.
Collective Behavior. 4. Pattern Recognition. I.
Rodrigues Junior, José Fernando, orient. II. Título.

Gabriel Spadon de Souza

Caracterizando padrões de mobilidade e descrevendo
comportamento coletivo em grafos georreferenciados por
meio de processamento analítico de redes complexas reais

Dissertação apresentada ao Instituto de Ciências
Matemáticas e de Computação – ICMC-USP,
como parte dos requisitos para obtenção do título
de Mestre em Ciências – Ciências de Computação e
Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e
Matemática Computacional

Orientador: Prof. Dr. José Fernando Rodrigues Júnior

USP – São Carlos
Agosto de 2017

*Dedicated to my grandfather Antônio Spadon and to the loving
memory of my grandmother Marileide Pardo Spadon.*

Acknowledgements

I would like to express my kindly gratitude to my advisor, Ph.D. Jose Fernando Rodrigues Junior, who wisely and patiently guided my dissertation work, providing plenty of academic opportunities by teaching me how to research. I am also grateful to my research group colleagues (GBdI) for the good times that we spent together these last years, which was of significant importance to motivate my work and myself. I kindly thank my bachelor's professors from Sao Paulo State University (UNESP) that provided me with all the means to go further during my undergraduate course, making possible for me to be accepted in a post-graduation program. Furthermore, my family and girlfriend had a fundamental importance in this work by always supporting me and believing in my dreams, no matter the difficulties that I faced to achieve them; they always stood by my side. My sincere thanks to my grandparents, who always encourage me to go further, despite not understanding the academic world. Moreover, my recognition to my hometown friends, the ones that granted me with incredible life experiences through remarkable years of friendship. Finally, yet importantly, I would like to thank CAPES that fully financed my master's degree.

*“Hey you, don’t tell me there’s no hope at all.
Together we stand, divided we fall.”
Pink Floyd (1979)*

RESUMO

SOUZA, G. S. **Caracterizando padrões de mobilidade e descrevendo comportamento coletivo em grafos georreferenciados por meio de processamento analítico de redes complexas reais**. 2017. 72 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

As cidades são sistemas complexos de interação social e de transporte. Suas estruturas podem ser usadas para modelar redes de mobilidade urbana — *i.e.* redes complexas que representam a geometria de uma cidade — permitindo a consecução de atividades analíticas para descoberta de padrões e para a tomada de decisão baseada em dados. A geometria da cidade carrega informações intrínsecas que auxiliam atividades relacionadas à análise de dados provenientes do cenário urbano. As informações inerentes a tais análises podem ser usadas para melhorar a qualidade de vida dos habitantes de uma região, ou para entender a dinâmica de centros urbanos. Diversos processos analíticos aplicados a tais cenários carecem de metodologias para analisar o padrão criminal e para identificar estruturas urbanas mal planejadas. Deste modo, este trabalho tem por objetivo prover meios para análise topológica de regiões criminais e para a identificação de inconsistências urbanas, as quais apontam para regiões que carecem de mobilidade e acesso para outras regiões de uma cidade. Neste sentido, foi desenvolvido um conjunto de procedimentos algébricos e algorítmicos capazes de revelar padrões e meios para compreensão e análise dos dados. Mais especificamente, foram desenvolvidos métodos de pré-processamento para transformar mapas eletrônicos georreferenciados em grafos que representam cidades, foi utilizado um conjunto métrico analítico e outro com base em processos epidêmicos para entender a dinâmica intrínseca à criminalidade de uma cidade, e por fim, foi desenvolvido um conjunto de formalismos e operações baseados em teoria dos conjuntos para identificar falhas no desenho das estruturas urbanas que impactam no acesso viário em centros urbanos. Os resultados deste trabalho versam sobre o desenvolvimento de novos métodos para preparar mapas na forma de redes de mobilidade urbana; na análise de crimes baseada em sua disposição espacial; no desenvolvimento de um modelo capaz de descrever a atividade criminal de uma cidade; e, em um conceito baseado na análise de regiões críticas identificadas a partir do desenho urbano.

Palavras-chave: Redes Complexas, Sistemas Urbanos, Comportamento Coletivo, Reconhecimento de Padrões.

ABSTRACT

SOUZA, G. S. **Characterizing mobility patterns and describing collective behavior in georeferenced graphs through analytical processing of real-world complex networks**. 2017. 72 p. Dissertation (Master of Science – Computer Science and Computational Mathematics) – Institute of Mathematics and Computer Sciences, University of São Paulo, São Carlos – SP, 2017.

Cities are complex systems of transportation and social activity; their structure can be used to model urban street networks — *i.e.* complex network that represents the geometry of a city — allowing analytical activities for data-driven decision-making. The geometry of a city holds intrinsic information that can support activities related to the analysis of the urban scenario; of higher importance is the use of such information to enhance the quality of life of its inhabitants and/or to understand the dynamics of an urban center. Several of these analytical processes lacks in-depth methodologies to analyze crime patterns and ill-designed urban structures, which can provide for public safety and urban design. Consequently, it is our goal to provide means for the structural and topological analysis of highly criminal regions of cities represented as complex networks, and for the identification of urban planning inconsistencies that point to regions that lack access from/to points of interest in a city. In this regard, we devised a set of algebraic and algorithmic procedures that are capable of revealing patterns and provide for data comprehension. More specifically, we introduced pre-processing techniques to transform georeferenced electronic maps into graph representations of cities; we used metric-based and epidemic processes to understand the dynamics of cities in what refers to criminality; finally, we introduced a novel set of formalisms and operations based on set theory to identify design flaws concerning access in urban centers. Our results refer to approaches to preprocess and prepare maps in the form of urban street networks; to the analyses of crimes based on their spatial disposition; to the development of a model to describe criminal activities; and, to the advance of a concept based on critical problems in the urban design.

Keywords: Complex Networks, Urban Systems, Collective Behavior, Pattern Recognition.

List of Figures

Figure 1 – The filtering process used to identify the relevant communities while discarding the mischaracterized ones. We sort the communities by their average criminality, and we eliminate the ones that have less than 100 nodes. The highlighted area indicates the communities that are, at the same time, highly populated and highly criminal.	35
Figure 2 – The most highly criminal region from San Francisco – CA, USA, identified by evaluating the crime density in the city from 01/01/2003 to 05/18/2016. Each colored area represents the occurrence of a crime type or the intersection of a crime type with another type of crime. In the legend, each CM (from 1 to 3) denotes a type — <i>assault</i> , <i>theft</i> and <i>minor crimes</i>	38
Figure 3 – 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. The three scenarios of interest to this work are represented by the graphs on the top of the image, which are ordered from left to right. For comparison, our model uses $p = 0.5$	49
Figure 4 – Redundancy Removal in a sub-region of the city of Sao Carlos – SP, Brazil.	60
Figure 5 – City distribution and characteristics. (a)(d) The in-out/total degree distribution of the primary and auxiliary network. (b)(e) The average nearest neighbor degree correlation histogram. (c)(f) The average target neighborhood degree, considering the source node degree.	61
Figure 6 – The identification of urban inconsistencies in a general case and in a real-world scenario. Also, despite the shape of the images, the perimeters are always convex.	62

List of Algorithms

Algorithm 1 – Dynamic Propagation of Urban Criminality.	47
Algorithm 2 – Redundancy Removal.	55
Algorithm 3 – Inconsistency Identification.	58

List of Tables

Table 1	– Communities with the highest crime average, identified from: (i) <i>topology</i> , (ii) <i>assaults</i> , (iii) <i>theft</i> , and (iv) <i>minor crimes</i> . Column <i>Average</i> denotes the crime average and <i>Total</i> denotes the number of nodes in each community.	39
Table 2	– The Homogeneity and Completeness scores used to evaluate the quality of each community set and to quantify how intrinsic is the criminality within them.	41
Table 3	– Redundancy Removal in five cities, indicating that the length of the edges is the same in every case and exhibiting the percentage of elements removed during the process.	59
Table 4	– The inconsistencies from the city of Sao Carlos concerning public facilities (i) Police Stations, (ii) Hospitals, and (iii) Schools, detected through Inconsistencies Identification.	62

Table of Contents

	Page
1 INTRODUCTION	23
1.1 Motivation	23
1.2 Problem	23
1.3 Hypotheses	24
1.4 Rationale	25
1.5 Contributions	26
2 COMPLEX NETWORK TOOLS TO UNDERSTAND THE BEHAVIOR OF CRIMINALITY	29
2.1 Introduction	29
2.2 Related Work	30
2.2.1 Mapping of Urban Crimes	31
2.2.2 Crime Analysis	31
2.3 Proposed Methodology	32
2.3.1 Background and Datasets	32
2.3.2 Mapping of Urban Crimes	33
2.3.3 Criminal Community Identification	34
2.3.4 Crime Analysis	36
2.3.4.1 Homogeneity	37
2.3.4.2 Completeness	37
2.4 Results and Discussions	38
2.5 Conclusion	42
3 BEHAVIORAL CHARACTERIZATION OF CRIMINALITY SPREAD IN CITIES	45
3.1 Introduction	45
3.2 Proposed Methodology	46
3.2.1 Crime Propagation Behavior Analysis	46
3.3 Results and Discussions	48
3.4 Conclusion	50

4	IDENTIFYING URBAN INCONSISTENCIES VIA STREET NETWORKS	51
4.1	Introduction	51
4.2	Related Work	52
4.2.1	Urban Space Characterization	52
4.2.2	Traffic Planning and Mobility Enhancement	53
4.2.3	Other Related Work	53
4.3	Proposed Methodology	54
4.3.1	Background and Datasets	54
4.3.2	Preprocessing Technique	54
4.3.3	Identification of Urban Inconsistencies	56
4.3.4	Perimeter set	56
4.3.5	Network-distance set	56
4.3.6	Categorizing Urban Inconsistencies	57
4.3.7	Inconsistencies Identification Algorithm	57
4.4	Results and Discussions	59
4.4.1	Assessment of Redundancy Removal Algorithm	59
4.4.2	Assessment of the Auxiliary Network	60
4.4.3	Identification of Urban Inconsistencies	61
4.5	Conclusion	63
5	CONCLUSION	65

Introduction

1.1 Motivation

Cities represented as complex networks can support analytical studies and experimentation [1]. This is due to the diversity of types and shapes of such networks, which represents the topological and geometrical characteristics of a city. Studies in these networks are potentialized when correlating the network structure with daily issues, such as lack of transportation, health care, and safety. These issues give rise to interesting problems that can be tackled by using network analysis tools [2]. For instance, among transportation problems, there can be slowness, congestion, lack of road access, and others. Considering public safety, we can list problems as the identification of highly criminal regions within a city and the analysis of similarity of criminal activities.

Along these lines, the motivation of this dissertation work is to inspect the urban space identifying patterns that can guide the processes of planning and design; such approach can improve urban-related issues. We addressed this topic seeking for ways to improve the city structure by acting over the network mobility and by understanding the social behavior of citizens. Based on complex networks, it is our goal to analyze the access in urban scenarios aiming to support decision-making activities to enhance people's life.

1.2 Problem

Considering the Computer Science point of view, urban planning is a decision-making task that aims to improve the well-being of citizens by enhancing a city at its structural level [3]. Processes that are related to urban planning can better allocate a private or public service; point to regions of the city that need intervention; and also, reveal planning mistakes from previous planning actions. For all those purposes, we use the distance of its streets, which represents the connections between elements in the urban space. It has a huge importance to the effectiveness of timing-critical activities [4], as ambulances and

police officers displacement, which are related to travel logistics, to the technology used for planning, and to physical, political, and financial constraints [5].

Research approaches in these areas have focused on transportation and access to facilitate the displacement in a city [6]. Several research groups provided with methods and methodologies to related tasks [7–9]. They consider geometric and topological characteristics to characterize and analyze different cities. However, cities are evolving according to time, opening possibilities for constant intervention. As a consequence, the methods found in the literature still lack the potential to characterize the urban scenario, especially when considering the large electronic maps available in today’s georeferenced technologies. It is our contention that the topological and geometrical analysis of the street network can provide for better routing, positioning, resource allocation, and other activities that influence the citizens’ lives.

1.3 Hypotheses

Considering urban-related problems and cities represented in the form of complex networks, it is desirable to develop techniques to enhance processes of planning and of designing urban structures. Considering such issues, in this dissertation work, we developed three different methodologies that compose our contributions, which are outlined in the following.

- We analyzed the city by the topology of the complex network and through a crime dataset. Our intention was to characterize the overlap of crime occurrences, which helped us to identify criminal communities. These communities are regions that point to neighborhoods that need better police intervention. This study followed the hypothesis that: **by employing network mapping techniques, allied to distance-based properties of graphs, it is possible to trace the relationship between areas that are highly criminal within a city.**
- We further studied crime propagation by means of an epidemic propagation model. We considered the criminal communities, outlying regions that disseminate criminality to neighboring areas. Such approach had the intention to describe how crimes propagate when police intervention is effective in a given region. Accordingly, the hypothesis of this research is that **epidemic propagation models, allied to network mapping techniques and community-related measures, can aid in the characterization of criminal behavior and dispersion in a city.**
- We identified that the geometry and the disposition of the streets can interfere in the way that services are available to citizens. Consequently, the position of these services jeopardizes the access and mobility of the street network. In this sense, we put together geometrical features extracted from the topology of a network,

georeferenced information, and graph measures to analyze nodes that are suffering from the lack of urban planning in cities; such nodes are referred to as inconsistencies. Inconsistencies are related to a set of referential nodes, that can be public or private services of a city. For instance, when referential nodes are police stations, the inconsistencies can be used to identify low-access regions that can be turned in propitious neighborhoods for criminal activity. This assumption is because these regions are less accessible when considering a police station and, consequently, they tend to be less surveilled. The hypothesis of this investigation is that **geometric and geographic measurements derived from street networks can aid in the identification of urban inconsistencies observed on ill-designed urban structures.**

1.4 Rationale

Following we provide evidence that the topic of research is active by citing relevant works. The advancements and challenges found in the literature justify our interest and guides our approaches.

Complex networks have long been used in activities related to the analysis of urban systems, starting from the representation of its structures up until the mobility pattern of its citizens. Considering these scenarios, some authors [10] have focused on metric and analytical methods, while others [11, 12] covered the city planning and design activities.

Along with these lines, Porta, Crucitti, and Latora [13] perform a detailed study of centrality in complex networks of cities. The authors introduced a centrality-based methodology, which is defined and implemented over 1-square-mile street sub-regions. The 1-square-mile is a sampling procedure that provides for larger scalability over non-optimized algorithms. In this sense, their proposal is limited because it focuses on local characteristics, but it is not able to characterize macro-structural features like those observed in cities.

The low scalability of methods caused by sample strategies can be seen in other works, like the one by Crucitti, Latora, and Porta [14] who identified common characteristics between self-organized and planned cities and the one by Porta *et al.* [15] that used complex networks to infer the location of economic activities in cities.

Other authors [16, 17] covered points related to the optimization of transportation, mainly with the allocation of resources in street networks. Notwithstanding, to solve issues related to the urban scenario, several authors have been considering data derived from the structural conditions of cities [18], the intense traffic of vehicles [19] and collective behavior of their citizens [20, 21]. This is the case of Viana and Fontoura [22], who explored the impact of fast long-range connections to improve the traffic between two points. Their study differs from the beliefs of our proposal as we consider just one way to go from

one point to the other, while they consider heterogeneous transportation means in the locomotion processes.

Also, there is the research of Travencolo and Costa [23], who defined the metric concept of accessibility in terms of complex networks derived from cities. Their proposal focused on measuring access by using entropy and it differs from ours as they consider the access as a global property of the network, while we consider it as a property of a pair of nodes.

All issues covered by these works are related to mathematical and geometrical properties derived from the network topology, such as the shortest path [24, 25] and centrality indices [26, 27] that are specialized considering the network geometry [28, 29], spatial arrangement [30, 31], and also the attributes attached to network elements [32, 33].

In a particular perspective, Strano *et al.* [34] investigate the network's geometrical attributes using centrality measures. The authors report that cities share architectural similarities due to their quasi-planarity. However, their methods focused in seeking information about similarities, but they were not proficient in identifying irregularities in the city design, which is the desired result in urban planning processes.

The urban analysis is not limited to the previous works. For instance, in [35] the authors investigated the centrality of cities to understand its traffic and in [36] they analyzed the centrality of the elements acquired from the city representation. Considering the literature on complex networks, here we use spatial and topological metrics, contributing to the state of the art in the georeferenced network analysis. We show that our methods are able to characterize the urban space according to perspectives not encompassed before.

1.5 Contributions

The study described in this document explores three different fronts (see Section 1.3) of graph techniques applied to the processing of urban data. The fronts consider stages of data preprocessing, data representation, algorithmic processing, and the systematization of results. The results of these works provided means for the understanding of urban structures, allowing the identification of problems and the mining of solutions.

In Chapter 2 [37], we used a complex network created from georeferenced data to identify patterns and behaviors through criminal data. We have analyzed the urban scenario from the perspective of its topology and crimes. Our contributions come in using a complex network to represent and analyze the real-world space, in being independent of socioeconomic information, and in processing an in-depth methodology to map and analyze crimes — which is done by using similarity-based and cluster-based measures.

Subsequently, in Chapter 3 [38] we extended our previous methodology in order to

analyze crimes dissipation by means of the topology of a complex network. We present a four-phase methodology that discusses ways to identify highly criminal areas, that analyze the uniformity of crimes, identify behaviors and patterns regarding their sparsity, and is able to portray synthetic cases of criminality regarding different degrees of crime prevention. Our methods are validated via the analysis of the scenarios of crime dispersion, where we investigate and interpret the behavior of our set of measures.

Lastly, we introduce in Chapter 4 [39] concepts related to Urban Inconsistencies — nodes that lack access and mobility considering a set of referential nodes; we discussed mathematical and computational processes to improve the design of urban agglomerates. For this purpose, we considered a set of distance-based measures gathered from street networks to describe its structure and mobility. This investigation contributes in a variety of areas and thorough processes to deal with urban networks, including extraction, pre-processing, mathematical modeling and algorithmic innovation; it brings light to the processes that involve computer-assisted urban design and the mobility enhancement of urban systems.

This dissertation is composed by a *collection of articles* and it is organized as follow: Chapter 2 introduces our first study “Complex Network Tools to Understand the Behavior of Criminality in Urban Areas” that describes ways to identify criminal communities and their behaviors; Chapter 3 discusses our second study “Behavioral Characterization of Criminality Spread in Cities”, which extends our methodology for crime identification and proposes a way to model its dispersion in a city; Chapter 4 discusses our third study “Identifying Urban Inconsistencies via Street Networks” that brings light to the process of identifying urban planning mistakes by using a set of distance-based metrics from the network topology; the conclusions come in Chapter 5.

Complex Network Tools to Understand the Behavior of Criminality in Urban Areas

2.1 Introduction

Complex network has long been used to model social behavior, spatial patterns, and urban structures. It is capable of representing neural connections, data transmission schemes, electric power grids, airports, as well as street, rail and subway networks [40–48]. Street networks, which are complex networks that represent cities, are able to describe flow relationships and social behavior in an urban zone [11].

As dynamic organisms, cities evolve (shrink or enlarge) with respect to their space and population, causing changes in their structures. Such structures are of high relevance as they directly influence the flow within the cities and, consequently, the behavior of their citizens. Nevertheless, violence and social disorder may also be related to the urban structure of a given city. For instance, low-flow areas tend to be propitious for crime events, since such areas end up being less surveilled when compared to high-flow areas. For that reason, crime events usually occur within regions that can be detected based on their structural and behavioral properties [49]. In this context, it is our contention that, by means of computational techniques, it is possible to identify and understand the crime dynamics in an urban network. As so, this chapter provides a methodology for aiding the analysis of urban criminality — *e.g.* how specific crimes occur in certain areas and how their types are related to each other by being in neighboring areas — *i.e.* by being similar.

The hypothesis of this work is that *by employing network mapping techniques, allied to distance-based properties of graphs, it is possible to trace the relationship between areas that are highly criminal within a city.* Our assumption is that similar crimes occur in adjacent regions. Such crimes emerge from, among other factors, the urban organization,

and their inter-similarity carries correspondence to the distance from one region to the other. Technically, our proposal is based on: (i) community detection algorithms applied to datasets of isolated crime types; (ii) pattern and similarity analyses, which compare crimes occurred in distinct regions within the same city. Based on those techniques, we propose a threefold methodology consisted of: **(i) Mapping of Urban Crimes** – we describe how to combine a city’s georeferenced urban structure with crime records into a complex network with potential for analytical tasks; **(ii) Criminal Community Identification** – we show how to detect criminal communities by exploring geographical as well as structural properties; **(iii) Crime Analysis** – we methodologically compare crime patterns by identifying relationships between their communities, especially in the face of diverse crime types.

Notwithstanding, we are aware that criminality is a sociological dilemma. Cautiously, our work is limited to a specific perspective, centered on the similarities between georeferenced interacting crimes [37]. The corresponding contributions, therefore, are limited to the understanding of criminality in the context of the urban organization, approaching issues related to: (i) *an approach to identify highly criminal areas*; (ii) *the characterization of the city space by inferring the homogeneity of its crimes*; (iii) *the identification of city sectors that present behaviors related to the sparsity of their crimes*; and (iv) *the delineation of the city space by identifying regions that are related, acting as crime spots for different crime types*.

The remainder of the chapter is structured as follows. Section 2.2 presents the related work, discussing how our methodology stands out from theirs. Section 2.3 describes the dataset used in this study and the methodology proposed. Section 2.4 discusses the results obtained from the application of our methodology to the crimes dataset, which describes a real scenario from the street network of the North American city of San Francisco. Finally, Section 2.5 presents the conclusion.

2.2 Related Work

The use of complex networks allows obtaining insights regarding the urban planning process of a city, particularly when considering public safety. Therefore, it is important to perform compelling crime analyses considering the real-world space. This is because the understanding of criminal patterns and their distribution is the beginning to prevent crimes and to contain their dispersion.

Several studies in the literature have dealt with the topics of urban networks and criminality. This section describes some of these works organized into two categories related to the phases of our proposal: **Mapping of Urban Crimes**, which focuses on representing raw crime data with a complex network, and **Crime Analysis**, which refers

to identifying criminal behaviors and patterns by analyzing urban street networks.

2.2.1 Mapping of Urban Crimes

Spicer *et al.* [50] describe a theoretical framework capable of mapping urban crimes into related locations in a city. Their work discusses the pros and cons of mapping methods. To represent a city, they use street networks and GIS software, describing methods for mapping georeferenced elements into a complex network. However, their methods associate a crime with the nearest node or edge based on address geocoding, not having any relation with graph measures. Moreover, all methods are superficially introduced, not presenting any formal fundamentals.

Shinode and Shinode [51] aim at introducing a search-window method to analyze urban crimes in street networks. They represent a search window as a sub-area that concentrates a high number of crimes in small regions; besides that, the crimes are used to identify clusters spatially and temporally. The authors validate their methods through an empirical case study. However, their criminal data derive from 911-calls made in 1996; that is, they used a dataset that carries few details, and that is outdated, especially when one considers the conclusions claimed by the authors. Such information implies a superficial analysis of the crime behavior and the ways to prevent it.

Those previous works aim at identifying highly criminal areas to predict future criminal events. They use several properties of the data to map them into a node, an edge or a combination of both. However, they use non-graph measures to build a city scenario, which fails in characterizing the real world. Our approach, on the other hand, benefits from modeling the real-world space via a georeferenced complex network. Such network allows us to merge two georeferenced grids, one describing crimes and the other outlining city elements, which enables us to characterize crimes and extract their patterns.

2.2.2 Crime Analysis

White *et al.* [52] represent criminal activities via a georeferenced complex network, through which they compare criminal demographics and identify criminal communities. The demographic regions are created by clustering the urban areas using socioeconomic borders, whereas the crime network is formed by linking crime locations that are close according to the Euclidean distance. To identify the criminal communities, the authors employ a label propagation technique. Their results show that the crime network can be constructed without requiring information about individual crimes. However, their crime network does not take the urban topology into account, so their comparison is based only on socioeconomic information.

Rey *et al.* [53] employ spatiotemporal techniques to represent and analyze urban

crimes. Their approach provides a way to quantify neighborhood criminality and is able to identify patterns from the data. Moreover, they discuss spatial crime analyses as a way to determine the influence of a crime on its surroundings, analyzing an entire city and providing an evaluation of its crimes. However, their work is based on outdated data, derived from a police district in the period from 2005 to 2009. Besides, their methods rely on black-box GIS technologies, used to summarize criminal areas by geocoding each crime into a quarter-mile grid cell that has residential units.

Other approaches include the proposals of Galbrun *et al.* [25], who focus on crime mapping to estimate the probability of a crime occurrence within a street segment; of Fitterer *et al.* [54], who use a statistical model to predict crimes in the city of Vancouver; and of Bogomolov *et al.* [55], who predict *crime hotspots* in London by using mobile data. Our work differs from theirs because (i) *we consider crimes within a more complete and up-to-date period*; (ii) *we map crimes directly, i.e. there is no space summarization in the crime mapping phase*; and (iii) *we benefit from complex network theory, which are a reliable toolset to design georeferenced data*.

Considering real crime events and a real-world city, like those presented in this section, we introduce ways to map and analyze crimes, supported by criminal communities extracted from complex networks, allowing outcomes far more detailed and advanced than the former works. We not only show how to map crimes into georeferenced complex networks but also how to analyze their impact on the city, as well as the relationships among different crime types. Mainly, we describe results from various points of view, for multiple crime types, and based on several interrelated metrics.

2.3 Proposed Methodology

This work aims at analyzing the impact of the urban structure on city criminal events, providing insights to enhance the city planning in highly criminal areas. To achieve that goal, we propose a methodology with three phases: **(i) Mapping of Urban Crimes**; **(ii) Criminal Community Identification**; and **(iii) Crime Analysis**. This section describes such phases according to an application example.

2.3.1 Background and Datasets

We refer to an urban street network as a directed georeferenced graph $G = \{V, E\}$ composed of a set E of $|E|$ edges (street segments) and another set V of $|V|$ nodes (streets intersections). We refer to an edge $e \in E$ as an ordered pair $\langle i, j \rangle$, $i \in V$ and $j \in V$, in which i is named *source* and j is named *target*. Each node $v \in V$ has coordinates l_{at} and l_{on} , where l_{at} denotes the node's latitude and l_{on} the longitude. We refer to a set C , $C \cap V = \emptyset$, of crimes, each with coordinates l_{at} and l_{on} within the area of the city. Each

crime c has a value that belongs to a domain of 39 crime types — the most common types are *assault*, *theft*, and *minor crimes*.

Graph Source

Our approach uses electronic cartographic maps extracted from the *OpenStreetMap* (OSM) [56] platform. OSM provides maps representing spatial elements, *e.g.* points, lines, and polygons, as objects. Such objects, which are abstractions from the real-world geographical space, are represented by a triple of attributes that uniquely identifies them in their set, granting each object an identification number and georeferenced coordinates. These objects allow representing OSM data as a georeferenced graph.

Crime Source

The crime dataset C comes from the *San Francisco OpenData* (SFO) initiative ¹, a central clearinghouse for information about the city of San Francisco – CA, USA. SFO provides data concerning public issues. We use a set of events that corresponds to the period between Jan 1st, 2003 through May 5th, 2016. It has 1,916,911 instances of 39 crime types.

2.3.2 Mapping of Urban Crimes

The first phase of our methodology receives the crime dataset as input, denoted as set C , where each element $c \in C$ corresponds to a crime event occurred in a specific latitude and longitude. We proceed by mapping the crime events to the nodes of the urban street network, represented as the graph $G = \{V, E\}$. Given a crime c , its corresponding network node is the closest node v , $v \in V$, as given by the Euclidean distance calculated from the coordinates of c and v . This distance refers to the real length between elements through the Earth's surface, which is derived from the law of cosines and defined as:

$$d_{ij}^E = \mathcal{R} \times \cos^{-1}(\sin(l_{at}^i)\sin(l_{at}^j) + \cos(l_{at}^i)\cos(l_{at}^j)\cos(\Delta_{ij}^{lon})) \quad (2.1)$$

where l_{at}^i and l_{at}^j are latitudes, Δ_{ij}^{lon} is the difference between longitudes l_{on}^i and l_{on}^j , $\mathcal{R}$ is the earth's radius (6,371km), and d_{ij}^E denotes the Euclidean distance (E) between points i and j . All values are represented in radians [57]. With respect to the georeferenced network, d_{ij}^E also represents an edge's weight, which refers to the street length (in meters) that connects the streets' intersections (*i.e.* nodes) i and j .

Next, we enrich the network representation with a new property, making it reflect the geographical distribution of crimes. The result is a new set of nodes V' , with each node $v \in V'$ having a set with three properties $prop(v) = \{l_{at}, l_{on}, |C_v|\}$, where l_{at} is the latitude, l_{on} is the longitude, and $|C_v|$ is the quantity of crimes mapped to node v .

¹ Available on <data.sfgov.org>.

Formally, the set of crimes mapped to a node v , denoted as C_v , is defined as:

$$C_v = \{c \in C \mid d_{vc}^E < d_{wc}^E, \forall w \in V, v \neq w\} \quad (2.2)$$

Note that this phase is performed separately for each kind of crime. For example, if performing the mapping for crime type *assault*, we get a version of the graph in which the nodes have a property indicating the number of crimes that are categorized as an *assault*. After processing each kind of crime, we get a new property in the nodes. In a mapping perspective, each processing yields a new graph that differs with respect to the properties of the nodes; formally, $\mathfrak{G} = \{G_0, G_1, \dots, G_n\}$, where each $G = \{V', E\}$, $\forall G \in \mathfrak{G}$.

2.3.3 Criminal Community Identification

Intuitively, a community is understood as a group of nodes that have a probability of connecting to each other that is greater than the probability of connecting to nodes out of the group [58]. Formally, consider $G = \{V', E\}$ as a city graph and $p(i, j)$ as the probability of nodes i and j to define a connection; a community set is understood as a subset of nodes $\mathfrak{C}$. A node i is considered a member of the community $\mathfrak{c} \in \mathfrak{C}$ if its probability of connecting to a node $j \in \mathfrak{c}$ is greater than the probability of connecting to a node j' in another community $\mathfrak{c}' \in \mathfrak{C}$, as follows:

$$p(i, j) \geq p(i, j') \mid i \notin \mathfrak{c}, i \notin \mathfrak{c}', \forall j \in \mathfrak{c}, \forall j' \in \mathfrak{c}', \mathfrak{c} \neq \mathfrak{c}'$$

The elements of a community define a connected induced subgraph $G(\mathfrak{c})$ in which each node is reachable from all other nodes of the same community. Formally, $G(\mathfrak{c}) = \{V'', E'\}$, $V'' \subseteq V' \iff \{\forall i \in \mathfrak{c}, \forall j \in \mathfrak{c} \mid d_{ij} < \infty\}$, where $d_{ij} : V' \times V' \rightarrow \mathbb{R}$ is an arbitrary distance function which returns the distance from any given pair of nodes.

Next, we proceed with the identification of communities. To identify them concerning each crime type, we opted for the tool Nerstrand [59], which carries out a fast multi-thread detection. This tool, which yields results broadly accepted in the literature, identifies communities without any information about how many of them exist in the city. The input for this phase is an undirected graph, whereas the output is a set of number-labeled communities. The tool allows the user to attach weights to the graph elements; for our application, the edges' weights are the distance between two distinct nodes, and nodes' properties are the number of mapped crimes.

Our dataset comprises 39 crime types, and our techniques and results are generalizable for any number of crime types. For those reasons, we worked only with the three most recurrent ones, which comprise almost 40% of our dataset and represent: (i) assault, *i.e.* to inflict injury on a person intentionally; (ii) theft, *i.e.* to take a property of a person's possession by stealth, with no brute force; and (iii) minor crimes, such as offenses that do not involve any loss or injury. They represent, respectively, 167,832 (8.76%), 392,338

(20.47%) and 204,451 (10.67%) crimes. As previously explained, for each crime type, we perform a mapping over the graph, each one with a different number of crimes mapped into their nodes. Further on, we discuss that each crime-mapped graph can produce a specific criminality scenario; that is, when considering each crime type, we get a different set of communities, each one characterized by that kind of crime and by the topology of the network.

The next step, after identifying communities, is to filter out the less relevant ones for each type of crime. From the perspective of criminality analysis, relevance here means choosing the highly criminal communities. This approach reduces the excessive number of identified communities, as well as characterizes the most relevant ones. Accordingly, we worked with the *top five highly criminal communities*, which were identified by using two variables: (i) the highest criminal average, *i.e.* the average number of crimes per node; (ii) community size threshold, which prioritizes communities containing at least a specified number of nodes. For the purpose of this work, the threshold has been set to 100 nodes, since communities with at least 100 nodes are spatially well distributed, and can better represent the regions of the city. This filtering process is depicted in Figure 1, which shows the average criminality on the x -axis and the number of nodes on the y -axis. In the figure, we describe the less relevant communities as the mischaracterized ones, and we show where the most relevant ones are, considering the previously defined threshold.

Figure 1 – The filtering process used to identify the relevant communities while discarding the mischaracterized ones. We sort the communities by their average criminality, and we eliminate the ones that have less than 100 nodes. The highlighted area indicates the communities that are, at the same time, highly populated and highly criminal.

2.3.4 Crime Analysis

In this last phase, with the communities already detected, we proceed to analyze their interrelationship and interaction. Therefore, here, we consider the communities in pairs of types; for example, we analyze communities of assault and theft to verify whether they occur together or if one type of crime makes the other more intense. The idea is to obtain insights about their geospatial dependency, pointing to regions that share space of more than one crime type and, consequently, are potentially more dangerous. To this end, this phase has two stages. The first stage obtains the geospatial similarity between pairs of crime types, which allows determining how close are their communities and how related are their criminal activity. The second stage computes the Homogeneity and Completeness scores, which are useful to understand the crimes' behaviors by evaluating specific aspects of each community set. Combined, both stages lead to a more meaningful analytical scenario, accomplishing the contributions mentioned at the end of Section 2.1.

Measuring the Similarity of Communities

The similarity enables us to characterize the relationship of distinct events based on their spatial behavior, more specifically their spatial distribution. Such measure is based on their distances and considers the possibility of a non-criminal region containing crimes as well because a non-criminal area may have its behavior affected by a close criminal area.

We proceed with the computation of the similarity measure by considering the previously detected top five highly criminal communities. The similarity of crime events is based on the distance between elements belonging to any two sets of communities $\mathfrak{E}$ and $\mathfrak{F}$, where each one represents a distinct crime type. Note that, at this point, after communities were detected based on the topology of the city, we proceed using only the nodes and their positioning — *i.e.* we do not consider the edges nor the city topology for this processing. Such additional information is not necessary since we consider only the position of highly criminal communities in order to characterize the geospatial relationship of crime types. Hence, we define the similarity $\mathcal{S}_{\mathcal{K}}(\mathfrak{E}, \mathfrak{F}, \mathfrak{C}) \in [0, 1]$; the closer to 0, the more spatially apart are the sets of nodes:

$$\mathcal{S}_{\mathcal{K}}(\mathfrak{E}, \mathfrak{F}, \mathfrak{C}) = 1 - \frac{\sum_{\forall u \in \mathfrak{E}} \sum_{\forall v \in \mathfrak{F}} d_{uv}^E}{\max(\sum_{\forall i \in \mathfrak{E}} \sum_{\forall j \in \mathfrak{C}} d_{ij}^E)} \quad (2.3)$$

Such measure identifies, at the same time, the similarity between intra- and inter-community elements. The result 1 denotes that the elements within community $\mathfrak{E}$ are highly close to each other and to the elements of community $\mathfrak{F}$, intra- and inter-similar respectively, and 0 denotes the opposite.

Identifying the Behavior of Crimes

The characterization of communities identified in a city can be complemented by the Homogeneity and Completeness scores. They are derived from an entropy-based cluster evaluation [60], which works by analyzing the presence and absence of criminality in the

network's communities. Such scores quantify how intensely each crime type manifests across a community set $\mathfrak{C}$. To this end, the Homogeneity score *evaluates the criminality* in a community by identifying the criminal distribution among their nodes. Further, the Completeness score *quantifies how intrinsic is the criminality* in a community by measuring their distribution among all the communities, both regarding a specific crime type.

2.3.4.1 Homogeneity

Regarding the top-five criminal communities, it is desired to know what are their predominant aspects. For the purpose of achieving a better assessment of criminal communities, the Homogeneity score is able to measure how uniform, *i.e.* homogeneous, a community is. It is defined by $\mathcal{H}(\mathfrak{C}) : \mathfrak{C} \times \mathfrak{C} \rightarrow [0, 1]$ as follows:

$$\mathcal{H}(\mathfrak{C}) = 1 - \left\{ \left[\sum_{i=1}^{|\mathfrak{C}|} \sum_{j=0}^{|Q|} \frac{|\mathfrak{C}_{i,j}|}{|V'|} \times \log_2 \left(\frac{|\mathfrak{C}_{i,j}|}{|\mathfrak{C}_i|} \right) \right] \div \left[\sum_{k=0}^{|Q|} \frac{|\mathfrak{C}_k|}{|V'|} \times \log_2 \left(\frac{|\mathfrak{C}_k|}{|V'|} \right) \right] \right\} \quad (2.4)$$

where $|V'|$ is the number of nodes in all the communities of set $\mathfrak{C}$ (the top five in our example) of a crime type, and $Q = \{1, 0\}$, which, in our scenario, indicates the nodes with and without crimes, respectively. Even if calculated separately, both scenarios need to be considered to measure the Homogeneity score of a community set. We used $\mathfrak{C}$ to denote the set that contains all communities and $\mathfrak{C}_i \in \mathfrak{C}$ to indicate a single community. Finally, $\mathfrak{C}_{i,j}$, $j \in Q$, represents nodes in community i in which crimes are present or absent, whereas $\mathfrak{C}_k$, $k \in Q$, represents the same idea but among all communities. For instance, $\mathfrak{C}_k$, $k = 0$, represents all nodes in all communities that have no crime, while $\mathfrak{C}_{i,j}$, $i = 1$, $j = 1$, indicates the nodes from community 1 that have crime(s).

The Homogeneity score equals 0 when the presence and absence of crimes are proportionally similar across all the communities. Contrarily, the value 1 denotes that all communities in $\mathfrak{C}$ have crimes in all their nodes or no crime at all.

It is noteworthy that the Homogeneity score solely is insufficient to provide the needed comprehension about the criminal behavior in a city. This is because a single community can have all its nodes characterizing either crime occurrences or no crime at all, which, in turn, could yield the same Homogeneity score. For this reason, we also employ the Completeness score that, combined with the Homogeneity score, can better describe the distribution of crimes across a city. By analyzing both scores, we can effectively compare different areas, regardless of their high or low criminality. Their purpose is to identify the communities' behaviors, mainly their tendency for high/low scores to a given crime type.

2.3.4.2 Completeness

If crimes do not concentrate in a community, it is likely that there are multiple communities with criminal nodes spread across the city, demanding surveillance at multiple criminal

spots. Alternatively, if crimes concentrate, they tend to be grouped within a single community, requiring police patrols on specific regions. To determine if crimes are more or less concentrated, we employ the Completeness score, which is directly obtained from the whole community set. The Completeness score equals 0 when the presence of crimes is totally scattered, *i.e.* equally absent across all communities — their occurrence is totally distributed. On the other hand, the higher the value, the more concentrated are the criminal and safe zones. This is defined in $\mathcal{C}(\mathfrak{C}) : \mathfrak{C} \times \mathfrak{C} \rightarrow [0, 1]$ as:

$$\mathcal{C}(\mathfrak{C}) = 1 - \left\{ \left[\sum_{j=0}^{|Q|} \sum_{i=1}^{|\mathfrak{C}|} \frac{|\mathfrak{C}_{i_j}|}{|V'|} \times \log_2 \left(\frac{|\mathfrak{C}_{i_j}|}{|\mathfrak{C}_j|} \right) \right] \div \left[\sum_{k=1}^{|\mathfrak{C}|} \frac{|\mathfrak{C}_k|}{|V'|} \times \log_2 \left(\frac{|\mathfrak{C}_k|}{|V'|} \right) \right] \right\} \quad (2.5)$$

Hence, if safe and criminal zones are totally apart, the Completeness score of a community set equals 1. Conversely, if criminality has a proportionally equivalent participation in all communities, the score is 0.

2.4 Results and Discussions

This section presents and discusses the results of the experiments regarding the three phases of our methodology over the city of San Francisco.

Figure 2 – The most highly criminal region from San Francisco – CA, USA, identified by evaluating the crime density in the city from 01/01/2003 to 05/18/2016. Each colored area represents the occurrence of a crime type or the intersection of a crime type with another type of crime. In the legend, each CM (from 1 to 3) denotes a type — *assault*, *theft* and *minor crimes*.

Influence of Crime Mapping

The first experiment aimed at analyzing the impact of the urban crime mapping (first phase of our methodology) on the criminal community identification process. This process was carried out twice: once considering the topological graph, *i.e.* the graph that considers

only nodes and edges, and once over the network after the urban crime mapping, regarding the newly-included edges' weights (distances in meters) and nodes' weights (number of crimes). As a result, we obtained the number of 134 communities from the topological graph and, when considering the information introduced by the crime mapping phase, the numbers of communities were: 210 for *assault*, 215 for *theft* and 211 for *minor crimes*.

As a first result, we noticed the discrepancy between the numbers of criminal communities for each graph. Such difference is due to the presence/absence of crime-related data derived from the crime mapping phase. The Nerstrand algorithm considers the crime-related data (the nodes' and edges' weights), ensuring that the nodes within each identified community share similar characteristics. This behavior is not the same when considering the topological graph. Since there is no crime information attached to its nodes and edges, the algorithm examines only the connections between nodes. Therefore, considering only the network's topology, there is no guarantee that the crimes will be optimally grouped among the identified communities. These findings show that identifying communities through a topological graph is an inaccurate way to represent crime relationships in a city.

Communities' Design Aspects

The second experiment aimed at comparing the results from the previous analysis by considering their design aspects. Table 1 presents the most highly criminal communities obtained directly from the network's topology and from the three graphs generated via the Crime Mapping process. The table shows twelve communities sorted by crime average, among which are the top five criminal ones for each type.

Table 1 – Communities with the highest crime average, identified from: (i) *topology*, (ii) *assaults*, (iii) *theft*, and (iv) *minor crimes*. Column *Average* denotes the crime average and *Total* denotes the number of nodes in each community.

	Communities identified from:							
	<i>Topology</i>		<i>Assault</i>		<i>Theft</i>		<i>Minor Crimes</i>	
	Average	Total	Average	Total	Average	Total	Average	Total
00	1063.00	2	362.00	2	4030.00	3	545.92	12
01	157.67	788	230.59	17	245.00	2	457.00	2
02	116.30	802	55.26	316	201.00	17	356.41	17
03	110.75	4	47.44	226	60.15	281	281.91	67
04	108.80	5	41.00	2	49.74	277	140.13	142
05	77.93	515	34.69	87	47.08	156	110.15	188
06	77.50	2	34.06	149	37.88	66	73.80	5
07	65.86	539	29.53	238	32.25	4	72.99	143
08	44.17	744	27.42	142	25.92	296	71.37	286
09	43.88	1129	22.14	248	25.77	342	68.50	4
10	36.28	635	15.52	281	21.79	307	58.72	32
11	35.88	518	14.20	54	21.10	371	54.43	144

The table describes a higher crime average in communities derived only from the network topology. Such higher average occurs because, in this context, there is a lower number of communities, each one containing a higher number of criminal nodes.

As shown in the table, the number of crimes per community still stands out when comparing the graph from topology and the “complete” graphs — *i.e.* the graphs generated by the Crime Mapping process. This is because the complete graph used no attribute that characterizes a crime in the community identification process, so there is no way to assure that they are related just by being in the same community. Therefore, hereafter we discuss only the graphs with crime-related data.

Regarding such graphs, Table 1 shows a significant number of communities with few nodes. This is evidence that the crimes are not evenly distributed in the network. Therefore, if someone eventually attempted to eradicate or prevent crimes in the entire city based on characteristics of a local region (a single community), such attempt would fail because the crimes are not concentrated in a single criminal community.

San Francisco’s Crime Neighborhood

To allow analyzing a group of communities containing a significant number of crimes, we have selected the region from the city of San Francisco that comprises the highest crime indices. This region contains the five communities, as mentioned earlier, of each crime type, which are among the twelve communities shown in Table 1. The top five communities of each type also meet the requirement regarding the number of nodes, previously determined as a threshold of 100 nodes (see Section 2.3.4). Specifically, considering the table, the top five communities for *assault* are in lines 2, 3, 6, 7 and 8; the top five for *theft* are in lines 3, 4, 5, 8 and 9; and the top five for *minor-crimes* are in lines 4, 5, 7, 8 and 11. Figure 2 depicts the entire selected area of San Francisco, as well as highlights each community (and each geospatial intersection of communities) with respect to their crime types.

To quantify the relationship between communities in the city space, we used Equation 2.3 to obtain the similarity between communities. Through this measure, we compared each pair of crime types, computing the similarity between community sets $\mathfrak{C}$ of one type and community sets $\mathfrak{F}$ of another type. By doing so, we achieved a similarity value $\mathcal{S}_K(\mathfrak{C}, \mathfrak{F})$ of 0.66 for assault *vs.* theft, 0.65 for assault *vs.* minor crimes and 0.58 for theft *vs.* minor crimes. Such results suggest that there is a relationship between crimes of different types, especially when comparing assault *vs.* theft and assault *vs.* minor crimes, since the values, 0.65 and 0.66 can be considered significant enough if one ponders over what these numbers represent — that the majority of the area of these two communities might be more dangerous due to the occurrence of more than one crime type.

More specifically: **(i)** it is very likely that criminal activities occur at intersections of multiple crime types, which we call *crime hubs* — regions that attract or disseminate criminality; **(ii)** those crimes are spatially related, so it is possible that they have been

caused by the same people or gang; (iii) places, where more than one crime type occurs, are high-priority regions for public safety improvement.

Communities' Intrinsic Criminality

The previous section discussed the geospatial disposition of different crime types with respect to intersections between criminal regions. This section, in turn, brings results concerning the Homogeneity and Completeness scores for the communities in the most highly criminal area of San Francisco. This experiment aimed at measuring how intrinsic was the criminality within the neighborhoods of San Francisco during the studied period. To understand this measure, recall that the number of crimes of some nodes is zero. For the *assault* type, there are 1,071 nodes, of which 132 (12.33%) contain at least one crime and 939 (87.68%) contain none. For *theft*, there are 903 nodes, of which 133 (14.73%) are criminal and 770 (85.27%) are not criminal. For *minor crimes*, there are 1,352 nodes, of which 177 (13.09%) present crimes and 1,175 (86.91%) do not. These values and percentages refer to the entire community set for each crime type.

The presence and absence of crimes were considered to calculate the scores presented in Table 2, which shows that *assault* has the lowest values for both measures. Regarding Homogeneity, the value indicates that this crime type presents a reduced concentration of criminal nodes within each community. This is straightly inferable because, besides the low Homogeneity value indicating an inexpressive occurrence of crimes, the actual number of crimes is very low; a pattern that holds for other types of crime as well.

Table 2 – The Homogeneity and Completeness scores used to evaluate the quality of each community set and to quantify how intrinsic is the criminality within them.

Crime type	Homogeneity	Completeness
<i>Theft</i>	0.015588	0.004159
<i>Minor Crimes</i>	0.014014	0.003441
<i>Assault</i>	0.013557	0.003234

Since the value for Completeness is also the closest to 0, we can reliably assume the occurrence of crimes to be scattered across all the *assault* communities. The *theft* type has values close to 0 as well, but the highest ones when compared to the values of other crime types. The greater difference corresponded to the Completeness score, whose value suggests a geospatial crime concentration more elevated. Regarding

Completeness (see Section 2.3.4), such higher concentration tends to occur in a smaller set of communities rather than across all of them, which is more likely for the *assault* type.

A higher crime concentration means that a particular crime type is less likely to intersect with another crime type. This is because, in such a case, the criminal nodes would occupy a smaller area in the city space. Furthermore, such a crime type could be considered easier to prevent, since its occurrences would be in a more compact area of the city. On the other hand, a lower concentration of a crime type means that it exerts a

stronger influence on the network as a whole, being easier to propagate to remote areas and demanding global approaches both to eradicate it and to prevent it.

Remarks

We focused on analyzing the behavior of criminality in urban areas through complex-network tools. Due to the difficulty to acquire well-organized crime data, this study was limited to the city San Francisco, a very representative urban scenario. Nevertheless, it is noteworthy that our methodology is adequate for any real-world city, provided the data is available. We demonstrated that the distances between the nodes of a city and the data concerning criminal occurrences can be used in the task of analyzing the criminality of a region, which is done through computational and mathematical methods. Nonetheless, we are aware that criminality is a sociology problem to be further analyzed from the perspective of social sciences. This is the reason why we did not attempt to explain the reasons behind the criminal instances; rather, we traced a scenario from the perspective of the relationships between georeferenced interacting crimes. This was also the case of the work of Bogomolov *et al.* [55], which presents the closest methodology as compared to our approach. In their work, the authors do not take social variables into account; still, they are able to predict crime instances from demographic information, despite not explaining the reasons beneath. Complementarily, our work hypothesizes on how crime patterns are geographically related, bringing up another dimension to allow decision making over cities. All of our findings were compared, when possible, through traditional analytical methods applied over data made available by the San Francisco Open Data (DataSF)² initiative.

2.5 Conclusion

In this chapter, we proposed a threefold methodology for identifying and analyzing criminal spatial patterns in urban street networks, which refers to our first publication [37]. Our methodology is based on the hypothesis that *by employing network mapping techniques, allied to distance-based properties of graphs, it is possible to trace the relationship between areas that are highly criminal within a city*. To demonstrate our methodology, we analyzed criminal communities from real crime data representing *assaults*, *thefts* and *minor crimes* in the city of San Francisco – CA, USA. The methodology comprises the following phases: **(i) Mapping of Urban Crimes**, **(ii) Criminal Community Identification**, and **(iii) Crime Analysis**. The latter employs the well-established Homogeneity and Completeness scores to analyze the identified criminal communities more deeply. Our main achievement confirms the hypothesis of our work, allowing us to state that different crime types share

² <https://data.sfgov.org/Public-Safety/>

common spaces, characterizing areas that lack strategies for crime prevention, and that particular crime types are sparser than others in the city space.

The highlighted contributions are: **(i)** the use of a complex network to represent the real-world space, enabling an ample analysis of the city; **(ii)** independence from socioeconomic information, allowing the analyses of crimes based on their spatial disposition solely; **(iii)** the assessment of the impact of criminal regions, considering the similarity between distinct crime types and their Homogeneity and Completeness scores.

Finally, as future work, we intend to continue expanding our methodology developing tools to analyze crimes and their types considering a temporal perspective. Our aim is to provide insights about the criminal behaviors in a city, allowing to determine how a crime spreads throughout the city and when interventions were made by the governors were effective to prevent them, using time windows and comparing the results with socioeconomic data.

Behavioral Characterization of Criminality Spread in Cities

3.1 Introduction

Complex networks are proficient at representing complex systems, from neural networks to subway systems [61]. 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 [49]. 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 [50, 51] and analysing [52, 53] crimes.

In this chapter, we present our second work, which is stimulated by the discoveries that we presented in Chapter 2. Considering our previous contributions, we expanded upon the techniques used to map and analyze information from the North American city of San Francisco – CA, corresponding to our second publication [38]. This 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 highly 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.*

Considering that different crime types are related by being close to each other [37], 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.

3.2 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*. Our dataset consists of an electronic map of San Francisco — CA¹ and a list of 1,916,911 georeferenced criminal incidents from Jan 1st, 2003 to May 5th, 2016². To validate our methodology, we employed Homogeneity ($\mathcal{H}$) and Completeness ($\mathcal{C}$) scores, derived from an entropy-based cluster evaluation [60]. The early three phases have already been reported in Spadon *et al.* [37], and, therefore, we only detail the main contribution of this chapter, which focuses on the analyses of crime propagation in urban streets.

3.2.1 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.* [62], the Susceptible-Infected-Susceptible (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 $S \rightarrow I$, becoming infected as well. The latter represents the opposite, when an infected node recovers and becomes susceptible $I \rightarrow S$.

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, $N = S(t) + I(t)$ holds true for all values of t , where N is the number of nodes, $S(t)$ is the number of susceptible individuals, and $I(t)$ is the number of infected nodes. Given this notion, two deterministic equations can be derived. First, $\frac{dP_I}{dt} = +(pP_I)P_S - qP_I$ describes the infected nodes per time, whereas $\frac{dP_S}{dt} = -(pP_I)P_S + qP_I$ describes the susceptible nodes; where P_c , $c \in \{I, S\}$, represents the percentage of nodes that are infected or susceptible, pP_I is the force of infection, and p and q are the infection and recovering probabilities, respectively.

¹ Provided by <www.openstreetmap.org>.

² Available at <data.sfgov.org>.

Algorithm 1 – Dynamic Propagation of Urban Criminality.

```

1: procedure CRIMINALITYPROPAGATION( $G, C, l$ )  $\triangleright$  directed and weighted complex
   network, set of urban crimes, percentage of human interference
2:    $I \leftarrow \{ \}$ ;  $S \leftarrow \{ \}$ 
3:   for all  $v \in G(V)$  do
4:     if  $|C_v| > 0$  then
5:        $I \leftarrow I \cup \{v\}$   $\triangleright$  The seeds are marked as infected.
6:        $v_p \leftarrow 0.5$ ;  $v_q \leftarrow 0.5$   $\triangleright v_p$  is the infection probability;  $v_q$  is the recovering one.
7:     end if
8:     if  $|C_v| \leq 0$  then
9:        $S \leftarrow S \cup \{v\}$   $\triangleright$  The others are marked as susceptible.
10:       $v_p \leftarrow \emptyset$ ;  $v_q \leftarrow \emptyset$   $\triangleright$  Notice that such nodes are inert.
11:    end if
12:  end for
13:  for all  $i \in I$  do
14:    for all  $n \in \text{neighborhood}(i)$  do
15:      if ( $n \notin I$ ) and ( $\text{rand}([0, 1]) < i_p$ ) then
16:         $I \leftarrow I \cup \{v\}$   $\triangleright$  The node is marked as infected.
17:         $S \leftarrow S - \{v\}$   $\triangleright$  Also, it is unmarked as susceptible.
18:         $n_p \leftarrow i_p \times (1 - l)$   $\triangleright$  It inherits the probability of its infector
19:         $n_q \leftarrow 1 - n_p$   $\triangleright$  by considering the infection loss  $l$ .
20:      end if
21:    end for
22:  end for
23:  return  $I$   $\triangleright$  The algorithm returns a set of infected nodes.
24: end procedure

```

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 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.* [37]. 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.

3.3 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 3.2.1). 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 3 using $p = 0.5$.

Figure 3 – 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. The three scenarios of interest to this work are represented by the graphs on the top of the image, which are ordered from left to right. For comparison, our model uses $p = 0.5$.

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.

3.4 Conclusion

In this chapter, we propose a technique for assessing crime behavior in street networks considering the perspective of the criminality spread; such contribution constitutes our second publication [38]. 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.

Identifying Urban Inconsistencies via Street Networks

4.1 Introduction

Complex systems that exchange information through their entities' relationships can be described as complex networks. Such networks are derived from a variety of real-world systems, from neuronal connections to the distribution of power grids [61]. They can be used to represent cities, by considering the streets' intersection as entities and the streets as their relationships [13].

Through the extraction of network's features and the analysis of its properties, it is possible to computationally model urban problems [11], serving as a meaningful indicator for city planners. For instance, the identification of bottlenecks in the flow helps in the understanding of how a city works [63] and to decide where and how to improve its design. Additionally, such properties can indicate low-access regions, the ones that lack access *from* or *to* others in a city. Moreover, they can indicate potential places where facilities should be allocated or reallocated.

The motivation to portray a city as a complex network comes from the desire to explain how it's mobility works and how it influences the city's livability. In this regard, several works have focused on such problem [2], but until now, there has been a lack of a clear, transparent and precise methodology for employing complex networks in a complete process of analyzing the city, *i.e.* from data preparation to an in-depth analysis from a geometrical perspective. Furthermore, the former works lack techniques to assist in the planning and design process, by not being able to identify low-access regions that are critical to the displacement in a city.

In this sense, we propose a methodology that uses topological and geometrical measures in the task of analyzing the location and access to different types of facilities

in a city. The contributions here described refer to our third publication [39], which discusses techniques to detect urban planning inconsistencies. The hypothesis of this work is that *geometric and geographic measurements derived from street networks can aid in the identification of urban inconsistencies observed on ill-designed urban structures*. Our assumption is that the mobility is derived from the access indices of a city, which can be identified by measuring the distances among the locations of the city facilities. Based on those assumptions, our work introduces the following contributions:

1. **Scalability:** we contribute to the extraction, preparation, preprocessing, and processing of street networks, in ways beyond those of former works;
2. **Methodology:** we introduce methods, based on the network topology and geometry, which recognize low-access regions that are critical to the city planning and design process; and,
3. **Real-world application:** we quantitatively characterize inconsistencies of the Brazilian city of Sao Carlos, discussing their meanings and implications.

The remainder of this chapter is organized as follows. Section 4.2 presents related works that are relevant to our proposal. In Section 4.3, we introduce our methods, demonstrating how to evaluate the access by identifying the inconsistencies of a city. The results are presented in Section 4.4, where we show how to account critical regions, discussing how to improve access to the majority of the nodes in a real-world city. Finally, the conclusion and final remarks come in Section 4.5.

4.2 Related Work

Preceding our contributions, in this section, we discuss related work comparatively relevant to our proposal. In Subsection 4.2.1, we present works that focus on the network's properties to identify cities' characteristics, whereas Subsection 4.2.2 introduces the ones that concentrate on the cities' patterns to enhance their mobility. Finally, other relevant related work came in Subsection 4.2.3.

4.2.1 Urban Space Characterization

The motivation for our methodology is the desire to identify low-access regions through complex networks. Such regions, if taken into account in the planning and design process, can enhance decision-making activities and, consequently, improve the mobility of a city as a whole.

In this sense, Porta, Crucitti, and Latora [13] performed an in-depth investigation of centrality measures in cities. They introduced a methodology aiming to identify urban

structures that impact upon spatial cognition and collective behaviors. In a different perspective, our scope covers topological and geometrical analyses to characterize urban spaces by considering a set of referential nodes, *i.e.* facilities, of a city.

Strano *et al.* [34] investigated the network's geometrical attributes, highlighting the fact that cities share structural similarities due to their quasi-planarity. Whereas, Scellato *et al.* [64] explained how it is possible to identify routes that affect retails, land-use separation, and collective behaviors. However, none of them was proficient in identifying regions that lack access when compared to others, the desired result to support the planning and the design.

4.2.2 Traffic Planning and Mobility Enhancement

The desire to characterize the urban scenario comes from daily issues such as slowness, traffic jams, and lack of access. Such issues impact on the citizens' lives by negatively affecting their mobility. Despite the notable studies on the topic, little has been done to enhance low-access regions. They focused on the flow of the streets, while the access criterion was put aside.

This is the case of Costa *et al.* [10] that evaluated the effectiveness of the underground systems in connecting scattered places. They considered that citizens use different kinds of transportation to move along a city; in contrast, our work is based on a sole transportation mean. This is because we consider transportation as part of a service, which is allocated in a facility and demands displacement between a pair of vertices by the streets of a city.

Viana and Costa [22] examined the importance of fast long-range connections between two remote locations, aiming to improve the time displacement in a city. These connections are far larger than what is reasonable in the urban space because long-range streets are attached to geographical and financial constraints, which would require different research approaches.

4.2.3 Other Related Work

The analysis of ill-designed urban structures and its related problems is not limited to the previous works. For instance, Crucitti, Latora, and Porta [65] identified common characteristics between self-organized and planned cities. Zhang, Zeng, Jia and Li [66] focused on the optimal placement of a set of facilities to provide the lowest travel time to all network nodes, while Zhang, Cao, Liu and Huang [67] worked in the development of an optimal facility location plan, and also, there is the work of Travençolo and Costa [23] that describes the accessibility in the context of complex networks. In this regard, our proposal consists in using topo-geometrical measures to characterize low-access regions,

contributing to state-of-the-art on street analysis.

4.3 Proposed Methodology

Our methodology starts with formal notations, which are required for the understanding of our proposal, followed by the description of our data. Next, in Subsection 4.3.2 we introduce a novel pre-processing technique. In Subsection 4.3.3 we characterize the concept of urban inconsistency and, subsequently, in Subsection 4.3.6, we describe its categories. Finally, in Subsection 4.3.7 we describe our algorithm to evaluate such inconsistencies considering a city in the form of a street network.

4.3.1 Background and Datasets

Required concepts

Along the text, we refer to a distance-weighted directed graph $G = \{V, E\}$ as a complex network ¹, which is composed of a set V of $|V|$ nodes, and another set E of $|E|$ edges. An edge $e \in E$ is an ordered pair $\langle u, v \rangle$, in which $u \in V$ is titled *origin* and $v \in V$ is titled *destination*, $u \neq v$. By being a *street network*, its elements are georeferenced, and, therefore, each node $v \in V$ has two coordinates $\{\mathcal{L}_{at}^v, \mathcal{L}_{on}^v\}$, the latitude and longitude, respectively. Also, we use the term *primary* to refer to a network derived from a map, and *auxiliary* to define a post-processed network. Furthermore, *mobility* and *access* are terms used to denote the displacement capability in a city. The term *path* refers to an ordered set of distance-weighted directed edges that connects two nodes, such path has the length corresponding to the weighted sum of its edges. Still, the path of minimum length, also between two nodes, which is called as shortest path, is referred here as *network distance*.

Data source

Our work benefits from electronic cartographic maps that are extracted from the *OpenStreetMap* ² (OSM) platform. OSM provides files that represent the spatial elements of a city as points, lines, and polygons. Such elements are abstractions of the real-world and are composed of an identification number and georeferenced coordinates.

4.3.2 Preprocessing Technique

Given a map provided by OSM, the first step is to derive a street network from it. To do so, we considered streets as edges and street intersections as nodes. In addition, we

¹ Despite different, in this chapter, the terms graph, complex network and street network, are equivalent.

² Available in www.openstreetmap.org.

Algorithm 2 – Redundancy Removal.

```

1: procedure REDUNDANCY REMOVAL( $G$ )  $\triangleright$  directed and weighted complex network
2:   for all  $v \in V$  do
3:     if  $k_{in}(v) = 1$  and  $k_{out}(v) = 1$  then
4:        $i \leftarrow get\_in\_neighbor(v)$ 
5:        $o \leftarrow get\_out\_neighbor(v)$ 
6:       if  $\{ \langle i, o \rangle \notin E \mid i \neq o \}$  then
7:          $E \leftarrow E \cup \{ \{i, o\}, d_{iv}^E + d_{vo}^E \}$ 
8:          $V \leftarrow V - \{v\}$ 
9:       end if
10:    end if
11:    if  $k_{in}(v) = 2$  and  $k_{out}(v) = 2$  then
12:       $I \leftarrow get\_in\_neighborhood(v)$ 
13:       $O \leftarrow get\_out\_neighborhood(v)$ 
14:      if  $\{ \langle i, o \rangle \notin E \mid \forall i \in I, \forall o \in O, i \neq o \}$  then
15:         $E \leftarrow E \cup \{ \{o, i\}, d_{iv}^E + d_{vo}^E \} \cup \{ \{i, o\}, d_{ov}^E + d_{vi}^E \}$ 
16:         $V \leftarrow V - \{v\}$ 
17:      end if
18:    end if
19:  end for
20:  return  $G$ 
21: end procedure

```

pondered the edges' weight with their real-world distance in meters. Later, we noticed that the resulting network was filled with redundant elements due to the geometrical characteristics of the city.

This is the case, for example, of roundabouts, which are translated into polygons with a high number of nodes and edges. The redundant data usually raise the required resources to process these networks, especially the ones that represent capital cities. Besides, street segments that are not relevant negatively influence algorithms that depend on the topology of the network. Therefore, to scale down the network, we designed an algorithm that removes redundancies at the same time that it maintains the network's characteristics.

Our algorithm merges nodes that do not deprive the network of its primary feature, *i.e.* the distances of its paths. The algorithm starts by identifying nodes with total degree two and others with total degree four — in a street network, the degree of a node measures the number of incoming and outgoing streets. Such nodes correspond to the ones in the middle of a directed path. We consider the selected nodes as redundant if they are transitional, that is, if they simply break a street into segments without defining a street intersection; and, if there is not an edge that connects the nodes that communicate with the transitional one. Hence, when the transitional node is removed, a new edge is created to communicate the nodes that the transitional one was connected with. These steps are algorithmically described by Algorithm 2.

The algorithm is linear in the number of nodes $O(|V|)$. Considering that nodes and edges are removed within the iterations of the algorithm, it cannot be parallelized. However, it runs in less than a second over a network of a 200,000-inhabitant's city.

4.3.3 Identification of Urban Inconsistencies

The aim of our methodology is to characterize urban spaces of a city from the point of view of the location of its facilities. To this end, we start by tracking a set $\mathbb{F}$ of facilities, and, subsequently, identifying two sets of nodes in the surroundings of each facility $f \in \mathbb{F}$.

4.3.4 Perimeter set

The first set corresponds to the closest nodes to a facility f according to the geodesic (Euclidean) distance, as defined by Equation 4.1. Such distance d_{ij}^E is measured between nodes i and j over the Earth's surface, which is derived from the *Spherical Law of Cosines*.

$$\mathbf{V}_f^E = \{v \in V \mid d_{vf}^E < d_{vf'}^E, \forall f \in \mathbb{F}, f' \in \mathbb{F}, f \neq f'\} \quad (4.1)$$

The geodesic distance by means of georeferenced coordinates is defined by Equation 4.2 considering nodes i and j , where $\mathcal{L}_{at}^i$ and $\mathcal{L}_{at}^j$ are the latitudes, $\Delta_{\mathcal{L}_{on}^{ij}}$ is the difference between the longitudes $\mathcal{L}_{on}^i$ and $\mathcal{L}_{on}^j$, $\mathcal{R}$ is the earth's radius (6,371 km), all values expressed in radians.

$$d_{ij}^E = \mathcal{R} \times \cos^{-1}(\sin(\mathcal{L}_{at}^i)\sin(\mathcal{L}_{at}^j) + \cos(\mathcal{L}_{at}^i)\cos(\mathcal{L}_{at}^j)\cos(\Delta_{\mathcal{L}_{on}^{ij}})) \quad (4.2)$$

Notice that, given the set of network nodes V and a set of facilities $\mathbb{F}$, every node $v \in V$ pertains to the perimeter set of a single facility $f \in \mathbb{F}$. That is, according to Equation 4.1, the sets $\mathbf{V}_f^E$ and $\mathbf{V}_{f'}^E$, $\forall (f, f') \in \mathbb{F}, f \neq f'$, are mutually disjoint.

4.3.5 Network-distance set

The second set corresponds to the nodes closest to a given facility f via network-distance from any node v to f , and it is defined by Equation 4.3, where d_{ij}^N is the network distance from node i to node j . Notice that, the sets $\mathbf{V}_f^N$ and $\mathbf{V}_{f'}^N$, $\forall (f, f') \in \mathbb{F}, f \neq f'$ are mutually disjoint; that is, a node $v \in V$ can be network-closest to one facility $f \in \mathbb{F}$ at a time.

$$\mathbf{V}_f^N = \{v \in V \mid d_{vf}^N < d_{vf'}^N, \forall f \in \mathbb{F}, f' \in \mathbb{F}, f \neq f'\} \quad (4.3)$$

In view of the fact that the network-distance set is directed, the *network-distance set TO a given facility f* might not be the same as the *network distance set FROM a given facility f* . Hence, we define the network-distance set *from* a facility $f \in \mathbb{F}$ to the nodes in V as:

$$\bar{\mathbf{V}}_f^N = \{v \in V \mid d_{fv}^N < d_{fv'}^N, \forall f \in \mathbb{F}, f' \in \mathbb{F}, f \neq f'\} \quad (4.4)$$

The essence of our work is that the nodes that are closest to a facility according to the network distance, and at the same time, are not closest to this same facility according to the geodesic distance are evidence of network inconsistencies. These are considered inconsistencies because the network is supposed to render streets that provide access through shortest distances, which becomes a false premise due to ill-conceived urban decisions. In the following, such inconsistencies are defined by set operations, which characterizes categories of inconsistencies.

4.3.6 Categorizing Urban Inconsistencies

In the whole urban scenario, there are three displacement cases used to categorize groups of inconsistencies, which are all based on the flow direction. In the *first* one, the flow is from the node v to the facility f ; in the *second*, the flow is from the facility f to the node v ; and, in the *third* case, the flow is mutual between v and f . These three situations lead to the following inconsistencies definitions that are based on the directed characteristic of the street network:

1. **Inward Inconsistency:** nodes that are geodesic-closest to a facility, but network-closest (from v to f , as given by Equation 4.3) to another; this case corresponds to the set given by:

$$\Phi_f^I = V_f^E - V_f^N \quad (4.5)$$

2. **Outward Inconsistency:** the same as an inward inconsistency, but considering the opposite direction (from f to v , as given by Equation 4.4):

$$\Phi_f^O = V_f^E - \bar{V}_f^N \quad (4.6)$$

3. **Absolute Inconsistency:** nodes that are, simultaneously, inward and outward inconsistencies, which corresponds to the following intersection:

$$\Phi_f^A = \Phi_f^I \cap \Phi_f^O \quad (4.7)$$

Notice that, inconsistencies are nodes apart from the normality of the network mobility, as so, these nodes represent *anomalies* derived from the mobility scenario of a given city. Through the concept of urban inconsistencies and the definitions of its categories, it is feasible to devise an algorithm capable of recognizing such inconsistencies, as accomplished with Algorithm 3.

4.3.7 Inconsistencies Identification Algorithm

The idea of the algorithm is to locate low-access regions by considering the nodes geodesic and network distances, to a set of facilities $\mathbb{F}$ that are within the city. It is noteworthy

that our algorithm evaluates an inconsistency category at a time, and, therefore, we used $c \in \{I, O, A\}$ to denote an arbitrary category that is being analyzed. We use the variables f_n and f_e to denote the closest facility to a given node v considering the network and geodesic distances.

The functions *NetworkDistance* and *GeodesicDistance* are used to identify the closest facility between f_n/f_e and f , which are compared in lines 9 and 12. We confirm if a node is an inconsistency in line 16, where we check if the f -minimum by network distance (f_n) and the f -minimum by geodesic distance (f_e) are equal or not. If the node v is inconsistent, by being wrongly closer to f_n when it should be closer to f_e , v is an inconsistency of f_e — see line 17. In line 20, the algorithm returns as output a set of inconsistencies, which are grouped by facility $V_f^c \subseteq V^c$, $\forall f \in \mathbb{F}$. Furthermore, the sets V_f^c and $V_{f'}^c$, $\forall (f, f') \in \mathbb{F}$, $f \neq f'$, are mutually disjoint; the meaning of that is, a node is an inconsistency to one, and only one, facility.

Algorithm 3 – Inconsistency Identification.

```

1: procedure INCONSISTENCYIDENTIFICATION( $G, \mathbb{F}, c$ )
2:   for all  $f \in \mathbb{F}$  do                                      $\triangleright$  Initialize the empty inconsistency set of size  $\mathbb{F}$ 
3:      $V_f^c \leftarrow \{ \}$ 
4:   end for
5:   for all  $v \in V$  do
6:      $f_n \leftarrow \emptyset$                                       $\triangleright v \in V$  has no closest  $f \in \mathbb{F}$  yet
7:      $f_e \leftarrow \emptyset$ 
8:     for all  $f \in \mathbb{F}$  do
9:       if NetworkDistance ( $v, f, c$ ) < NetworkDistance ( $v, f_n, c$ ) then
10:         $f_n \leftarrow f$ 
11:      end if
12:      if GeodesicDistance ( $v, f$ ) < GeodesicDistance ( $v, f_n$ ) then
13:         $f_e \leftarrow f$ 
14:      end if
15:    end for
16:    if  $f_n \neq f_e$  then
17:       $V_{f_e}^c \leftarrow V_{f_e}^c \cup \{v\}$                       $\triangleright$  Node  $v$  should be closer to  $f_e$  than to  $f_n$ 
18:    end if
19:  end for
20:  return  $V^c$ 
21: end procedure

```

Algorithm 3 runs in $O(|V||\mathbb{F}|^3)$ in the average case, where $|\mathbb{F}|$ is the number of public services and $|V|$ is the number of nodes, $|\mathbb{F}| \ll |V|$. In the worst case, where $|\mathbb{F}| = |V|$, the complexity would be $O(|V|^4)$. It is straightly parallelized, running in a minute over a 200,000-inhabitant's city.

4.4 Results and Discussions

Our experiments start by validating the Redundancy Removal (RR) algorithm (see Subsection 4.3.2). This process precedes the identification of urban inconsistencies (see Subsection 4.3.3) because all the subsequent experiments are carried out by the post-processed, *i.e.* auxiliary, network.

4.4.1 Assessment of Redundancy Removal Algorithm

The algorithm comes to be useful if it does not deprive the network mobility. Such mobility is measured by the edge distance that connects all nodes of the network. In this sense, it is enough if we evaluate the total edge length of a city ($\mathcal{W}$), which is given by the summation of the lengths of all their edges. The summation must be the same in both primary and auxiliary networks to certify that the paths between the nodes do not change due to the removal of redundancies. It is clear that the network will have fewer elements, but the nodes that prevail will be connected to longer edges.

As explained before, networks with redundant elements require more resources to be processed, especially networks derived from capital cities. In this sense, in the first experiment, we preprocessed five cities to assess whether the lengths were conserved. The results are presented in Table 3, where $\mathcal{W}$ denotes the total edge length, in addition to the percentage (%) of redundancies removed, both in terms of the primary and auxiliary network.

Table 3 – Redundancy Removal in five cities, indicating that the length of the edges is the same in every case and exhibiting the percentage of elements removed during the process.

Country	Primary			Auxiliary			%	City
	Nodes	Edges	$\mathcal{W}$	Nodes	Edges	$\mathcal{W}$		
Abu Dhabi	658,550	1,173,246	74×10^6	181,444	434,428	74×10^6	72.44	<i>Dubai</i>
Calif. USA	355,094	691,466	35×10^5	95,703	243,330	35×10^5	73.04	<i>San Francisco</i>
Canada	802,406	1,633,681	66×10^6	203,175	525,883	66×10^6	74.67	<i>Toronto</i>
Denmark	58,126	116,301	37×10^5	20,627	46,289	37×10^5	64.51	<i>Aarhus</i>
England	2,013,291	3,969,028	13×10^7	572,405	1,286,465	13×10^7	71.56	<i>London</i>

When analyzing the data in the table, it is possible to see that the length is the same in every case; that is, no relevant information was lost, so the networks preserved their mobility properties by keeping the distance between their nodes. Also, it is possible to see that the majority of the nodes ($> 60\%$) of all the cities comes to be redundancies, which means that results acquired from primary networks are dubious. This is because many measurements rely on how the city is connected and in the distance of the streets that communicate their nodes, which cannot be correctly accounted via primary networks.

Considering now that the RR algorithm assures the mobility aspects of the network, we applied it to the city of Sao Carlos, a 200,000-inhabitant's city, which corresponds

to our real-world case study. The primary network contained 12,464 nodes and 26,378 edges; after processing, we got an auxiliary network with 5,984 nodes and 16,338 edges. The reduction was of 51.98% in the number of nodes, and of 38.06% in the number of edges, which means 57.46% of all elements. These processes are depicted regarding a city sub-region in Figure 4a and 4b.

Figure 4 – Redundancy Removal in a sub-region of the city of Sao Carlos – SP, Brazil.

4.4.2 Assessment of the Auxiliary Network

Afterward, to understand the impact of the algorithm to the properties of the graph we evaluated the characteristics of the auxiliary network. First, the results from degree distribution, depicted in Figure 5a and 5d, show that the number of nodes with degree 2 and 4, which refers to paths, were reduced, and the portion of nodes with degree 6 and 8 that form city blocks — regions that are surrounded by streets — were increased significantly. This is evidence of that the primary network is more propitious to an inaccurate degree characterization by the high number of redundancies.

In a similar fashion, in Figure 5b and 5e, we analyzed the mobility property by observing the average neighborhood connectivity. Each cell of the 3D-histogram indicates the number of edges with a given in-degree in the x -axis, and out-degree in the y -axis. It is possible to see that the mobility was enhanced because the main diagonal became darker after the preprocessing. The reason is that the auxiliary graph has fewer low-degree nodes, characterizing a higher probability of movement in fewer steps with the same weight.

Finally, Figure 5c and 5f shows the degree correlation of the networks, where is possible to notice that the degree of the nodes connected to others with degree 1 changed significantly in the auxiliary network. This is due to the removal of redundancies, which

causes the nodes that communicate with the redundant one to connect straightly to nodes with higher degrees.

All the analysis seen from above, are results that state that the RR algorithm reduces the number of redundant nodes at the same time that it reveals the real topology of the network. As a consequence, we discuss results from urban inconsistency just in the auxiliary network.

Figure 5 – City distribution and characteristics. (a)(d) The in-out/total degree distribution of the primary and auxiliary network. (b)(e) The average nearest neighbor degree correlation histogram. (c)(f) The average target neighborhood degree, considering the source node degree.

4.4.3 Identification of Urban Inconsistencies

At the last step, we tested the Inconsistency Identification Algorithm 3, over the auxiliary network derived from the city of Sao Carlos. To this end, we selected three kinds of facilities in the city. We chose facilities that have implications for public security, health and education, being respectively represented by police stations, hospitals, and schools. In addition, each one has a different type of assistance to citizens, so they represent different categories of inconsistencies. That being said, consider that for police stations, hospitals, and schools, we have inward, outward, and absolute inconsistency, respectively. These results are presented in Table 4, where each category of inconsistency was evaluated independently, so there is no influence between different kinds of facilities.

Figure 6 – The identification of urban inconsistencies in a general case and in a real-world scenario. Also, despite the shape of the images, the perimeters are always convex.

By analyzing the data, it is possible to conceive that the number of inconsistencies is directly related to the number of facilities. The reason is that inconsistencies are more probable to be allocated at the intersection of the perimeter boundaries. As a consequence, by having more facilities, there will be more intersections with the boundaries of the others, which will lead to an increased number of inconsistencies. A general case is illustrated in Figure 6a, where the inconsistencies are found where the perimeters intersect, while Figure 6b depicts a real case of inconsistency found in the city of Sao Carlos, which is related to a Hospital position.

Considering that inconsistency emerges within the intersection of different perimeters, we notice as well that they cannot be eradicated without altering the network topology. Nonetheless, the notion of urban inconsistency describes the lack of access in a city. Such inconsistencies can serve as guidance to enhance a city by improving its design. Hence, this scenario opens a challenge to find feasible locations to facilities that reduce, rather than eliminate, inconsistencies, and, at the same time, that provide a finer access to the population.

Table 4 – The inconsistencies from the city of Sao Carlos concerning public facilities (i) Police Stations, (ii) Hospitals, and (iii) Schools, detected through Inconsistencies Identification.

Urban Inconsistencies																	
<i>i-th Facility</i>	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	<i>Total</i>
Police St.	32	4	86	191	—	—	—	—	—	—	—	—	—	—	—	—	342
Hospital	13	2	12	19	30	49	145	39	12	43	72	95	28	—	—	—	559
Schools	15	77	43	71	114	3	8	15	78	51	38	15	56	8	60	11	663

4.5 Conclusion

In this chapter, we focus on identifying low-access regions through a methodology able to analyze urban structures considering a set of facilities. Such results refers to the ones from our third publication [39]. The methodology that we proposed is based on the hypothesis that *geometric and geographic measurements derived from street networks can aid in the identification of urban inconsistencies observed on ill-designed urban structures*. To demonstrate our methodology we proposed a set of formalisms, and two algorithms named Redundancy Removal and Inconsistency Identification. Our achievements confirm the hypothesis of our work, allowing us to state that the city of Sao Carlos lacks well-positioned facilities and that low-access regions hamper the mobility *from* or *to* facilities that provide public assistance. The highlighted contributions are: **(i)** a novel concept based on critical problems in the urban design, which are caused by mispositioning facilities in a city; **(ii)** new approaches to preprocess and prepare maps in the form of street networks, as well as a method to analyze urban inconsistencies; also, **(iii)** the analyses of the discussed methods by evaluating the Brazilian city of Sao Carlos.

Our approach has inherent intricacies, what renders it several future works. **(1)** Considering the Redundancy Removal algorithm when removing elements of the network we are altering the geometry and shape of the city that it represents; it is necessary, henceforth, to develop visualization techniques that can handle the structural simplification of the network. **(2)** The proposed method to identify Urban Inconsistencies is binary; further algebra can aid in identifying the severity of the network inconsistency in a continuous, rather than binary, manner. **(3)** Not always an inconsistency can be avoided, this is because there are other variables that can influence the inconsistency occurrence. For this reason, this concept may need to be used in conjunction with others, comparing each mathematical inconsistency together with the real-world reason that makes a certain node inconsistent. **(4)** For simplicity purposes, we considered displacements starting and ending in nodes of a street network, which are streets intersections in the real-world. Intuitively, a facility may be on the side of a road, but not at a node. The generalization might be achieved by inducing new nodes for specific processing.

Conclusion

This dissertation described works to deal with the problem theme outlined in Chapter 1, which encompasses three sub-problems. Each one culminated in a single work with a particular set of methods that contributed with novel methodologies concerning problems from urban structures. Our approaches not only focused on mobility and transportation, but also in behavioral characteristics that can be investigated from the perspective of the topology of the network. Thus, it was our goal to enhance the capacity of analysis over these networks to aid decision-making activities that focus on city transportation and public safety.

In Chapter 2 we introduced steps to map and assess urban areas that suffer from high indices of criminal activities. Our work was conducted via the hypothesis that **by employing network mapping techniques, allied to distance-based properties of graphs, it is possible to trace the relationship between areas that are highly criminal within a city**. Accordingly, we described a systematic methodology that proved able to infer regions that are more criminal, consequently, validating the corresponding hypothesis. We highlight the following contributions: (i) the use of a complex network to represent the real-world space; (ii) the analyses of crimes based on their spatial disposition; and, (iii) the assessment of the impact of criminal regions, considering the similarities traced between distinct crime types.

In Chapter 3 we extended the toolset proposed in Chapter 2, considering that criminal activity might propagate as an epidemic. Our assumption was that the epidemic loses strength on a distance basis. Also, we considered that human interference can narrow the propagation by applying crime prevention policies. We conducted this work considering the hypothesis 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**. Our results validated our hypothesis by showing that our methodology can be used in planning actions for public safety. The highlighted contributions are: (i) the analyses of crimes regarding their georeferenced

coordinates; (ii) the development of a model to describe criminal activity; and, (iii) the illustration of human interference on criminal regions of a city.

Chapter 4 introduces the concept of Urban Inconsistency observed on ill-designed urban structures. Such concept consists of a set of algebraic methods and geometric features extracted from the complex network, which was used to highlight discrepancies found in urban planning activities. Such work was based on the hypothesis that **geometric and geographic measurements derived from street networks can aid in the identification of urban inconsistencies observed on ill-designed urban structures**. We proved the hypothesis through algorithms to pre-process and analyze the street network from the perspective of the placement of its public services. With regard to this work, we highlight: (i) a novel concept based on critical problems in the urban design; (ii) new approaches to pre-process maps in the form of street networks; and, (iii) the analyses of the methods by evaluating the Brazilian city of Sao Carlos.

References

- [1] BETTENCOURT, L. M. The kind of problem a city is. **Santa Fe Institute**, Birkhauser Heidelberg, p. 175–187, 2013. Citation on page 23.
- [2] COSTA, L. F.; OLIVEIRA, O. N.; TRAVIESO, G.; RODRIGUES, F. A.; BOAS, P. R. V.; ANTIQUEIRA, L.; VIANA, M. P.; ROCHA, L. E. C. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. **Advances in Physics**, Informa UK Limited, v. 60, n. 3, p. 329–412, jun 2011. ISSN 0001-8732. Citations on pages 23 e 51.
- [3] CORBURN, J. Bringing Local Knowledge into Environmental Decision Making: Improving urban planning for communities at risk. **Journal of Planning Education and Research**, SAGE Publications, v. 22, n. 4, p. 420–433, jun 2003. Citation on page 23.
- [4] FRUMKIN, H.; FRANK, L.; JACKSON, R. J. **Urban sprawl and public health: Designing, planning, and building for healthy communities**. [S.l.]: Island Press, 2004. Citation on page 23.
- [5] FURTADO, B. A.; SAKOWSKI, P. A. M.; TÓVOLI, M. H. **Modeling complex systems for public policies**. Ipea. Brasília: Instituto de Pesquisa Econômica Aplicada (Ipea), 2015. 436 p. ISBN 9788578112486. Citation on page 24.
- [6] FOLTÊTE, J.-C.; PIOMBINI, A. Urban layout, landscape features and pedestrian usage. **Landscape and Urban Planning**, Elsevier BV, v. 81, n. 3, p. 225–234, jun 2007. Citation on page 24.
- [7] BETTENCOURT, L. M. A. The Origins of Scaling in Cities. **Science**, American Association for the Advancement of Science (AAAS), v. 340, n. 6139, p. 1438–1441, jun 2013. Citation on page 24.
- [8] BETTENCOURT, L. M. A.; LOBO, J.; STRUMSKY, D.; WEST, G. B. Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities. **PLoS ONE**, Public Library of Science (PLOS), v. 5, n. 11, p. e13541, nov 2010. Citation on page 24.
- [9] BETTENCOURT, L. M. A.; LOBO, J.; HELBING, D.; KUHNERT, C.; WEST, G. B. Growth, innovation, scaling, and the pace of life in cities. **Proceedings of the National Academy of Sciences**, Proceedings of the National Academy of Sciences, v. 104, n. 17, p. 7301–7306, apr 2007. Citation on page 24.
- [10] COSTA, L. F.; TRAVENÇOLO, B. A. N.; VIANA, M. P.; STRANO, E. On the efficiency of transportation systems in large cities. **EPL (Europhysics Letters)**, IOP Publishing, v. 91, n. 1, jul 2010. Citations on pages 25 e 53.
- [11] PORTA, S.; LATORA, V.; WANG, F.; STRANO, E.; CARDILLO, A.; SCELLATO, S.; IACOVIELLO, V.; MESSORA, R. Street Centrality and Densities of Retail and Services in Bologna, Italy. **Environment and Planning B: Planning and Design**, SAGE Publications, v. 36, n. 3, p. 450–465, 2009. Citations on pages 25, 29 e 51.

- [12] STRANO, E.; NICOSIA, V.; LATORA, V.; PORTA, S.; BARTHÉLEMY, M. Elementary processes governing the evolution of road networks. **Scientific Reports**, Springer Nature, v. 2, mar 2012. Citation on page 25.
- [13] PORTA, S.; CRUCITTI, P.; LATORA, V. The Network Analysis of Urban Streets: A Primal Approach. **Environment and Planning B: Planning and Design**, SAGE Publications, v. 33, n. 5, p. 705–725, oct 2006. Citations on pages 25, 51 e 52.
- [14] CRUCITTI, P.; LATORA, V.; PORTA, S. Centrality measures in spatial networks of urban streets. **Physical Review E - Statistical, Nonlinear, and Soft Matter Physics**, American Physical Society (APS), v. 73, n. 3, mar 2006. Citation on page 25.
- [15] PORTA, S.; LATORA, V.; WANG, F.; RUEDA, S.; STRANO, E.; SCELLATO, S.; CARDILLO, A.; BELLÍ, E.; CARDENAS, F.; CORMENZANA, B.; LATORA, L. Street Centrality and the Location of Economic Activities in Barcelona. **Urban Studies**, SAGE Publications, v. 49, n. 7, p. 1471–1488, nov 2011. Citation on page 25.
- [16] LOWRY, M. B.; BALLING, R. J. An approach to land-use and transportation planning that facilitates city and region cooperation. **Environment and Planning B: Planning and Design**, SAGE Publications, v. 36, n. 3, p. 487–504, 2009. Citation on page 25.
- [17] HUANG, B.; ZHANG, W. Sustainable Land-Use Planning for a Downtown Lake Area in Central China: Multiobjective Optimization Approach Aided by Urban Growth Modeling. **Journal of Urban Planning and Development**, American Society of Civil Engineers (ASCE), v. 140, n. 2, p. 04014002, jun 2014. Citation on page 25.
- [18] MASUCCI, A. P.; STANILOV, K.; BATTY, M. Limited Urban Growth: London's Street Network Dynamics since the 18th Century. **PLoS ONE**, Public Library of Science (PLoS), v. 8, n. 8, aug 2013. Citation on page 25.
- [19] KACZYNSKI, A. T.; KOOHSARI, M. J.; STANIS, S. A. W.; BERGSTROM, R.; SUGIYAMA, T. Association of Street Connectivity and Road Traffic Speed With Park Usage and Park-Based Physical Activity. **American Journal of Health Promotion**, SAGE Publications, v. 28, n. 3, p. 197–203, jan 2014. Citation on page 25.
- [20] SOPAN, A.; REY, P. J.; SHNEIDERMAN, B. The Dynamics of Web-Based Community Safety Groups: Lessons Learned from the Nation of Neighbors. **IEEE Signal Processing Magazine**, Institute of Electrical and Electronics Engineers (IEEE), v. 30, n. 6, p. 157–162, nov 2013. Citation on page 25.
- [21] EDWARDS, D.; GRIFFIN, T. Understanding tourists' spatial behaviour: GPS tracking as an aid to sustainable destination management. **Journal of Sustainable Tourism**, Informa UK Limited, v. 21, n. 4, p. 580–595, may 2013. Citation on page 25.
- [22] VIANA, M. P.; COSTA, L. F. Fast long-range connections in transportation networks. **Physics Letters, Section A: General, Atomic and Solid State Physics**, Elsevier BV, v. 375, n. 15, p. 1626–1629, apr 2011. Citations on pages 25 e 53.

-
- [23] TRAVENÇOLO, B.; COSTA, L. F. Accessibility in complex networks. **Physics Letters, Section A: General, Atomic and Solid State Physics**, Elsevier BV, v. 373, n. 1, p. 89–95, dec 2008. ISSN 0375-9601. Citations on pages [26](#) e [53](#).
 - [24] MARUHASHI, K.; SHIGEZUMI, J.; YUGAMI, N.; FALOUTSOS, C. EigenSP: A More Accurate Shortest Path Distance Estimation on Large-Scale Networks. In: **2012 IEEE 12th International Conference on Data Mining Workshops**. [S.l.]: Institute of Electrical and Electronics Engineers (IEEE), 2012. Citation on page [26](#).
 - [25] GALBRUN, E.; PELECHRINIS, K.; TERZI, E. Urban navigation beyond shortest route: The case of safe paths. **Information Systems**, Elsevier BV, v. 57, p. 160–171, apr 2016. Citations on pages [26](#) e [32](#).
 - [26] VRAGOVIĆ, I.; LOUIS, E.; DÍAZ-GUILERA, A. Efficiency of informational transfer in regular and complex networks. **Physical Review E - Statistical, Nonlinear, and Soft Matter Physics**, American Physical Society (APS), v. 71, n. 3, mar 2005. Citation on page [26](#).
 - [27] GIL, J. Street network analysis “edge effects”: Examining the sensitivity of centrality measures to boundary conditions. **Environment and Planning B: Planning and Design**, SAGE Publications, may 2016. Citation on page [26](#).
 - [28] CARDILLO, A.; SCELLATO, S.; LATORA, V.; PORTA, S. Structural properties of planar graphs of urban street patterns. **Physical Review E - Statistical, Nonlinear, and Soft Matter Physics**, American Physical Society (APS), v. 73, n. 6, jun 2006. Citation on page [26](#).
 - [29] CORCORAN, P.; JILANI, M.; MOONEY, P.; BERTOLOTTO, M. Inferring semantics from geometry. In: **Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS’15**. [S.l.]: Association for Computing Machinery (ACM), 2015. Citation on page [26](#).
 - [30] BARTHÉLEMY, M.; FLAMMINI, A. Modeling Urban Street Patterns. **Physical Review Letters**, American Physical Society (APS), v. 100, n. 13, apr 2008. Citation on page [26](#).
 - [31] ZHONG, C.; ARISONA, S. M.; HUANG, X.; BATTY, M.; SCHMITT, G. Detecting the dynamics of urban structure through spatial network analysis. **International Journal of Geographical Information Science**, Informa UK Limited, v. 28, n. 11, p. 2178–2199, may 2014. Citation on page [26](#).
 - [32] SCRIPPS, J.; NUSSBAUM, R.; TAN, P.-N.; ESFAHANIAN, A.-H. Link-Based Network Mining. In: **Structural Analysis of Complex Networks**. [S.l.]: Springer Nature, 2010. p. 403–419. Citation on page [26](#).
 - [33] MYRONENKO, S.; WENGER, A.; ATMAZHOV, S. Capacity analysis of the street and road network of modern regional center. **Odes’kyi Politechnichniy Universytet. Pratsi**, Odessa National Polytechnic University, n. 3, p. 21–25, dec 2016. Citation on page [26](#).
 - [34] STRANO, E.; VIANA, M.; COSTA, L. F.; CARDILLO, A.; PORTA, S.; LATORA, V. Urban Street Networks, a Comparative Analysis of Ten European Cities. **Environment**

- and Planning B: Planning and Design**, SAGE Publications, v. 40, n. 6, p. 1071–1086, dec 2013. Citations on pages [26](#) e [53](#).
- [35] LÄMMER, S.; GEHLSSEN, B.; HELBING, D. Scaling laws in the spatial structure of urban road networks. **Physica A: Statistical Mechanics and its Applications**, Elsevier BV, v. 363, n. 1, p. 89–95, apr 2006. ISSN 0378-4371. Citation on page [26](#).
- [36] PORTA, S.; CRUCITTI, P.; LATORA, V. The network analysis of urban streets: A dual approach. **Physica A: Statistical Mechanics and its Applications**, Elsevier BV, v. 369, n. 2, p. 853–866, sep 2006. Citation on page [26](#).
- [37] SPADON, G.; SCABORA, L. C.; OLIVEIRA, P. H.; ARAUJO, M. V. S.; SOUSA, E. P. M.; TRAINA-JR, C.; MACHADO, B. B.; RODRIGUES-JR, J. F. Complex network tools to understand the behavior of criminality in urban areas: the case of San Francisco. In: LATIFI, S. (Ed.). **Information Technology - New Generations: 14th International Conference on Information Technology**. [S.l.]: Springer, 2018. p. 493–500. ISBN 978-3-319-54978-1. Citations on pages [26](#), [30](#), [42](#), [45](#), [46](#) e [48](#).
- [38] SPADON, G.; SCABORA, L. C.; ARAUJO, M. V. S.; OLIVEIRA, P. H.; SOUSA, E. P. M.; TRAINA-JR, C.; MACHADO, B. B.; RODRIGUES-JR, J. F. Behavioral Characterization of Criminality Spread in Cities. In: . [S.l.]: Elsevier BV, 2017. v. 108, p. 2537 – 2541. ISSN 1877-0509. International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland. Citations on pages [26](#), [45](#) e [50](#).
- [39] SPADON, G.; GIMENES, G.; RODRIGUES-JR, J. F. Identifying Urban Inconsistencies via Street Networks. In: . [S.l.]: Elsevier BV, 2017. v. 108, p. 18 – 27. ISSN 1877-0509. International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland. Citations on pages [27](#), [52](#) e [63](#).
- [40] CENTOLA, D. The Spread of Behavior in an Online Social Network Experiment. **Science**, American Association for the Advancement of Science (AAAS), v. 329, n. 5996, p. 1194–1197, sep 2010. ISSN 0036-8075. Citation on page [29](#).
- [41] GONZÁLEZ, M. C.; HIDALGO, C. A.; BARABÁSI, A.-L. Understanding individual human mobility patterns. **Nature**, Springer Nature, v. 453, n. 7196, p. 779–782, jun 2008. Citation on page [29](#).
- [42] ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Error and attack tolerance of complex networks. **Nature**, Springer Nature, v. 406, n. 6794, p. 378–382, jul 2000. Citation on page [29](#).
- [43] CHEN, G.; DONG, Z. Y.; HILL, D. J.; ZHANG, G. H.; HUA, K. Q. Attack structural vulnerability of power grids: A hybrid approach based on complex networks. **Physica A: Statistical Mechanics and its Applications**, Elsevier BV, v. 389, n. 3, p. 595–603, feb 2010. ISSN 0378-4371. Citation on page [29](#).
- [44] BAGLER, G. Analysis of the airport network of India as a complex weighted network. **Physica A: Statistical Mechanics and its Applications**, Elsevier BV, v. 387, n. 12, p. 2972–2980, may 2008. ISSN 0378-4371. Citation on page [29](#).
- [45] BUHL, J.; GAUTRAIS, J.; REEVES, N.; SOLÉ, R. V.; VALVERDE, S.; KUNTZ, P.; THERAULAZ, G. Topological patterns in street networks of self-organized urban settlements. **The European Physical Journal B - Condensed Matter and**

-
- Complex Systems**, Springer Nature, v. 49, n. 4, p. 513–522, feb 2006. ISSN 1434-6036. Citation on page [29](#).
- [46] BULLMORE, E.; SPORNS, O. Complex brain networks: graph theoretical analysis of structural and functional systems. **Nature Reviews Neuroscience**, Springer Nature, v. 10, n. 4, p. 312–312, mar 2009. Citation on page [29](#).
- [47] ANGELOUDIS, P.; FISK, D. Large subway systems as complex networks. **Physica A: Statistical Mechanics and its Applications**, Elsevier BV, v. 367, p. 553–558, jul 2006. ISSN 0378-4371. Citation on page [29](#).
- [48] SEN, P.; DASGUPTA, S.; CHATTERJEE, A.; SREERAM, P. A.; MUKHERJEE, G.; MANNA, S. S. Small-world properties of the Indian railway network. **Physical Review E - Statistical, Nonlinear, and Soft Matter Physics**, American Physical Society (APS), v. 67, n. 3, p. 036106, mar 2003. Citation on page [29](#).
- [49] DERYOL, R.; WILCOX, P.; LOGAN, M.; WOOLDREDGE, J. Crime Places in Context: An Illustration of the Multilevel Nature of Hot Spot Development. **Journal of Quantitative Criminology**, Springer Nature, v. 32, n. 2, p. 305–325, jan 2016. ISSN 0748-4518. Citations on pages [29](#) e [45](#).
- [50] SPICER, V.; SONG, J.; BRANTINGHAM, P.; PARK, A.; ANDRESEN, M. A. Street profile analysis: A new method for mapping crime on major roadways. **Applied Geography**, Elsevier BV, v. 69, p. 65–74, apr 2016. ISSN 0143-6228. Citations on pages [31](#) e [45](#).
- [51] SHIODE, S.; SHIODE, N. Network-based space-time search-window technique for hotspot detection of street-level crime incidents. **International Journal of Geographical Information Science**, Informa UK Limited, v. 27, n. 5, p. 866–882, may 2013. ISSN 1365-8816. Citations on pages [31](#) e [45](#).
- [52] WHITE, S.; YEHLE, T.; SERRANO, H.; OLIVEIRA, M.; MENEZES, R. The Spatial Structure of Crime in Urban Environments. In: **Lecture Notes in Computer Science**. [S.l.]: Springer Nature, 2015. v. 8852, p. 102–111. ISBN 9783319151670. Citations on pages [31](#) e [45](#).
- [53] REY, S. J.; MACK, E. A.; KOSCHINSKY, J. Exploratory Space-Time Analysis of Burglary Patterns. **Journal of Quantitative Criminology**, Springer Nature, v. 28, n. 3, p. 509–531, nov 2011. ISSN 0748-4518. Citations on pages [31](#) e [45](#).
- [54] FITTERER, J.; NELSON, T.; NATHOO, F. Predictive crime mapping. **Police Practice and Research**, Informa UK Limited, v. 16, n. 2, p. 121–135, oct 2014. ISSN 1561-4263. Citation on page [32](#).
- [55] BOGOMOLOV, A.; LEPRI, B.; STAIANO, J.; OLIVER, N.; PIANESI, F.; PENTLAND, A. Once Upon a Crime. In: **Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14**. [S.l.]: Association for Computing Machinery (ACM), 2014. v. 1, p. 427–434. ISBN 9781450328852. ISSN 0198-9715. Citations on pages [32](#) e [42](#).
- [56] HAKLAY, M.; WEBER, P. OpenStreetMap: User-Generated Street Maps. **IEEE Pervasive Computing**, Institute of Electrical and Electronics Engineers (IEEE), v. 7, n. 4, p. 12–18, oct 2008. ISSN 1536-1268. Citation on page [33](#).

- [57] KONSTANTOPOULOS, T. **Introduction to projective geometry**. [S.l.]: Dover Publications, 2012. 1–24 p. ISBN 9780486141701. Citation on page 33.
- [58] BARABÁSI, A.-L. **Network Science**. [S.l.]: CUP, 2016. Citation on page 34.
- [59] LASALLE, D.; KARYPIS, G. Multi-threaded modularity based graph clustering using the multilevel paradigm. **Journal of Parallel and Distributed Computing**, Elsevier BV, v. 76, p. 66–80, feb 2015. ISSN 0743-7315. Citation on page 34.
- [60] ROSENBERG, A.; HIRSCHBERG, J. V-measure: A conditional entropy-based external cluster evaluation measure. In: **Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language (EMNLP-CoNLL’07)**. [S.l.: s.n.], 2007. v. 1, n. June, p. 410–420. Citations on pages 36 e 46.
- [61] BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D. Complex networks: Structure and dynamics. **Physics Reports**, Elsevier BV, v. 424, n. 4-5, p. 175–308, feb 2006. Citations on pages 45 e 51.
- [62] PASTOR-SATORRAS, R.; CASTELLANO, C.; MIEGHEM, P. V.; VESPIGNANI, A. Epidemic processes in complex networks. **Reviews of Modern Physics**, American Physical Society (APS), v. 87, n. 3, p. 925–979, aug 2015. Citation on page 46.
- [63] KERNER, B. Theory of Congested Traffic Flow: Self-Organization Without Bottlenecks. In: **Proceedings of the 14th International Symposium on Transportation and Traffic Theory**. [S.l.: s.n.], 1999. p. 147–171. Citation on page 51.
- [64] SCELLATO, S.; CARDILLO, A.; LATORA, V.; PORTA, S. The backbone of a city. **The European Physical Journal B - Condensed Matter and Complex Systems**, Springer Nature, v. 50, n. 1-2, p. 221–225, feb 2006. ISSN 1434-6028. Citation on page 53.
- [65] CRUCITTI, P.; LATORA, V.; PORTA, S. Centrality in networks of urban streets. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, AIP Publishing, v. 16, n. 1, p. 015113, mar 2006. Citation on page 53.
- [66] ZHANG, T.; ZENG, Z.; JIA, T.; LI, J. Examining the amenability of urban street networks for locating facilities. **Physica A: Statistical Mechanics and its Applications**, Elsevier BV, v. 457, p. 469–479, sep 2016. Citation on page 53.
- [67] ZHANG, W.; CAO, K.; LIU, S.; HUANG, B. A multi-objective optimization approach for health-care facility location-allocation problems in highly developed cities such as Hong Kong. **Computers, Environment and Urban Systems**, Elsevier BV, v. 59, p. 220–230, sep 2016. Citation on page 53.