DEVELOPING A DYNAMIC PREDICTIVE POLICING SYSTEM

TUĞRUL CABİR HAKYEMEZ

BOĞAZİÇİ UNIVERSITY

2022

DEVELOPING A DYNAMIC PREDICTIVE POLICING SYSTEM

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Management Information Systems

by

Tuğrul Cabir Hakyemez

Boğaziçi University

2022

Developing a Dynamic Predictive Policing System

The thesis of Tuğrul Cabir Hakyemez

has been approved by:

Assoc. Bertan Badur _____
(Thesis Advisor)

Prof. Sona Mardikyan _____

Dr. Mustafa Gökçe Baydoğan _____

Assoc. Prof. Adem Akbıyık _____
(External Member)

Dr. Selçuk Kıran _____
(External Member)

January 2022

DECLARATION OF ORIGINALITY

I, Tuğrul Cabir Hakyemez, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature……

Date……

ABSTRACT

Developing a Dynamic Predictive Policing System

The retrospective predictive policing techniques are atheoretical and therefore remain incapable of sensing the changing crime risk across the streets. In this study, we aim to develop a dynamic predictive policing system that capitalizes on theory-based risk indicators. The sample includes all the theft and robbery incidents in Chicago between 2014-2019. In the first step, pipelining bivariate network K analysis and segmented regression, we introduce novel distance-aware risk functions that operationalize spatiotemporal crime risk around the selected urban features (i.e., bus stop, fast food restaurant, gas station, grocery store, pub). In the second step, we develop various network-based predictive policing methods using graph-based deep learning algorithms (i.e., GraphWavenet, Spatiotemporal Graph Convolutional Networks). These methods generate weekly and intraday hotspot predictions. We complement these methods with various theory-based risk indicators including a risk score devised from the novel risk functions, 311 calls, park events, and cooccurring crime incidents. The results showcase that crime risk around urban features varies across space, time, and crime types. Furthermore, this risk is found to be significantly correlated with the regional socioeconomic characteristics. Another important result shows that incorporating theory-based indicators improved the performance of the retrospective methods up to 68%. Amongst the algorithms, GraphWavenet is found to outperform its counterparts in the majority of the prediction models with an accuracy as high as 80%. The proposed system helps law enforcement agents in planning their operations efficiently by pinpointing the micro geographical units with relatively higher risks in the next time step.

ÖZET

Dinamik bir Öngörücü Polislik Sistemi Geliştirme

Geriye dönük öngörücü polislik teknikleri teorik değildir ve bu nedenle sokaklarda değişen suç riskini algılamada yetersiz kalmaktadır. Bu çalışmada, teoriye dayalı risk göstergelerinden yararlanan dinamik bir tahmine dayalı polislik sistemi geliştirmeyi amaçlıyoruz. Örnek, 2014-2019 yılları arasında Chicago'daki tüm hırsızlık ve soygun olaylarını içermektedir. İlk adımda, ağ K analizi ve segmentli regresyon tekniklerini birleştirerek, seçilen kentsel özellikler (örn. otobüs durağı, fast food restoranı, benzin istasyonu, bakkal, pub) etrafında uzamsal-zamansal suç riskini işlevselleştiren yeni, mesafeye duyarlı risk fonksiyonlarını tanıtıyoruz. İkinci adımda, grafik tabanlı derin öğrenme algoritmalarını (yani GraphWavenet, Spatiotemporal Graph Convolutional Networks) kullanarak çeşitli ağ tabanlı tahmine dayalı polislik yöntemleri geliştiriyoruz. Bu yöntemler haftalık ve gün içi etkin nokta tahminleri üretir. Bu yöntemleri, yeni risk işlevlerinden, 311 çağrıdan, park olaylarından ve birlikte meydana gelen suç olaylarından tasarlanmış bir risk puanı dahil olmak üzere çeşitli teori tabanlı risk göstergeleriyle tamamlıyoruz. Sonuçlar, kentsel özellikler etrafındaki suç riskinin mekâna, zamana ve suç türlerine göre değiştiğini göstermektedir. Ayrıca, bu riskin bölgesel sosyoekonomik özelliklerle önemli ölçüde ilişkili olduğu bulunmuştur. Bir diğer önemli sonuç, teoriye dayalı göstergelerin dahil edilmesinin geriye dönük yöntemlerin performansını %68'e kadar iyileştirdiğini göstermektedir. Algoritmalar arasında, GraphWavenet'in tahmin modellerinin çoğunda, %80'e varan doğrulukla benzerlerinden daha iyi performans gösterdiği bulunmuştur. Önerilen sistem, bir sonraki

zaman adımında nispeten daha yüksek risklere sahip mikro coğrafi birimleri belirleyerek kolluk kuvvetlerinin operasyonlarını verimli bir şekilde planlamalarına yardımcı olur.

CURRICULUM VITAE

NAME: Tuğrul Cabir Hakyemez

DEGREES AWARDED

PhD in Management Information Systems, 2022, Boğaziçi University

MA in Management Information Systems, 2015, Sakarya University

BA in Industrial Engineering, 2011, Sakarya University

AREAS OF SPECIAL INTEREST

Spatiotemporal Data Analysis, Urban Hotspot Prediction, Spatial Crime Research, Data Mining, Machine Learning, Deep Learning, Geographical Information Systems,

PROFESSIONAL EXPERIENCE

Graduate Assistant, Chang School of Continuing Education, Ryerson University, 2020 – Present

Research Assistant, Department of Management Information Systems, Boğaziçi University, 2016 – 2017

Research Assistant, Department of Management Information Systems, Sakarya University, 2012 – 2016, 2017-2020

PUBLICATIONS

International Journal Articles

Hakyemez, T. C., & Badur, B. (2021). Crime Risk Stations: Examining Spatiotemporal Influence of Urban Features through Distance-Aware Risk Signal Functions. ISPRS International Journal of Geo-Information, 10(7), 472. (SCI-E).

Hakyemez, T. C. & Mardikyan, S. (2021). The interplay between institutional integration and self-efficacy in the academic performance of first-year university

students: A multigroup approach. The International Journal of Management Education, 19(1), 100430. (SSCI)

Akar, E., Hakyemez, T. C., Bozanta, A., & Akar, S. (2022). What Sells on the Fake News Market? Examining the Impact of Contextualized Rhetorical Features on the Popularity of Fake Tweets. Online Journal of Communication and Media Technologies, 12(1), e202201. https://doi.org/10.30935/ojcmt/11278 (ESCI)

Ozturan, M., Alacam, S., Coskun-Setirek, A., Hakyemez, T. C., Kebapci, H., Ozturk, E., & Tasan, E. (2017). A Citizen-Centric Integrated Information System for Municipalities. Journal of Public Administration and Governance, 175-197.

Akkaya, A., Hakyemez, T. C., Kutbay, E., & Bayraktar, B. K. HOW DOES STUDENTS'LEARNING (ACHIEVEMENT) RELATE TO THEIR LEVEL OF WE-INTENTION AND THEIR LEARNING METHODS ON FACEBOOK? Yönetim Bilişim Sistemleri Dergisi, 2(2), 61-69.

International Conference Proceedings

Hakyemez, T. C., Bozanta, A., & Coşkun, M. (2018). K-Means vs. Fuzzy C-Means: A Comparative Analysis of Two Popular Clustering Techniques on the Featured Mobile Applications Benchmark. Proceedings of 5th International Management Information Systems Conference, (s. 1-21). Ankara.

Hakyemez, T. C., & Coskun, E. (2014). A Data Mining Approach towards Predicting the Academic Success of the Students based on Demographic Features and Highschool Education Parameters: Results of a Preliminary Study at Sakarya University. Global Interdisciplinary Business-Economics Advancement Conference (GIBA), (p. 708-713). 2014.

Hakyemez, T.C., Cekici, T., & Gundogan, T. (2013). Extreme Logistics: Comparison with Traditional Logistics Services and an Overview of the Situation in Turkey. IX. Logistics and Suply Chain Congress. Kayseri.

ACKNOWLEDGEMENTS

After a long journey of six years, the day has finally come: writing this acknowledgement is the finishing touch on my dissertation. It has been a rewarding period for me, not only scientifically, but also personally. Preparing this thesis has had a transforming impact on me. I would like to take a moment to give credits to those who have contributed so much to this study during this period.

I would like to express my deep gratitude to my thesis advisor Assoc. Prof. Bertan BADUR for his unwavering support, wise feedbacks, enthusiastic encouragement. I've always felt him on my side. This thesis would not be possible without his extremely competent supervision.

I would like to send my thanks to the thesis committee –Prof. Sona MARDİKYAN, Dr. Mustafa Gökçe BAYDOĞAN, and Assoc. Prof. Adem AKBIYIK and Dr. Selçuk KIRAN for caring to be a part of the committee, their constructive criticism, and valuable comments.

I am particularly grateful for the support I received from all my friends, colleagues and the precious professors of the department of Management Information Systems of Bogazici University.

I would also like to thank The Bogazici University Scientific Research Projects Fund (Project #15385), for providing me with the funding I need to conduct this research.

I would like to thank my brothers Yavuz Selim HAKYEMEZ and Oğuz Kağan HAKYEMEZ, my beloved nephews Ege Kağan HAKYEMEZ, Işık TAŞÇEVİREN, Ayşegül Aybike HAKYEMEZ, Zeynep Ada HAKYEMEZ, and Oğuz Alp

HAKYEMEZ, and my sister and brother in-laws, Fulya and Mehmet TAŞÇEVİREN, for their unconditional support, encouragement and love.

I am beyond grateful to my mother Fatma Afitalp HAKYEMEZ, and my father Ali HAKYEMEZ for their unconditional love and support throughout my entire life. I would not go this far without their unwavering supports. They are always there for me, which is something that cannot be provided by anyone but the parents. And special thanks to my mother and father in-laws, Nuray and Ersin BOZANTA, for their support especially when I was struggling to wrap up the thesis. The free times they created for me were valuable beyond measure.

Above all, it was my lovely wife, Aysun BOZANTA HAKYEMEZ, who firmly held my hands when I was lost many times searching for the light at the end of the tunnel. This thesis would not be even possible without her loving and calming support. And, last but not least, I would like to send my deepest love to my beloved son, Kerem Ilgaz HAKYEMEZ, from the bottom of my heart. Although he pitched in much later than his mother, his shining presence gave me the strength I very much need when things seemed utterly hopeless.

With love and hope.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 1

CHAPTER 2 LITERATURE REVIEW 13

2.1. Spatial units in predictive policing 13

2.2. Criminological background 16

2.3 The methods used in predictive policing 27

2.4 DL-based crime prediction methods 31

2.5. This study 35

CHAPTER 3 METHODOLOGY 42

3.1. Study setting 42

3.2. Unit of analysis 44

3.3. Creating graphs from the street network 46

3.4. Developing distance-aware risk signal functions (DRSFs) 52

3.5 Developing DL-based predictive crime hotspot prediction models, ... 61

CHAPTER 4 RESULTS 83

4.1. Descriptive analysis 83

4.2. DRSFs 95

4.3. Predictive crime hotspot mapping algorithms.....112

CHAPTER 5 DISCUSSION.....123

5.1. DRSF.....123

5.2. Enhancing DL-based predictive crime hotspot mapping with theory-based event surfaces.....126

CHAPTER 6 135

CONCLUSION.....135

6.1. Research implications 138

6.2. Practical implications.....140

6.3. Limitations 141

6.4. Future research.....143

APPENDIX A. 144

APPENDIX B. 145

APPENDIX C. 146

REFERENCES.....147

LIST OF TABLES

Table 1. The Key Research 41

Table 2.. The Summary of the Resulting Street Graph of Chicago 48

Table 3. The Statistics for the Resulting Subnetworks 50

Table 4. The PCA Results of CD 55

Table 5. Experimental Setup 81

Table 6. The Chicago Crime Dataset 82

Table 7. The Daily and Shift Results of the Autoregressive Robbery Models 112

Table 8. The Daily and Shift Results of the Autoregressive Robbery Models 113

Table 9. The Daily and Shift Results of the Robbery Models with SSRS 114

Table 10. The Daily and Shift Results of the Theft Models with SSRS 114

Table 11. The Daily and Shift Results of the Robbery Models with Feature Counts 115

Table 12. The Daily and Shift Results of the Theft Models with Feature Counts 115

Table 13. The Daily and Shift Results of the Robbery Models with Other Crime (I.E., Theft) 116

Table 14. The Daily and Shift Results of the Theft Models with Other Crime (i.e., Robbery) 117

Table 15. The Results of Daily and Shift Robbery Models with 311 Calls 117

Table 16. The Results of Daily and Shift Theft Models with 311 Calls 118

Table 17. The Results of Daily and Shift Robbery Models with Park Events 119

Table 18. The Results of Daily and Shift Theft Models with Park Events 119

Table 19. The Results of Predictive Crime Hotspot Mapping Algorithms 122

LIST OF FIGURES

Figure 1. The sides of Chicago (Adapted from City of Chicago Data Portal).....44

Figure 2. A snapshot of the city of Chicago on Open Street Map 45

Figure 3. A sample street network with enumerated intersections 46

Figure 4. An example of (a) undirected graph and (b) directed graph (adapted from Fionda & Palopoli, 2011).....47

Figure 5. Snapping a point to a segment (Diener,2015) 49

Figure 6. The side street networks (**a**) CS, (**b**) FSES, (**c**) FNS, (**d**) NWS, (**e**) FSWS, (**f**) SWS, (**g**) NS, (**h**) SS, and (**i**) WS.....51

Figure 7. Crime distribution across the spatiotemporal model 53

Figure 8**.** The distribution of the selected urban features across the sides of Chicago 54

Figure 9. A sample DRSF. 59

Figure 10. Temporal crime distribution a) Robbery, b) Theft 64

Figure 11. Calculating SSRS (a) WD (b) WE (c) FS (d) SS (e) TS 65

Figure 12. Reformulating crime prediction on a street network as a task of GSP (Zhang & Cheng, 2020).....68

Figure 13. The impact of smoothing coefficient on crime count representation: (a) 0.1 (b) 0.5 (c) 0.9 69

Figure 14. Graph Wavenet framework (Source: Shleiffer et al.,2019).....75

Figure 15. STGCN framework (Yu, Yin, & Zhu, 2018).....78

Figure 16. Internal structure of LSTM (Olah, 2015) 80

Figure 17. Total crime incidents in Chicago between 2015 and 2019.....84

Figure 18. Crime distribution across (a) a week (b) day.....84

Figure 19. Total crime counts across the sides of Chicago (a) between the years 2015 and 2019 (b) across week (c) day 86
Figure 20. Yearly trends of the selected crime types 87
Figure 21. Weekly crime distribution (a) robbery (b) theft 87
Figure 22. Shift crime distribution (a) robbery (b) theft 88
Figure 23. Yearly (a) robbery (b) theft counts 89
Figure 24. Weekly (a) robbery (b) theft distribution across the sides 90
Figure 25. Shift (a) robbery (b) theft distribution across the sides of Chicago 91
Figure 26. A sample crime distribution across the street networks 92
Figure 27. Number of street segments across the sides of Chicago 93
Figure 28. % of crime-free segments across the sides of Chicago between 2015 and 2019 94
Figure 29. Yearly segments % having 50% of total incidents 95
Figure 30. Network K result of the default model for robbery and pubs in FSW 96
Figure 31. Network K result of the (a) weekday (b) weekend model for robbery and pubs in FSW 96
Figure 32. Network K result of the (a) FS (b) SS (c) TS model for robbery and pubs in FSW 97
Figure 33. Bus stops RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models. 100
Figure 34. Fast-food restaurant RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models. 102
Figure 35. Gas station RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models. 104

Figure 36. Grocery store RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models.....106

Figure 37. Pub RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models.....108

Figure 38**.** Correlograms between concentrated disadvantage (CD) and the RSIS values of gas stations in (a) default, (b) weekday, (c) weekend, (d) FS, (e) SS, and (f) TS models.....110

Figure 39. Correlograms between concentrated disadvantage (CD) and the RSSS values of gas stations in (**a**) default, (**b**) weekday, (**c**) weekend, (**d**) FS, (**e**) SS, and (**f**) TS models.....111

LIST OF ABBREVIATIONS

ANN: Artificial Neural Networks

CD: Concentrated Disadvantage

CPT: Crime Pattern Theory

CS: Center side

DL: Deep Learning

DRSF: Distance-aware Risk Signal Function

FNS: Far North Side

FS: First Shift

FSES: Far SouthEast Side

FSWS: Far SouthWest Side

IVA: Intensity Value Analysis

LRSS: Local Risk Signal Score

NS: North Side

NWS: NorthWest Side

RAT: Routine Activity Theory

RCT: Rational Choice Theory

RSIS: Risk Signal Intensity Score

RSSS: Risk Signal Strength Score

SS: Second Shift

SS: South Side

SSRS: Street Segment Risk Score

SWS: SouthWest Side

TS: Third Shift

TW: Train Window

SC: Smoothing Coefficient

SBW: Spatial BandWidth

LR: Learning Rate

WS: West Side

CHAPTER 1

INTRODUCTION

It was a decade ago when predictive policing was listed as one of the top 50 inventions of the year by Times Magazine (Grosmann et al.,2011). Since then, this term has become a buzzword amongst public safety researchers and practitioners. Much of this excitement came from the possibility of anticipating where, when, and by/to whom a crime would occur. By definition, predictive policing refers to a set of "analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions" (Perry, 2013). These techniques entail four main targets: crime occurrences, offenders, perpetrators' identities, and victims of crimes. From these types, the last three have thus far received bitter criticisms in many countries worldwide due to the violations they commit against personal data privacy. This is because the algorithms they used largely feed on large amounts of personal data (e.g., Uberti, 2021; Cockrell, 2021). As an example, a controversial case was reported from Israel where the national intelligence agency uses a predictive policing algorithm that criminalizes individuals solely based on their tweets it label as risky (Nashif & Fatafta, 2017). Such cases put a huge question mark on the legitimacy of predictive policing algorithms that often end up in profiling individuals. In this sense, focusing on the prediction of time and place, rather than offender, points to a more ethical research area. This crime-centric approach is the approach we adopted in developing a predictive policing system. From this point on, predictive policing

will thus be used to refer to the techniques that are specialized in predicting when and where, rather than by whom, crime will occur.

Predictive policing techniques enable timely security interventions by pinpointing potential places of future crime. The predictions generated by these techniques serve to the optimal planning of police patrol routes in every time step by allowing the prioritization amongst the places on these routes based on their predicted crime risk. These informed routes thus help increase the presence of law enforcement agents in places with relatively higher crime risks. As a result, this increased presence decreases the overall crime rate by deterring the offenders from offending (Cohen & Felson, 1979). Noticing their large benefits in operational efficiency and public safety, police departments around the world have launched predictive policing projects and thus far achieved significant drops in crime rates. As a striking example, Santa Ana Police Department saw an 11% decrease in theft incidents within a six-month trial period by using commercial software, PREDPOL (Huet, 2015). In a more recent case, Kent Police Department reported a 6% decrease in street violence crime rates in the first four months of using PREDPOL (Smith, 2018). Another commercial software developed by Rutgers Public Security Center (RTMdx) also achieved significant drops in the crime rates reported across the U.S. In two consecutive years, 2013 and 2014, it helped police departments to reduce motor vehicle thefts by 43% in Colorado Springs, gun violence incidents by 35% in New Ark, and aggravated assault by 12% in Kansas City (Kennedy, Caplan, & Piza, 2015). Evident from these facts and figures, predictive policing provides a data-driven innovative technology to law enforcement agents across the world in their combat against crime.

Predicting where and when a crime will occur is also of great value in terms of operational efficiency. This prediction ability predicates upon a risk-based approach that assigns ordinal risk values to places for a specified period. This approach allows the prioritization amongst the places to be patrolled, hence enacting more effective policing practices. Instead of piling the police forces in the historical crime hotspots, it informs patrol dispatching decisions about where and when the next crime event is most likely. Having been armed with this decision support, the security agents achieve highly targeted security interventions with shorter patrol routes and fewer police officers (Perry, 2013; Kennedy, Caplan, & Piza, 2015). Largely due to the substantial amount of money and resource it saved, predictive policing projects have gained significant momentum right after the 2008 crisis in the U.S. when the government significantly curbed public funds that were allocated to public institutions. Furthermore, the government created various incentives to promote the usage of predictive policing systems (Space and Naval Warfare Systems Center Atlantic, 2013). Having noticed the large benefits of predictive policing, many governments around the world followed suit and launches similar projects. For example, the Japanese government has led the development of a national predictive policing system project to enhance public security during the 2021 Tokyo Olympics (South China Morning Post, 2018). Similarly, the U.K. government has embarked upon a similar project with a university-police collaboration (Dunning, 2017). Max Planck Institut in Germany has also been working on the development of a highly sophisticated predictive policing system (Max Planck Institut, 2018). The ever-growing competition in developing predictive policing systems amongst countries showcases how influential these systems will be on the future of policing.

Accurately identifying existing and future crime places across a region is the most crucial task in predictive policing. The key factor in this task is to capture the nonrandom crime distribution in an area. The nonrandomness is largely ensued by the existence of crime hotspots accommodating highly clustered crime. In a seminal work, Sherman, Gartin, & Buerger (1989) found that half of the calls to police were placed from only 3% of the addresses in Minneapolis. Based on that insight was developed hotspot policing that aims to prevent crime by dispatching police forces to these hotspots. Despite a large body of empirical research showcasing its effectiveness in crime prevention (for a systematic review, (Braga, Papachristos, & Hureau, 2012)), hotspot policing often makes a false assumption that the locations of hotspots remain stable in the future. It, therefore, remains incapable of predicting the displacement in crime hotspots. A remarkable fix to that problem was offered by Prospective Hotspot Mapping (Bowers, Johnson, & Pease, 2004) that proposes to place a spatiotemporal crime risk bandwidth around a crime location. This idea of placing a spatiotemporal risk bandwidth is predicated upon a well-established near-repeat phenomenon (Morgan, 2001) that posits that crime in one place elevates risk in neighboring places for a time window. Accordingly, this method fits a risk kernel that decays crime risk based on spatial and temporal distance from a crime location and time. The calculated risk values for each place then form a crime risk surface representing the risk values of each place within a geographical area of interest. The places with the highest risk in this surface are called “hotspots” and techniques that predict this risk surface on a map are called predictive hotspot mapping techniques (Bowers, Johnson, & Pease, 2004). These techniques cater to predictive policing by highlighting the places in an area with higher crime risks in the next time step.

Therefore, effort should be made towards developing predictive crime hotspot maps as an application of predictive policing.

The main debate revolving around predictive crime hotspot mapping methods revolves around the shape of the crime hotspots. A common strategy in these methods is to divide an area into a series of equally sized grids. But this strategy brings about several problems. First, sizes of these grids are often selected arbitrarily, which causes Modifiable Areal Unit Problem (MAUP) that refers to changing outcomes of a spatial analysis based on the size of the selected units (Openshaw, 1981). For instance, a predictive model using the whole city as a unit of analysis almost always achieves 100% accuracy. But this is of little value. On the other hand, the models producing micro-level predictions (e.g., street segment) may have poor predictive accuracy but still are of higher practical value for law enforcement agents due to providing highly specific predictions. For this, there should be some rationale behind the selection of grid size selections, rather than arbitrary practices. Second, human activity moves through street networks in urban contexts rather than through unrealistic grids. Ignoring physical constraints operating on human activity may fail to capture the behavior of crime risk as a product of interactions between motivated offenders and suitable targets (Xu & Griffiths, 2017). Third, police patrol routes are heavily constrained to street networks. Grid-shaped hotspots may sometimes fall in areas beyond the reach of street networks (e.g., deep forests, sea, lake, etc.). These impractical units not only complicate patrol route planning but also distort risk calculations by artificially increasing the number of areas in an area (Rosser, Davies, Bowers, Johnson, & Cheng, 2017). An optimal

crime hotspot prediction method should thus consider these constraints and represent crime risk in a realistic manner.

Another important problem in predictive crime hotspot mapping is the univariate analysis that assumes future hotspots can be predicted solely based on retrospective crime data. Despite having delivered impressive predictive performances in several studies (e.g., Bowers, Johnson, & Pease, 2004; Mohler, Short, Brantingham, Schoenberg, & Tita, 2011; Rosser, Davies, Bowers, Johnson, & Cheng, 2017), this type of analysis is atheoretical (Groff & La Vigne, 2002), and often fail to sense the looming change in crime risk levels (Gorr & Olligschlaeger, 2002). Therefore, crime hotspot prediction methods should be enhanced with theory-based indicators that are able to send early warning signals before a crime occurs (Groff & La Vigne, 2002). Crime opportunity theories offer a solid theoretical framework for deriving the theory-based indicators. These theories are grounded on three main pillars: Routine Activity Theory (RAT) (Cohen & Felson, 1979), Rational Choice Theory (RCT) (Cornish & Clarke, 1987), and Crime Pattern Theory (CPT) (Brantingham & Brantingham, 1995; Brantingham & Brantingham, 1981). At the base level, the RAT formulates crime opportunity as a spatiotemporal convergence of three factors: a suitable target, a motivated offender, and the absence of a capable guardian. The RCT argues that this opportunity does not always lead to a crime event. Rather, it posits that this offending decision involves a rational calculation between risk and reward associated with the existing crime opportunity. Followingly, the CPT spatializes this calculated crime opportunity near two groups of urban features. The first group (i.e., crime generators) attracts crime by drawing larger crowds that probably contain larger numbers of offenders and victims. The second group of

features is known to provide a fertile ground for motivated offenders due to particular characteristics promoting criminal behavior. Supporting evidence for the elevated crime risk around these features was provided by a large volume of empirical research (e.g., Caplan, Kennedy, & Miller, 2011; Groff & Lockwood, 2014; Ratcliffe, 2012). At this point, we need to explicate the conceptual relationship between crime opportunity and crime risk to avoid any confusion amongst the readers in the future sections. To distinguish the opportunity surrounding an urban feature from the opportunity as offenders' subjective interpretation of a crime situation, Caplan et al. (2011) proposed to replace the former with a more objective one, crime risk. In a similar vein, we will use "risk" to refer to the opportunity surrounding an urban feature. In sum, an auxiliary risk surface that captures the environmental crime risk around urban features may improve the predictive ability of crime hotspot mapping methods. Therefore, an environmental crime risk surface should be integrated into these methods.

The degree of crime risk in an environment is quantified by the spatial influence of urban features on crime. A stylized fact in spatial crime research (Vandeviver & Bernasco, 2017) states that this influence is inversely related to distance: it is highest at the origin (i.e., local effect) and decays thereafter until it totally dissipates at some distance (i.e., spatial diffusion effect) (Wheeler, 2019). The total distance stretching to this dissipation point demarcates the spatial extent of an urban feature (Ratcliffe, 2012). Another determinant of spatial influence is human activity that, to a large extent, determines the magnitude of spatial influence and spatial extent. Accordingly, being exposed to greater levels of human activity not only magnifies the spatial influence of urban features but also extends their spatial extent (Groff,

2011). As a result, the environmental crime risk landscape throughout an area dynamically changes based on displacing human activity concentration in space and time. Previous research has reported significant spatial (e.g., Barnum, Caplan, Kennedy, & Piza, 2017) and temporal fluctuations (e.g., Haberman & Ratcliffe, 2015; (Bernasco, Ruiter, & Block, 2017) in the spatial influence of urban features. Additionally, this dynamic spatial influence also changes across crime types due to differential risk and reward definitions for different crime types (Clarke, 1995). For example, although large crowds may represent a fertile ground for theft, they may act as a shield against robbery that often takes the form of a physical struggle between the involved parties. As a result, a proper environmental risk indicator should capture spatiotemporal fluctuations in the spatial influence of urban features across crime types.

Using machine learning is not a new phenomenon in predicting crime hotspots. Over the last three decades, researchers have utilized various techniques including Artificial Neural Networks (ANN) (e.g., Corcoran, Wilson, & Ware, 2003, SVM (e.g., Kang & Kang, 2017), decision trees (e.g., Bogomolov et al., 2014) and random forests (e.g., Cavadas, Branco, & Pereira, 2015). After proving their worth in bringing innovative solutions to a wide range of real-world problems, the last decade has witnessed the rise of Deep Learning (DL) algorithms due to being widely adopted by researchers who would like to leverage their impressive predictive abilities (e.g, Zhuang et al., 2017; Duan et al., 2017; Kang & Kang, 2017). What was mainly found in these studies was the superiority of DL models over their traditional counterparts. However, a common problem in these studies is the grid-like units of analysis they use for crime prediction. These grid-based units are

problematic due to several reasons listed earlier in this chapter. Thus, employing network-based DL algorithms rather than grid-based ones in the task of crime hotspot prediction is highly needed.

This study consists of two independent yet interacting parts. The first part aims to develop a method that quantifies dynamic environmental crime risk around urban features across space, time, and crime types. To do so, we proposed a conceptual framework that views urban features as crime risk stations. These stations act like base stations. Different from the base stations, they broadcast crime risk signals, rather than radio signals, through street networks. The broadcasted signals represent the spatial influence and exist at varying strength levels throughout a coverage area (i.e., spatial extent). The strength is highest at the center (i.e., local effect) and recedes through nearby areas. The broadcasting performance of a crime risk station can be evaluated by the strength of its signal strength throughout a coverage area as well as how much of the signal strength is retained throughout the same area (i.e., spatial diffusion effect). Higher retention rates show higher signal strength. This framework hypothesizes that urban features have a crime-specific spatiotemporal influence manifesting itself in varying signal strength levels across space, time, and crime types. Based on that conceptual framework, we defined our novel Distance-aware Risk Signal Functions (DRSF). These functions operationalize the spatial influence of an urban feature on crime by calculating different types of crime density across various spatiotemporal models. From these functions, we devised two novel risk scores, namely Risk Signal Intensity Score (RSIS) and Risk Signal Strength Score (RSSS) to make within and between comparisons of spatiotemporal influences of urban features on different crime types. The former indicates the cumulative crime

density at a given spatial extent, the latter quantifies how much of the local effect is retained throughout the same spatial extent. Then, a spatiotemporal influence analysis tested the differences in these scores across various spatiotemporal models and crime types. Lastly, we examined the relationship between RSIS, RSSS, and socioeconomic characteristics through a correlation analysis.

In the second part, we first proposed a novel theory-based risk surface that consists of environmental crime risks of each street segment. This surface enhances DL-based prospective hotspot mapping methods by capturing dynamic environmental risk with a novel indicator Street Segment Risk Score (SSRS). Here, SSRS is a key component that bridges the first and second parts of the study. Based on DRSFs from the first part, SSRS captures the dynamic spatial influence of urban features on crime by aggregating their local effects at the street segment level. These local effects are operationalized by another novel risk indicator, the Local Risk Signal Score (LRSS). which is directly obtained from DRSFs. This indicator takes on the value of DRSF at the exact location of an urban feature. To capture the dynamic spatial influence, we calculated separate LRSS values for each selected crime type across temporal models. Next, we multiply these LRSS values of urban features (i.e., multiplicative effect) on a street segment to obtain the SSRS of that segment in a temporal model. For example, on weekdays, theft SSRS of a street segment that contains one grocery store and two pubs having LRSS values of 3 and 4 respectively would be 3x4x4=48. Similarly, on weekends, this value would increase to 196 as a result of the same values increasing to 4 and 7 (4x7x7) respectively. We then use these values to create an environmental risk surface to be incorporated into the DL-based predictive crime hotspot mapping methods. For the hyperparameter tuning, we described a two-step

procedure that gradually optimizes data representation and model performance. After finding the optimal crime data representation through a grid search approach, we experimented with various learning rates to optimize the model performance. In the last step, to highlight the impact of SSRS on predictive accuracy, we compared the performance of the model with SSRS with the autoregressive model and a static risk surface that only counts the number of selected urban features in the street segment. In addition to SSRS, we also created additional theory-driven event surfaces that can be incorporated into the predictive crime hotspot mapping algorithms. Included in these surfaces are the other crime, 311 calls, and park events. Each of these surfaces aims to capture the crime risk from a different aspect. Lastly, we performed a comparative performance analysis amongst the models complemented by theory-driven event surfaces to identify the surface that best senses the looming crime risk in an area. The contribution of this study can be summarized as follows:

- We introduced the Distance-Aware Risk Signal Function (DRSF) to model the dynamic spatial influence of an urban feature on crime within a given spatial extent.
- From these functions, we devised two novel spatial influence scores (i.e., RSIS and RSSS) that allow within and between urban features comparisons in terms of spatial influences across space, time, and crime types.
- This study fills an empirical gap in the methods using graph-based DL algorithms to predict crime hotspots by introducing two graph learning algorithms that were not adapted to the crime hotspot forecasting domain before.
- To the best of our knowledge, this is the first attempt to make shiftly crime hotspot predictions using network-based DL algorithms.

- We enhance the graph-based predictive crime hotspot algorithms with many theory-driven event surfaces including the novel SSRS, 311 requests, park events, and the other crime type.
- We propose a novel two-step hyperparameter tuning procedure that optimizes data representation and model performance gradually.
- It uses a novel squared quantile loss function to overcome the imbalance in the sparse crime dataset.

CHAPTER 2

LITERATURE REVIEW

This chapter is an extensive literature review that aims to provide answers to the research problems described in the previous chapter. First, we focus on the existing spatial units in the literature in that determining a suitable unit is of crucial importance for the validity of a crime hotspot prediction method. Second, we lay the theoretical groundwork for environmental crime risk across places. Third, we provide an overview of the existing crime hotspot prediction methods in the literature. Fourth, we summarize the DL-based crime prediction methods with a special emphasis on graph learning algorithms. Lastly, we position our study in the extant literature to highlight its contribution to the knowledge.

2.1. Spatial units in predictive policing

Selecting the optimal spatial unit is of paramount importance in predictive crime hotspot mapping algorithms since it draws the geographical boundaries of hotspots. In spatial crime research, spatial units are often categorized under three main groups: macro-level (e.g., cities, counties), meso-level (e.g., police beats, community areas, etc.), micro-level (e.g., parcels, households, etc.) (Johnson, Bowers, Birks, & Pease, 2009). In the first group, the primary concern is to provide the top managers with useful insights into the long-term planning of law enforcement activities across an area. The second group helps with the effective allocation of resources amongst the regional law enforcement units. The third group provides a micro understanding of why offenders prefer particular environmental settings over others. In an attempt to identify how these groups explain the crime variability in a city, Steenbeek & Weisburd (2016) addressed micro-units as having explained the most variability in crime incidents that occurred between 2001 and

2009 in Hague. This is a finding that can guide the unit selection in predictive crime hotspot mapping algorithms. Given the necessity of making predictions as specific as possible to increase the effectiveness of the security interventions, micro units appear to be the most suitable group to be used in crime prediction techniques.

Past researchers mainly employ grids as the units of analysis in their crime predictions (e.g., Cheng & Adepeju, 2014; Bowers, Johnson, & Pease, 2004; Steenbeek & Kreis, 2015; Caplan, Kennedy, & Miller, 2011; Rummens, Hardyns, & Pauwels, 2017; Malleson, Steenbeek, & Andresen, 2019). An open question in these units concerns the size at which crime should be aggregated (i.e., scale problem) (Weisburd, Bruinsma, & Bernasco, 2009). The answer to that question bears great importance for crime prediction models to avoid the MAUP (Openshaw, 1981) that refers to unstable model outcomes based on the selected unit size. If the size is too large, the crime risk will be distributed homogeneously across the area of interest, which may result in missing important localized crime risks. Furthermore, crime predictions in that area would be of very little practical value for predictive policing due to the lack of specificity. Likewise, if it is too small, data becomes so scarce that the prediction models cannot generate any unit-specific crime predictions (Oberwittler & Wikström, 2009). Additionally, the variability amongst the unit sizes inflicts the validity and comparability of the predictive models. Hunt (2016) reported significantly changing Predictive Accuracy Indexes (PAI), a measure frequently employed in evaluating the performance of crime prediction models, based on different unit sizes.

Existing research has thus far come up with various methods to overcome the unit selection problem. One line of research has circumvented the problem by proposing a data-driven approach that uses spatial crime clusters as spatial units instead of

predefined units. In one of the earliest studies, Ratcliffe (2005) proposed a method involving Random Nearest Neighbor and Monte Carlo Simulation. This method creates spatial crime clusters in a study area and monitors them for each and every time period. Similarly, Steenbeek & Kreis (2015) applied space-bounded Hierarchical Agglomerative Clustering Analysis to identify the boundaries of homogenous disorder subregions to be used as spatial units. In the same vein, (Johnson, Taylor, & Groff, 2015) analyzed the violent clusters through LISA (Anselin, 1995) and used the resulting clusters as the unit of analysis. The problem with adopting these units is their highly irregular shapes heavily dependent on the data. A potential problem arising from this irregularity would be to organize police patrol routes given the units that cover multiple police jurisdictions. Another problem would be the changing unit sizes based on crime density in a region. Dynamically changing cluster sizes may hamper meaningful comparisons between predictive models used in different periods.

Street segments (to two-sided street sections between intersections) have recently become a popular micro-unit amongst spatial crime researchers. The reason behind this growing popularity is that they are found small enough to capture spatial variations and large enough to avoid geocoding errors (Weisburd, Groff, & Yang, 2012). Furthermore, they are fairly realistic in terms of representing the true daily interactions between individuals in microsocial systems. Therefore, instead of "arbitrarily-shaped" grids (Rosser, Davies, Bowers, Johnson, & Cheng, 2017), researchers widely steered towards the street segments as the unit of analysis in their studies (e.g., Groff, 2011; Groff & Lockwood, 2014; Schnell, Grossman, & Braga, 2019). Using street segments as spatial units has a great advantage: it tracks human activity more accurately by utilizing network distance, rather than unrealistic euclidean distance. In addition, it accurately

reflects the constraints operating on the human movement. On the other hand, a potential challenge in selecting street segments as units of analysis may be data scarcity. As Weisburd (2015) describes in his well-known Law of Crime Concentration at Places, "for a defined measure of crime at a specific microgeographic unit, the concentration of crime will fall within a narrow bandwidth of percentages for a defined cumulative proportion of crime." (pp. 138). This law implies that most of the units will be "crime free". Testing how this situation affects the calculation of crime concentration, Levin, Rosenfeld, & Deckard (2017) noted a significant effect of including "crime free" places into the model on the calculated concentration. In sum, despite the potential data scarcity, street segments still appear to be the most suitable spatial unit in crime prediction amongst others.

2.2. Criminological background

The criminological background of the current study is built upon three main theory groups each explaining a different aspect of crime. The first group, repeat-near repeat victimization, discusses how crime incidents are interdependent in space and time. The second group, opportunity theories, explains how crime opportunity shapes spatiotemporal crime patterns across an area. In this group, a subsection will be devoted to explaining how urban features influence crime opportunity (risk) in near areas. The last group focuses on the role of social context in shaping crime opportunities in an area.

2.2.1. Repeat -Near repeat victimization

Repeat victimization theory hypothesizes a small number of individuals or households are exposed to a disproportionate amount of crime. Moreover, it posits that having been victimized in the past increases the chance of being victimized in the future (Pease, 1998). Past research has well documented the elevated victimization risks for one-time

victims compared to non victimized others (e.g., Polvi, Looman, Humphries, & Pease, 1990, Polvi, Looman, Humphries, & Pease, 1991; Kleemans, 2001; Short, D'orsogna, Brantingham, & Tita, 2009 ; Lantz & Ruback, 2017). What makes one-time victims far more vulnerable to further victimization has two different explanations: event dependence (Tseloni & Pease, 1996) and opportunity. The former views repeat victimization simply as a result of past victimization without giving much of a thought on the potential causes behind the victimization recurrence. The latter explains this repeat victimization as a consequence of the associated crime opportunity that a target represents for offenders. A number of qualitative studies that delve into the modus operandi of the offenders addressed a strong perceived opportunity in their risk-reward calculation before making offense decisions (e.g., Rengert & Wasilchick, 1985; Piquero & Rengert, 1999). Supporting evidence for this opportunity-based explanation was provided by a large volume of research that highlighted the target similarity, rather than identity, as the main factor in target selection (i.e., Bernasco & Nieuwbeerta, 2005; Bernasco, 2010). In sum, these findings support an opportunity-based explanation of repeat victimization patterns. This means that victimization risk may not necessarily be directed at the same individual or household but can be directed at similar others. Based on that opportunity-based explanation, Morgan (2001) defined near repeat victimization that refers to an elevated crime risk around a crime location for a time window. Followingly, using an infectious disease metaphor, Townsley, Homel & Chaseling (2003) tested this hypothesis across burglary incidents in Brisbane. What they found is more pronounced near-repeat patterns in neighborhoods with higher homogeneity in terms of building infrastructure. This is a finding that supports the view that the target similarity is what dominates offenders' decisions. Although burglary is

the most widely tested crime type, many other crime types were also found to exhibit these near-repeat patterns to varying degrees. Included in these types are motor vehicle thefts (e.g., Lockwood, 2012), robbery, and shooting (e.g., Youstin, Nobles, Ward, & Cook, 2011). The differential near repeat patterns across crime types have important implications for predictive policing methods that feed on spatiotemporal crime regularities. As an example, an elevated risk around a burglary location may continue to exist for two weeks within a 400 meters bandwidth in an area whereas the same elevated risk may be to exist only for one week within a 1000 meters bandwidth for motor vehicle thefts. This difference underlines the necessity of placing crime-specific spatiotemporal crime risk bandwidths around crime locations. In short,crime-specific near repeat victimization can capture spatiotemporal regularities better than a merely repeat victimization model does.

2.2.2. Opportunity theories

Opportunity theories explain crime incidents by focusing on the interactions of offenders with the environments in which they operate. The existing literature notes three main theories that form the theoretical basis of the opportunity theories. At the base lies the Routine Activity Theory (RAT) (Cohen & Felson, 1979) that formulates crime opportunity as a convergence of a suitable target with a motivated offender in space and time without a capable guardian present. Although this triangular formulation implies spatiotemporal crime patterns, it does not explicate when and where crime is likely to occur. Noting the possibility of situations where this formulation does not hold, the RCT (Cornish & Clarke;1987) addresses a rational risk-reward calculation made by offenders in their offending decisions. Accordingly, offenders will only offend when they reckon the associated reward would exceed the associated risk. Despite offering a plausible

explanation as to why not all crime opportunities result in actual crime incidents, the RCT still fails to explain where and when a crime is more likely to occur. On top of these theories, CPT (Brantingham & Brantingham, 1981; Brantingham & Brantingham, 1995) positions the crime opportunity onto an environmental backcloth that refers to a collection of external factors of the social environment, physical environment, people, behavior, activities, and timing (Groff, 2017). Then it continues to link the offending decision to the way an offender interprets the opportunity in this environmental backcloth. These interpretations are a product of a mental template that an offender builds on the familiarity with an environment he/she develops over time. The more an offender spends time in this environment traversing from one place to another, the more aware he/she becomes aware of the residing crime opportunities. In other words, they create an awareness space in that environment (Brantingham & Brantingham, 1981, Brantingham & Brantingham, 1995) A key factor that reinforces the indicated mental templates is past successful offending experiences of offenders and/or their peers. Deciphering these awareness spaces has always been of great scholarly interest. (e.g., Bernasco & Nieuwbeerta, 2005; Bernasco, 2010; Bernasco, Ruiter, & Block, 2017). This is because it promises large benefits in terms of characterizing the spatial preferences of offenders. Having detailed knowledge on the residing opportunities and risks as well as the escape routes, the offenders pose the greatest risk for the suitable targets in their awareness spaces (Curtis-Ham, Bernasco, Medvedev, & Polaschek, 2020). The CPT flags these encounters on pathways (i.e., streets) between activity nodes (i.e., the places people frequently visit to perform their daily routines). It further identifies two groups of activity nodes around which the chance of an encounter is much greater compared to others. The first group (i.e., crime generators) increases the crime opportunity by

attracting larger crowds that often contain a larger number of offenders and targets. The second group (i.e., crime attractors) does so due to having particular features promoting criminal behaviors. Past research has confirmed the elevated crime risk surrounding what can be considered either a crime generator or a crime attractor such as pub (e.g., Roncek & Maier, 1991; Roncek & Pravatiner, 1989; Ratcliffe, 2012; Xu & Griffiths, 2017; Groff & Lockwood, 2014), fast food restaurant (e.g., Bernasco & Block, 2011; Haberman & Ratcliffe, 2015), and grocery store (e.g., Barnum, Caplan, Kennedy, & Piza, 2017; Demeau & Parent, 2018). How these features attract crime to their vicinities is to be discussed in the following section that will detail the underlying mechanisms behind the spatial influence of urban features on crime.

2.2.3 The characteristics of spatial influence

The spatial influence of urban features on crime can be measured by the size and density of the surrounding crime clusters around them. Accordingly, crime clusters spreading across larger areas with greater crime densities indicate stronger spatial influences. Spatial extent refers to the size of indicated crime clusters and it demarcates the area where crime density can be formulated as a function of the distance from an urban feature. Human activity is a key factor in spatial influence: the exposure to higher levels of human activity for lengthened periods leads to higher exposure to potential offenders, which not only expands spatial extent but also increases its strength (Groff,2011). A large body of research has revealed that spatial influence decays with distance and entirely disappears at some point (e.g., Ratcliffe,2012; Groff, 2013; Groff & Lockwood,2014) as a consequence of the decreasing spatial interactions between urban features and crime (Rengert, Piquero, & Jones, 1999).

Past researchers have proposed different bandwidth sizes for spatial extents including census blocks (Roncek & Maier, 1991), a quarter-mile (Weisburd et al.,2012), or a completely arbitrary distance (Newton & Hirschfield, 2009). The problem with these disk-like bandwidths is that they often miss the fluctuations in crime density due to assuming a static spatial influence throughout a spatial extent. To characterize spatial influence more accurately, the subsequent researchers have brought different solutions to this problem by trying multiple bandwidths at finer resolution levels. They first placed ring-like spatial buffers of various sizes from 1.7 m to 457.2 m (Ratcliffe,2012; Groff,2011; McCord & Ratcliffe,2007; Xu & Griffiths,2017) around urban features. Next, they calculate the cumulative crime density at each buffer. Lastly, they selected the buffer that marks the end of the elevated crime density as the spatial extent of the urban feature. Comparing an average street block length (i.e., 122 m) and one-quarter mile as spatial buffers, Groff (2011) identified the threshold for the spatial influence of drinking places at 366 m for the first bandwidth and 402 m for the second. She also underlined the role of shorter bandwidths in characterizing spatial influence in that they better capture dynamically changing crime risk throughout an extent. Based on that insight, Xu and Griffith (2017) used bandwidths as small as 1.7 m and found diminishing crime density levels within a spatial extent of 304.8 m.

There are two main types of distance measurement in spatial influence analysis: euclidean distance (Felson & Boivin,2015; Bernasco & Block, 2010) or street network distance (Groff & Lockwood,2014; Xu & Griffiths,2017). Given the fact that human activity is heavily constrained to street networks in urban settings, the latter type is more suitable for representing distances between urban features and crime locations. Moreover, using Euclidean distance causes false-positive spatial relationships (Yamada

& Thill, 2004; Lu & Chen, 2007) or over-smoothed crime clusters (Tompson, Partridge, & Shepherd, 2009). In an attempt to examine the statistical significance between euclidean and network distance in an urban context, Maki & Okabe (2005) detected significant differences in distance calculations under 400 m. Also, a similar hypothesis was tested by Groff (2011) who examined the magnitude of association between bars and crime locations. She determined that the association by street distance was nearly three times higher than the ones identified by Euclidean distance. The studies using network distance have some limitations. First and foremost, they often aggregate crimes with relevant street segments (Groff &Lockwood,2014). This may be a huge problem depending on the length of a street segment. The distance difference between an urban feature at the center and a crime event at the edge would equal the midlength of the street segment given the crime is represented by the segment centroid. To improve spatial precision, Xu and Griffiths (2017) proposed a continuous network space where urban features and crime locations are represented by their exact locations. This approach significantly improved the accuracy of the spatial influence analysis by providing more precise distance measures.

2.2.4. Spatial influence in environmental backcloth

Spatial influence resides in an environmental backcloth (Brantingham & Brantingham, 1981;1993) that refers to a multitude of factors involving the social environment, physical environment, people, behavior, activities, and timing (Groff,2017). This backcloth points to the interplay between the physical environment (i.e., urban features and street networks) and human activity as the key driver of crime opportunities in an urban context. The physical environment routes human activity to different places throughout the day or the week, and thus dynamically changes the crime opportunity

landscape of a city. This fluctuating activity level validates a time geography perspective: the spatiotemporal rhythm of urban life (Pred, 1981) constrains individuals in many ways (Ratcliffe,2006) and regulates their movements in space and time. Accordingly, individuals have to be present at places (e.g., government offices, clothing stores) within their operating times to complete their tasks. These constraints lead individuals, regardless of being an offender or a target, to track common routes between activity nodes (e.g., home, work, shop). The dynamic change in the number of people on these common routes throughout a day or week is the underlying cause behind a periodical crime opportunity pattern around urban features (e.g., Haberman&Ratcliffe,2015; Bernasco, Ruiter, & Block, 2017; Corcoran, Zahnow, Kimpton, Wickes, & Brunsdon, 2021). The existing literature has reported a set of urban features that accommodate such periodical opportunities, such as schools, bus stops, and fast-food restaurants (e.g., Haberman&Ratcliffe,2015; Irvin-Erickson & La Vigne, 2015; Hart & Miethe, 2015; MacDonald, Nicosia, & Ukert, 2018).

Previous research has highlighted the localized spatial influence of urban features on crime (Barnum, Caplan, Kennedy, & Piza, 2017). This is a fact that manifests itself in a citywide j-shaped crime distribution (Eck, Clarke, & Guerette, 2007) that accrues from a highly asymmetrical crime exposure amongst urban features while controlling their types. To address this heterogeneity, Kinney, Brantingham, Wuschke, Kirk, & Brantingham (2008) utilizes the concept of" urban mosaics" (Timms, 1975) that characterizes cities as a set of mosaics having unique urban layouts, activities, and sociodemographic structures. Similarly, Hipp & Kim (2019) found that robbery risk is significantly smaller in the commercial district of Southern California in daytime than at night due to the presence of employees. Another important finding is the elevated

robbery risk in segments with a larger restaurant density on weekends. In a similar study, Stucky & Ottensmann (2009) revealed an increased spatial influence of retail stores on robbery risk in disadvantaged neighborhoods. These findings suggest the fact that that crime risk demonstrates a clear spatiotemporal pattern across a city. Therefore, an accurate crime risk operationalization should adopt a spatiotemporal approach.

The peculiar situational characteristics condition the periodical crime opportunity patterns across crime types. The primary situational crime prevention principle recommends differential security interventions to different crime types because of the obvious differences existing in associated crime opportunities (Clarke,1995). A straightforward example would be the crime opportunity represented by a vacant house with an open window. Here, the opportunity level would be completely irrelevant to robbers whereas it sets the ideal scene for a burglar. This conditioned opportunity implies differential spatial influences for urban features across crime types. In an empirical study, Groff & Lockwood (2014) reported the spatial influence of bars on disorder crimes to have nearly 1.5 times of what it has on violent and property crime within a short bandwidth. Within a 100 m bandwidth, Breetzke & Edelstein (2020) similarly showed that schools are exposed to two times more assault incidents than it is exposed to robbery incidents within the same bandwidth. In short, spatial influence operates differently on different crime types. Therefore, spatial influence should be quantified by crimes-specific indicators.

2.2.5. Ecological theories

The ecological theories explain crime variability across city regions based on the observed social and physical disorganization. Focusing on the former, Social Disorganization Theory (Shaw & McKay, 1942) views crime as an inevitable result of

dissolved social bonds amongst the members of a community. By undermining the trust and solidarity amongst the members, social disorganization weakens informal social control that acts as an important natural surveillance mechanism against criminal behavior. The deepened fear of victimization causes the members of this community to avoid any kind of social interaction with others (Kitchen & Williams, 2010). In her seminal work, Jacobs (1961) cited the social interactions between community members in shared movement spaces as the backbone of maintaining the public order. Therefore, the damaged social ties in a neighborhood impair the social order by deactivating the informal social control mechanism.

Social disorganization is often measured by sociodemographic indicators such as unemployment rate, education rate, and inverted median house income (e.g., Jones & Pridemore, 2019). The problem with using these indicators in crime prediction algorithms is their likelihood to calculate higher risks for disadvantaged areas. These inflated risk values may trigger biased security interventions resulting in disproportionate patrol concentration. This biased risk calculation is the most controversial issue that receives bitter criticisms from scholars questioning the ethics of predictive policing (e.g., Bennett Moses & Chan, 2018, Browning & Arrigo, 2021). Less controversial are the indicators derived from the Broken Windows Theory (Wilson & Kelling, 1982) that views broken windows and other unresolved social and physical disorders as serious signs of future crime incidents. The authors consider the indifference from the residents of a neighborhood to physical disorganization as a sign of a common disinterest of its residents in keeping their neighborhood decent. In conclusion, this indicated disinterest is hypothesized to represent social disorganization, hence causing higher crime risks. The broken windows theory is often measured by set

visible cues of disorganization such as 311 calls (e.g., Zhao & Tang, 2020) or misdemeanors (e.g., Cerdá et al., 2009).

A more recent theoretical endeavor is to reconcile ecological and opportunity theories. The CPT points to the effects of social environment on crime opportunity in the environmental backcloth (Brantingham & Brantingham, 1993; Brantingham & Brantingham, 1995). However, it was Wilcox, Land, & Hunt (2003) who first elaborated on the idea of crime contextualization by supplementing the traditional routine activity triangle with a multi contextual dynamic perspective. In their Multicontextual/Multilevel Criminal Opportunity Theory (MCOT), they posit that neighborhood opportunity context renders convergence of victims and offenders more likely in space and time. This opportunity context is heavily influenced by aggregate exposure to offenders, the concentration of suitable targets, and levels of collective guardianship. Similarly, Place in Neighborhood (PIN) framework underlines this interplay between place and neighborhood in crime (Wilcox & Tillyer, 2017). The main premises of this framework are (1) offenders make their decisions based on perceived risk, effort, and reward (2) this rational calculation is influenced by the general context of the surrounding neighborhood. These theories accentuated the hierarchical perspective towards crime risk by identifying possible neighborhood effects.

Briefly, crime demonstrates spatiotemporal regularities that enable the prediction of future incidents. At the base level, repeat near repeat theories provides a theoretical rationale in using past crime events to predict future events. Further explanations regarding what drives near-repeat patterns come from opportunity theories that ground crime in an urban context. Lastly, ecological theories bring a reasonable explanation to

cross-regional crime opportunity differences through the incorporation of physical and social disorganization.

2.3 The methods used in predictive policing

Fed by various data sources, predictive policing models are able to identify crime hotspots across an area before they emerge. The primary sources are crime datasets that describe a crime event with a set of features such as time, location, crime type. Multivariate prediction models incorporate external datasets about environmental configurations (e.g., land use) and sociodemographic characteristics (e.g., unemployment rate, education level). After data preprocessing, the researchers aggregate these datasets at various spatial levels such as equal-sized grids (Garnier, Caplan, & Kennedy, 2018), neighborhoods (e.g., Gerber, 2014), or street segments (e.g., Rosser, Davies, Bowers, Johnson, & Cheng, 2017). The goal is to predict the crime outcomes in each unit in the next time step. In this study, we follow a categorization that groups the models based on these outcomes. This categorization includes: Estimation of Crime Intensity based on Space-Time Interaction, Surveillance of Space-Time Clusters of Crime, Prediction of Crime Based on Environmental Factors, Prediction of Crime Counts and Possibilities (Ohyama & Amemiya, 2018).

2.3.1 Estimation of crime intensity based on space-time interaction

The beginning of empirical crime hotspot research was often marked by the seminal work of Sherman, Gartin, & Buerger (1989). In this study, Sherman and his colleagues found that nearly 50% of the total 323979 calls to the Minneapolis police department during 1985 came from only 3% of all addresses in the city. Although it opens up a whole new research avenue, the study is mostly criticized by following researchers for its basic assumptions on the stability and independence of hotspots (Kennedy, Caplan,

& Piza, 2015). As a matter of fact, empirical research has repeatedly shown that these hotspots demonstrate displacement and diffusion over time (e.g., Bowers & Johnson, 2003; Weisburd et al., 2006; Short, Brantingham, Bertozzi, & Tita, 2010). Based on the near-repeat phenomenon, Bowers, Johnson, & Pease (2004) proposed a fix to that problem with a prospective hotspot approach that postulates past crime locations can predict the future ones. In this approach, researchers use Kernel Density Estimation (KDE) method to predict crime intensity across units in the next time step. The crime intensity in a unit is calculated based on retrospective crime data. Then, these intensity values at units are placed onto a "risk surface". The units ranked at the top in terms of the calculated risk value are predicted to be crime hotspots in the next time step. The most popular methods in this group are ProMap (Johnson, Bowers, Birks, & Pease, 2009) and Self Exciting Point Process (SEPP) (Mohler, Short, Brantingham, Schoenberg, & Tita, 2011). Both methods calculate a crime risk surface based on retrospective crime data. They only differ in their risk calculations. The former utilizes spatiotemporal kernels that decay crime risk around crime locations in space and time. Based on the Epidemic Type Aftershock Sequence (ETAS) which is a widely popular method in earthquake prediction (Ogata, 1988), the latter calculates risk based on a function that takes the predecessor-successor relationship between crime occurrences into consideration. The main problem with these methods is that they only rely upon retrospective data in their risk calculations and do not incorporate any contextual factors. A group of subsequent research attempted to improve these methods through the inclusion of external variables such as tweets (e.g., Gerber, 2014), weather (e.g., Wang, 2015), or coexisting crime incidents (e.g., Mohler, 2014). Another problem in these methods is the grid-like spatial units whose sizes are often selected arbitrarily. The only

existing solution to that problem was proposed by Rosser, Davies, Bowers, Johnson, & Cheng (2017) who developed a method called Network KDE (NTKDE). Instead of using arbitrarily shaped grids, they propose street segments as units of analysis. In the same study, they reported a 20% improvement from using network-based KDE over the grid-based KDE. In the extant literature, there does not exist any study that brought a complete solution that both problems.

2.3.2 Surveillance of space–time clusters of crime

Past research widely used clustering techniques (e.g., k-means, nearest-neighbor hierarchical clustering) to identify the existing crime hotspots across an area (e.g., Andresen, Curman, & Linning, 2017; Mazeika & Kumar, 2017). Although these techniques are quite suitable for descriptive inferences; they are not well-suited for the crime prediction tasks. Being able to predict crime by dynamically tracking spatiotemporal crime clusters, Space-Time Scan Statistics (STSS) ((Kulldorff, Athas, Feurer, Miller, & Key, 1998)) and Prospective Scan Statistics (PSTSS) (Kulldorff, Heffernan, Hartman, Assunçao, & Mostashari, 2005) demonstrate two exceptions. The main idea behind these models is to scan the overlapping time windows across a geographic area. They identify significant emerging crime clusters when overlapping windows exceed the threshold at a specified intensity level (Ohyama and Amemiya, 2018). The clusters with unusual intensity can be interpreted as potential crime hotspots. These techniques allow the monitoring of crime risk in real-time. Yet, the shape of emerging clusters is not stable and change dynamically. As a result, these unstructured spatial units pose a great challenge to planning patrol routes.

2.3.3 Prediction of crime based on environmental factors

The most popular method in this category is Risk Terrain Modeling (RTM) (Caplan, Kennedy, & Miller, 2011). This method basically strives to incorporate the environmental backcloth (Brantingham & Brantingham, 1981) into crime prediction with the help of a multilayered map component. To do so, it operationalizes and standardizes each physical risk factor of environmental backcloth at a common terrain. Its main working principle is to calculate a Relative Risk Value (RRV) for each spatial unit by overlapping risk layers stacked on a map. Then it uses this RRV as a predictor of a logistic regression equation that predicts the probability of crime in that unit. As a result, the areas having higher probabilities are predicted to be crime hotspots. Empirical evidence has highlighted the superiority of RTM to other methods in terms of predictive accuracy (e.g., Drawve, Thomas, & Walker, 2016; Ohyama & Amemiya, 2018). In order to enhance RTM with sociodemographic indicators, Drawve, Thomas, & Walker, (2016) developed an Aggregated Neighborhood Risk Index (ANROC). This indicator is a composite score of RRVs and sociodemographic indicators in a neighborhood. It is used to monitor the crime risk in an area in each time step. In another complementary study, Chillar & Drawve (2018) extend RTM to examine how calculated RRVs change based on police division and shifts. The main problem with using RTM concerns its monolithic risk indicator. This indicator does not capture the fluctuating spatial influence of urban features on crime within a spatial extent. Furthermore, this indicator is not sensitive to spatiotemporal crime risk differences. For example, it assumes a citywide spatial influence of a pub. However, the magnitude of this spatial influence may significantly vary from one region to another (Wheeler & Steenbeek, 2021). Therefore, a proper

indicator of spatial influence should capture the localized spatial influence during the day or week.

2.3.4 Prediction of crime counts and possibilities

Spatiotemporal General Additive Model (ST-GAM) (Wang & Brown, 2012) is the primary method of this category. This method is unique in terms of its ability to incorporate both environmental and societal factors into crime prediction models. Its basic premise is that the probability of crime occurring in a spatial unit in the next time step can be calculated given all the crime-related features are provided to the model. In addition, ST-GAM allows the development of local models for subregions- Local Spatiotemporal General Additive Model (LST-GAM). A common problem with using LST-GAM, however, is the lack of an explanation regarding how to determine the optimal number of subregions.

In addition to ST-GAM, the regression-based methods may also be included in this group. These methods aggregate the indicators of the social and physical environment, and selected crime counts at selected spatial units. They then quantify the effects of these block attributes (e.g., number of bars, schools, concentrated disadvantage, etc.) on crime by their estimated coefficients in the fitted regression equations (e.g., Bernasco & Block, 2011; Bernasco, Ruiter, & Block, 2017; Jones & Pridemore, 2019). A general problem with the methods in this group is its incapability of reflecting intertwined relationships between the included factors.

2.4 DL-based crime prediction methods

Using machine learning techniques in crime hotspot prediction is not a new phenomenon (Olligschlaeger, 1997; Corcoran, Wilson, & Ware, 2003). Over the last three decades, researchers have leveraged various techniques within the realm of crime hotspot

prediction. The majority of these techniques exploit conventional machine learning algorithms such as ANN, SVM, logistic regression, and ensemble methods (e.g., Random Forest) (e.g., Kadar, Maculan, & Feuerriegel, 2019; Rummens, Hardyns, & Pauwels, 2017; Zhang, Liu, Xiao, & Ji, 2020). Recently, deep learning algorithms have also been added to the toolbox of the researchers and practitioners who are interested in predicting the future crime hotspots (e.g., Zhuang, Almeida, Morabito, & Ding, 2017; Kang & Kang, 2017; Huang, Zhang, Zheng, & Chawla, 2018; (Jin, Sha, Feng, Cheng, & Huang, 2021). The DL-based algorithms were often used to predict grid-shaped crime hotspots. These algorithms employ covariates by aggregating external factors (e.g., 311 calls, POI) at these grids. The problem with this approach is threefold. First, the sizes of these grids are often selected arbitrarily (Rosser, Davies, Bowers, Johnson, & Cheng, 2017). This arbitrary selection results in one of the most notorious problems in spatial analysis, MAUP (Cheng & Adepeju, 2014) that leaves the performance of these models unstable based on the grid size. Second, these units do not match with real world police patrolling practices that are constrained to street networks. The grids that do not fit into these networks complicate the patrol route planning (Rosser, Davies, Bowers, Johnson, & Cheng, 2017). Lastly, human movement in the urban context is heavily constrained to street networks as well (Groff & Lockwood, 2014). Placing unrealistic grids above a constrained network space does not reflect the actual behavior of crime risk. Acknowledging the need to accommodate the crime risk on street networks, Zhang & Cheng (2020) proposed a graph learning approach to predictive crime hotspot mapping. In the first step, they created a graph consisting of the midpoints of street segments located on a street network. To weight the edges in this graph, they used a Gaussian kernel weighting function that assigns inverse weights to the edges based on their length

in network distance. To overcome the challenges associated with learning from the sparse datasets in Deep Learning (DL), they smoothed crime counts recorded in segments in a way that they can be fed into graph-based DL models. Lastly, they interpreted these smoothed crime counts as graph signals that can be learned through graph learning algorithms. The predictive hotspot mapping then became the task of predicting crime hotspot in the form of graph signals in the next time step. This translation renders graph learning algorithms applicable to the field of predictive crime hotspot mapping. Surprisingly, the studies using graph learning algorithms in crime hotspot prediction are nonexistent, except GLDNet developed by Zhang & Cheng (2020).

2.4.1. Reformulating crime prediction as a graph learning task

Spatiotemporal graph learning algorithms have thus far proved efficient in solving a wide range of real-world problems across many domains. While many researchers have used it in traffic speed prediction (e.g., Shleifer, McCreery, & Chitters, 2019; Chen et al., 2020); others have focused on taxi demand prediction (e.g.,), exchange rate forecasting (e.g., Wu et al., 2020), and wind speed forecasting (e.g., Khodayar & Wang, 2018). Their applicability to the spatiotemporal crime prediction on street networks, however, remains underexplored, except a remarkable study (i.e., Zhang & Cheng, 2020). Given the commonalities existing between the aforementioned tasks and spatiotemporal crime prediction in terms of predicting graph-based data, examining the applicability of graph learning algorithms is not a trivial task. Because, in both cases, predictions are made for the spatial units that are located on a graph. Furthermore, the values to be predicted participate in spatiotemporally interdependent relationships with

each other. That is to say, the prediction of a unit in each timestep influences the predictions of its neighbors in the current and next time steps.

By contrast, the main difference between these studies exists in the input structures. The graph learning datasets often consist of observations with continuous values collected through sensors on a network with a certain time interval. For example, a popular benchmark dataset, METR-LA, stores traffic speed measurements at the sensors located on the Los Angeles highways with 5-minute intervals. On the other hand, crime datasets tally the event counts in each segment within a time interval. Unsurprisingly, most observations in these datasets are zero, producing extremely sparse datasets. The main challenge here is that DL algorithms are ill-suited to the prediction of sparse datasets. A viable solution was proposed by Zhang&Cheng (2020) who reframe crime hotspot prediction on a street network as a task of Graph Signal Processing (GSP). GSP is an umbrella term that encapsulates a set of related tasks for processing data on irregular graphs (Ortega, Frossard, Kovačević, Moura, & Vandergheynst, 2018). Zhang & Cheng (2020) described a clear mapping from the components of GSP to the problem of crime prediction on street networks. For the first component, a street network can be represented as a Graph (i.e., G): G= (V,E,W). The second refers to the values to be predicted (e.g., crime risk, traffic speed), and the last one involves the task of learning these signals. One challenge in crime hotspot prediction domain is to represent sparse crime time series in a way that they can be fed into GSP algorithms. Therefore, data representation is of paramount importance in terms of integrating graph learning algorithms into predictive crime hotspot mapping.

2.5. This study

The contribution of the current study can be grouped under two main fields. First, it develops novel Distance-Aware Risk Signal Functions (DRSF) that model the dynamic spatial influence of urban features on different crime types within a spatial extent across the day and week. Second, by using these functions, the current study proposes a novel network-based environmental risk surface that can inform the network-based predictive crime hotspot mapping methods using DL. The first group has contributed to the third group of predictive policing methods by developing nonmonolithic crime risk indicators. These indicators capture spatiotemporal fluctuations in spatial influence across crime types. The second group has filled an empirical gap in the first group by complementing predictive crime hotspot maps with an auxiliary risk surface that represents the dynamic environmental crime risk around urban features across street networks. In addition, we proposed three additional theory-based event surfaces to enhance the predictive ability of the models: 311 service calls, park events, and other crime types. It thus overcomes the main challenges that were mentioned in the first group of predictive policing methods. The following sections will detail the contributions in both groups.

2.5.1. Distance-aware risk signal functions

In this study, we introduced the concept of the "crime risk station". Using a base station analogy, this concept models the spatial influence of urban features on crime as broadcasted risk signals at changing strength levels (i.e., distance decay effect) throughout a coverage area (i.e., spatial extent). In doing so, we adopted a crime-specific spatiotemporal approach to illustrate how the signal strength changes based on space, time, and crime type. Our methodology coupled the methods of network K analysis with segmented regression. The former helped us identify the significant network-based

spatiotemporal crime clusters around urban features. The latter models the spatial influence for each significant spatiotemporal cluster as a function of street network distance. Xu and Griffiths (2017) previously used a similar methodology in their study where they attempted to measure the spatial influence of a set of urban features on gun violence in Newark, NJ. In this study, we have built upon this study by examining how the spatial influence of a wide set of selected urban features (i.e., bus stop, fast food restaurant, gas station, grocery store, pub) on robbery and theft risk varies across various spatiotemporal configurations in Chicago, IL.

The urban layout differences within and between cities complicate the selection of suitable spatial extents and pose one of the greatest challenges for comparative spatial influence analysis amongst urban features. A tenable solution to that problem was proposed by McCord and Ratcliffe (2009) who developed the Intensity Value Analysis (IVA). The IVA operationalizes spatial influence within a given radial bandwidth that can be considered as an assumed spatial extent. The intensity values were calculated by inversely weighting counts based on the distance from the center. Using fixed bandwidths then enacts comparative analysis of spatial influence between regions. In a similar way, we proposed fixed network bandwidths. In these bandwidths, we operationalize the spatial influence through the distance-aware risk signal functions. The risk signal functions are the fitted segmented regression equations that estimate the crime risk in near areas. Past researchers utilized segmented regression to identify the inflection points in the spatial influence of urban features (Ratcliffe,2012; Xu &Griffiths,2017). Differently, we used it to estimate the crime density at each point within the given network bandwidth. From these functions, we devised two risk indicators: risk signal intensity score (RSIS) and risk signal strength score (RSSS). The

former represents cumulative crime density at the assumed spatial extent. The latter quantifies the spatial diffusion effect with the percentage change between a local effect (i.e., crime density at the exact location) and the RSIS. For a local effect of four and a RSIS of three, the RSSS is -33.3. This value shows a 33.3% decrease in spatial influence. Next, we compared these scores for each crime type across various spatiotemporal models with nonparametric statistical tests, Kruskal-Wallis, and Wilcoxon Signed-Rank test. Lastly, we examined how spatial influence interacts with the socieconomic characteristics across the regions with a correlation analysis between the regional risk scores and concentrated disadvantages of these regions.

This study makes several contributions to the existing literature. First, it filters insignificant crime clusters around urban features with a novel approach using network K analysis. Thus, it ensures the validity of the spatial influence. Second, it introduces "Distance Aware Risk Signal Functions (DRSF)" that models spatial influence on a network-constrained space rather than representing it with a monolithic value in a euclidean space. Third, it develops two novel network-based risk scores (i.e., RSIS and RSS) that characterize spatial influence within an assumed spatial extent. Fourth, it brings a crime-specific spatiotemporal approach to the comparative spatial influence analysis. Lastly, it proposes a correlation analysis that tests the relationship between spatial influence and social context across space, time, and crime types.

2.5.2. Developing a predictive crime hotspot mapping method

This study introduces several network-based auxiliary event surfaces that enhance retrospective crime hotspot prediction techniques with theory-based crime risk indicators. The first surface developed represents the dynamic environmental crime risk across street segments. To capture the risk, we developed a novel score, Segment

Segment Risk Score (SSRS). This score is calculated by multiplying (i.e., multiplicative) the local effects of the urban features located on a street segment. The reason we chose a multiplicative model is to incorporate existing spatial interactions between the urban features within the same spatial neighborhood (He et al.,2020). For example, the spatial influence of an ATM will be much higher in closer proximities to a pub than its influence when it is located nearby a police station. We quantified the individual local effects through a novel indicator, Local Risk Score (LRS). This score measures the cumulative crime density at 1.7 m (5.5 feet) away from the urban feature. We obtained the LRSs from DRSFs that we described above. We calculated multiple LRS values for each selected urban feature across the selected crime types at two different temporal aggregation levels: intraday (i.e., First Shift (FS) (00:00-07:59), Second Shift (SS) (08:00-15:59), and Third Shift (TS) (16:00-23:59) and weekly (i.e., weekday and weekend). To calculate these values, we fitted five different DRSFs for each urban feature-crime pair at the temporal levels. These values then represent the dynamically changing local effects of urban features on selected crime types. Lastly, these values are used to calculate a cyclical SSRSs that recur on a daily or weekly basis. As an illustrative example, the weekday theft SSRS of a street segment that hosts one grocery store and two pubs having LRS values of 3 and 4 respectively would be 48 (3x4x4). On the other hand, the weekend theft SSRS for the same segment would be 196 given the values of grocery stores and pubs increase to 4 and 7 respectively.

In addition to an environmental risk surface, we have developed three theory-based auxiliary event surfaces: 311 calls, park events, and the coexisting crime incidents. The first surface represents the number of 311 service calls from a street segment in a given time step. The second surface represents the park events that take place on street

segments. To add temporal extent to these public events that increase the crime risk across units by concentrating human activity around particular areas (e.g., Ristea, Kurland, Resch, Leitner, & Langford, 2018), we smoothed the number of these events in a way that the crime risk spans the event duration. For example, a park event that would last the whole weekend will be smoothed into two days. The last surface tallies the counts of other crimes in a street segment during a period. For example, in predictive hotspot mapping for robbery, theft incidents are used to create an auxiliary event surface. Or vice versa. Other crime risk surface serves to improve the performance of crime hotspot prediction due to existing spatiotemporal dependencies between different types of crime patterns (Mohler,2014).

In the second step, we incorporated our auxiliary event surfaces into a set of graph learning algorithms (i.e., GraphWavenet, Spatiotemporal Graph Convolutional Network-STGCN) and a baseline (i.e., LSTM). For each algorithm, we created eight models for each crime type that produced daily and shift predictions with or without auxiliary event surfaces. We measured the predictive performance of these models with Mean Hit Rate (MHR) (Zhang & Cheng, 2020). This indicator denotes the mean percentage of crime incidents that fall in the predicted hotspots during test days. For hyperparameter tuning, we described a two-step procedure that involves optimizations of data representation and model parameters respectively. In the first step, we experimented with three different parameters related to data representation: time window, smoothing coefficient, and spatial bandwidth. The first parameter simply determines how many previous time steps should be used to predict the next one, the second and the third parameters adjust the extent of crime risk in space and time. In the second step, we experimented with 11 different values of learning rate (ρ) which is maybe the most important parameter in

hyperparameter tuning of deep learning algorithms (Bengio, 2012). Lastly, we compared the results. The current study defines itself as a graph-based predictive crime hotspot mapping method using DL algorithms. Differently, we enhanced graph DL-based crime prediction algorithms with theory-based auxiliary event surfaces that can inform the models on the upcoming changes in different crime risks across a day and week. Also, it adapted two different graph learning algorithms from the traffic speed forecasting domain into the predictive crime hotspot mapping. Another novelty is the shift model developed to predict intraday crime hotspots in an area of interest. Lastly, it introduced a hyperparameter tuning procedure that optimizes data representation and model parameters in two steps. In sum, this study is the first attempt to enhance graph-based predictive crime hotspot mapping algorithms with a set of theory-based auxiliary event surfaces.

Table 1. The Key Research

Study	Method	Unit of Analysis	External Variable	Prediction Horizon(s)
(Bowers, Johnson, & Pease, 2004)	KDE	Grid	-	2-days
Gerber (2014)	KDE	Grid	Topics derived from the tweets	Daily
(Rosser, Davies, Bowers, Johnson, & Cheng, 2017)	Network KDE	Street Segment	-	Daily
(Huang, Zhang, Zheng, & Chawla, 2018)	Deep Crime	Police District	POI, 311 calls	Daily
Zhang &Cheng (2020)	GLDNet	Street Segment	-	Daily
This study	*Graph Wavenet, STGCN*	*Street Segment*	*SSRS, 311 Calls, Park Events, Other Crime*	*Daily, Shift*

CHAPTER 3

METHODOLOGY

This chapter starts with a brief description of the study setting. Next, we outline how we extract the unit of analysis in this study. Then, we describe the procedure that we used to create graphs consisting of the extracted units of analysis. The following section describes the methodologies of two main parts in subsequent subsections. The first subsection details the methodology that we follow to develop our DRSFs. We begin by starting with the description of the datasets and continues with the introduction of the techniques we used (i.e., bivariate network K analysis and segmented regression). It concludes with the analytical framework that explains how we apply these techniques to these datasets in creating DRSFs. The second section concerns the methodological steps in developing a dynamic predictive crime hotspot mapping method enhanced with a set of theory-based auxiliary event surfaces. In the first step, we describe how we create the auxiliary event surfaces that will be incorporated into the crime hotspot prediction models. In the second step, we provide the details on the graph learning algorithms that we used to predict crime hotspots on a map. The following section is where we explained how we measured the performance of the selected algorithms. After explaining the two-step hyperparameter tuning procedure, the last section will detail the the experimental setup.

3.1. Study setting

With a population of approximately 3 million, Chicago is one of the largest cities in the U.S. The city's land coverage is 606.1 km2. Within this area are 200 neighborhoods and 77 community areas (About Chicago: Facts and Statistics, 2021). An additional regional

division are the sides of Chicago (i.e., administrative districts used for urban planning purposes). There are nine sides of Chicago: Center Side (CS), Far North Side (FNS), Far Southeast Side (FSES), Far Southwest Side (FSWS), North Side (NS), Northwest Side (NWS), South Side (SS), SouthWest Side (SWS), and West Side (WS). These sides have often been characterized by unique sociodemographic characteristics. As an example, the CS is the heart of the city accommodating the main business, shopping, and entertainment districts of the city. The FNS hosts the liveliest neighborhoods mainly populated by immigrants. And the NS contains the most affluent neighborhoods in the city. On the other end of the spectrum, WS and SS are the sides that chronically suffer from the highest social disadvantage levels in Chicago (Keating, 2008; Sampson, 2012). These dramatic differences amongst the sides bring about notable differences in terms of crime levels. In one of the earliest studies that examine the crime variability at the side level, Block (1993) found that 55% of gang-related homicide and 35% of nonlethal gang-related offenses occurred on the WS. Similarly, Schnell, Braga, & Piza, (2017) identified the largest violent crime hotspots on the WS and SS of the city. In a recent study, Rosser & Cheng (2019) revealed that a citywide crime prediction algorithm (i.e., SEPP) is not robust to localized crime patterns across the sides of Chicago. The citywide crime hotspot predictions, therefore, remain invalid on most of the sides. This is a study that underlines the necessity of developing regional crime hotspot prediction methods that can make realistic crime predictions across the sides. Therefore, we selected the sides of Chicago as our study area where we can implement our regional crime predictions. Figure 1 displays the sides and the related community areas of Chicago.

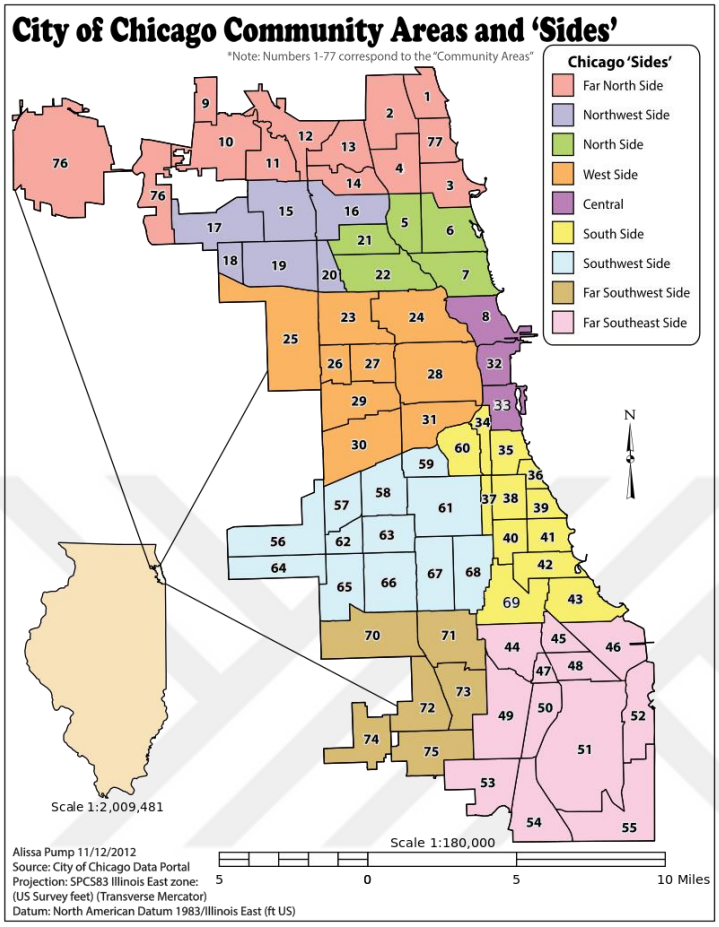


Figure 1. The sides of Chicago (Adapted from City of Chicago Data Portal)

3.2. Unit of analysis

The unit of analysis of the current study is the street segment that refers to the two-sided street part between two intersections (Weisburd, Groff, & Yang, 2012). Street segments represent optimal microsocial systems that truly reflect the daily interactions between the residents in an area. Also, they are small enough to avoid spatial aggregation errors and large enough to have a measurable and detectable variability in the number of crimes (Vandeviver & Steenbeek, 2019).

We extracted the street segments through the use of an open-source Geographical Information System Software, QGIS 3.6.2. First, we created an Open Street Map canvas

for Chicago. Figure 2 illustrates a sample street network retrieved from the Open Street Map.



Figure 2. A snapshot of the city of Chicago on Open Street Map

Second, we retrieved the street center lines in a shapefile format from City of Chicago Open Data portal and create another layer on top of the open street map. This layer basically consists of the line features that correspond to what we defined as street segments above. These features have many attributes such as segment ID, object IDs, GPS coordinates of involved intersections, length, community area, and type of the segment (e.g., street, alleyway, boulevard etc.). When illustrated on a map, this layer represents the street network of Chicago. Using this layer, we build a graph where each node corresponds to a street intersection. After enumerating the node numbers, we simplified the graph by removing the redundant edges between the nodes. This procedure produced a total of 36446 nodes. Then, we encoded each line (i.e., segment)

by the IDs of its start and end nodes. (e.g., 2333- 432). For this, we extracted the locations where the nodes and street network intersect, overlap, or cross with each other. This procedure resulted in a total of 56334 street segments with labels. In the last step, we enumerated the resulting street segments to be referred to later. For example, the ID of the segment with the label of 15895-22890 is 1, 23684-7643 is 2, so on so forth. Figure 3 displays a street network with enumerated intersections.



Figure 3. A sample street network with enumerated intersections

3.3. Creating graphs from the street network

A street network can be represented as a graph, G= (V, E, W). Here, V (i.e., vertices), V= {v1, v2,….. vn}, corresponds to the street intersections, E refers to the edges between these intersections (i.e., street segments), W is the edge weight and can be assigned to the length of the segment. If there exists order information between the edges (i.e., directionality), the network is “directed”, otherwise it is called “undirected”. The directionality could be an issue for the cars moving along a street network since all the streets may not be two ways in a street network. On the other hand, such limitations

do not apply to pedestrian movement in an urban context where people freely ebb and flow. Therefore, we created an undirected weighted graph to represent the street network in this study. Figure 4 illustrates the difference between directed and undirected graphs.

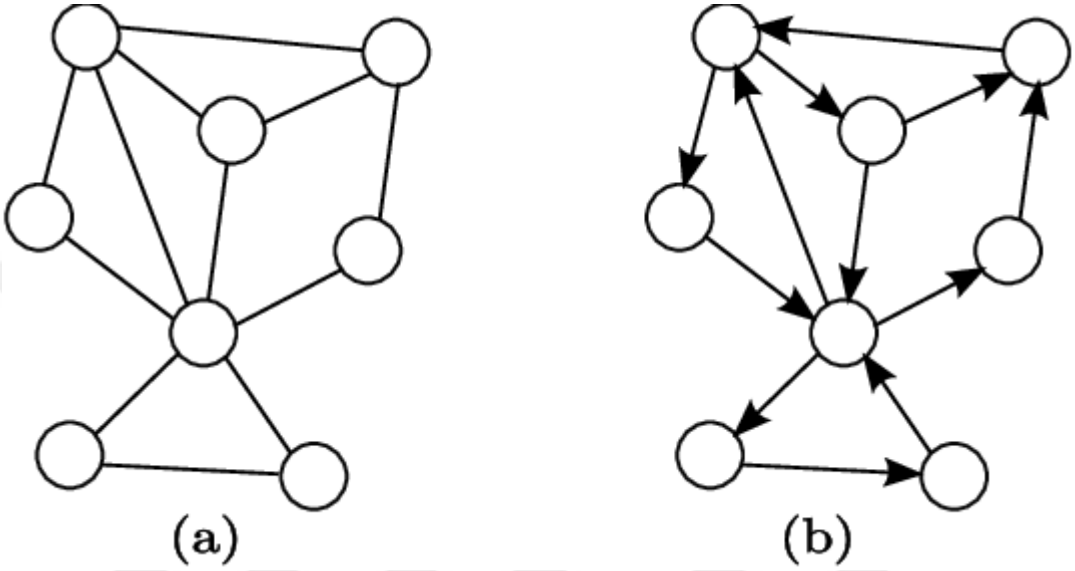


Figure 4. An example of (a) undirected graph and (b) directed graph (adapted from Fionda & Palopoli, 2011)

In the first step, we eliminated the duplicate segments to ensure the uniqueness of the street segments in the street network. In doing so, we retrieved the segments with the same start and end nodes. Their length is our criterion for selecting redundant segments. Accordingly, we randomly select from redundant segments having equal lengths. In case there are two different length values for the same segment, we selected the one with a shorter length since we observe the longer ones are often inputted with unreasonable values such as 8000 meters. Also, we removed many segments with a length of zero. Lastly, we eliminated the segments with self-recurring loops where the start and end nodes are the same. This preprocessing resulted in a street network with 56310 valid segments.

To create a graph from these segments, we used networkX (Hagberg et al., 2008), a python library specifically developed for creation, manipulation, and analysis of the networks. Based on the reviewed street segments that we labeled with the start and end intersections (i.e., nodes), we created an initial street network. One important point here is to make sure we have a fully connected network. This is because the disconnected networks amount to the existence of some nodes that are not accessible to some other nodes. These nodes together represent an independent subnetwork that can be called "components" of a network. Having a network with independent components may severely complicate the distance calculations (Okabe & Sugihara, 2012). Therefore, we extracted the largest component that accommodates most of the street segments to represent the street network. While examining the graph we initially created with 36446 street intersections and 56310 edges, we found 46 components. From these components, we selected the largest one that includes 36232 nodes. Table 2. summarizes the resulting graph.

Table 2.. The Summary of the Resulting Street Graph of Chicago

Measure	Value
#Nodes	36232
#Edges	56130
#Average Degree	3.0984

After creating the main graph, we geocoded the point features (e.g., crime locations). One challenge we faced in this task was the features that do not reside on the graph. As a solution, we used linear referencing that simply refers to a convention to store point

features in line segments. This method snaps all the GPS coordinates to the nearest segment of the graph. In the first step, we found the projection of a given coordinates on a line segment. Then we interpolated its coordinates on that segment. As a result, we obtained the corresponding coordinate of a point on that segment. This procedure allowed the calculation of network distances between different kinds of point features (e.g., crime, urban features). Figure 5 visualizes the snapping of a point.

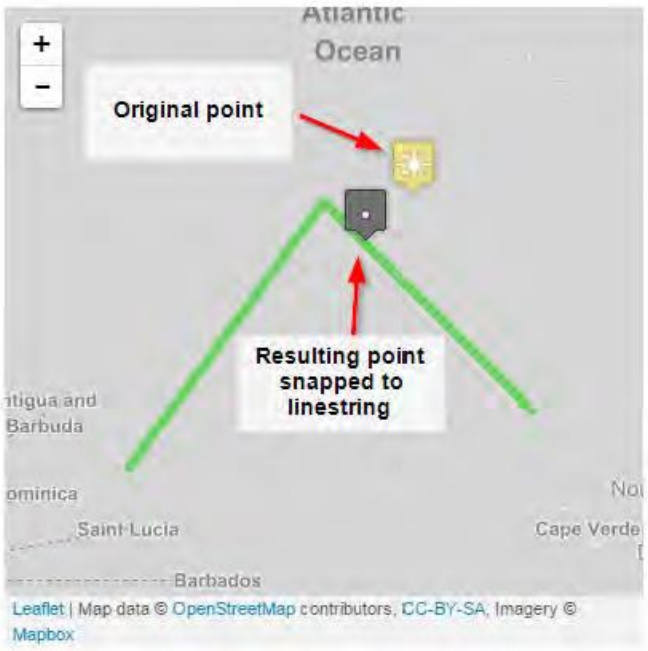


Figure 5. Snapping a point to a segment (Diener,2015)

3.3.1. Creating subgraphs for each side

In the current study, we extracted the subgraphs that represent the street networks of the sides of Chicago. To do so, we first performed spatial join operations on the street segments and the community areas. We retrieved the boundaries of community areas from the community areas boundary shapefile. As illustrated in Figure 1, one community area can only belong to one side. In that case, mapping street segments to the

community areas amounts to mapping them to the sides. Using these associations between the spatial units, we created separate side graphs. The spatial join operation was performed by using NNQGIS plug-in of QGIS 3.6.0 Software. Like what we did in creating the main graph, we obtained fully connected networks by eliminating all the street segments fallen outside the main component. Descriptive information is provided in Table 3. The resulting subgraphs are displayed in Figure 6.

Table 3. The Statistics for the Resulting Subnetworks

Sides	#Nodes	#Edges	Bounding Box Coordinates	Bounding Radius (m)	Network K Distance Chunks (m)
C	1680	2459	(41.91, −87.60; 41.84, −87.65)	9977.9	19.4
FN	5407	8151	(42.02, −87.63; 41.93, −87.93)	29611.9	57.7
FSE	4873	7227	(41.75, −87.52; 41.64, −87.66)	20438.6	39.8
FSW	3509	5308	(41.75, −87.63; 41.66, −87.74)	15932.5	31.0
N	2691	4145	(41.96, −87.62; 41.91, −87.73)	14001.2	27.2
NW	2693	4423	(41.96, −87.69; 41.91, −87.83)	16737.1	32.6
S	4040	6023	(41.85, −87.54; 41.74, −87.66)	15770	30.7
SW	5032	7976	(41.78, −87.62; 41.75, −87.80)	17380.9	33.8
W	6762	10395	(41.92, −87.63; 41.81, −87.80)	20760.5	40.4

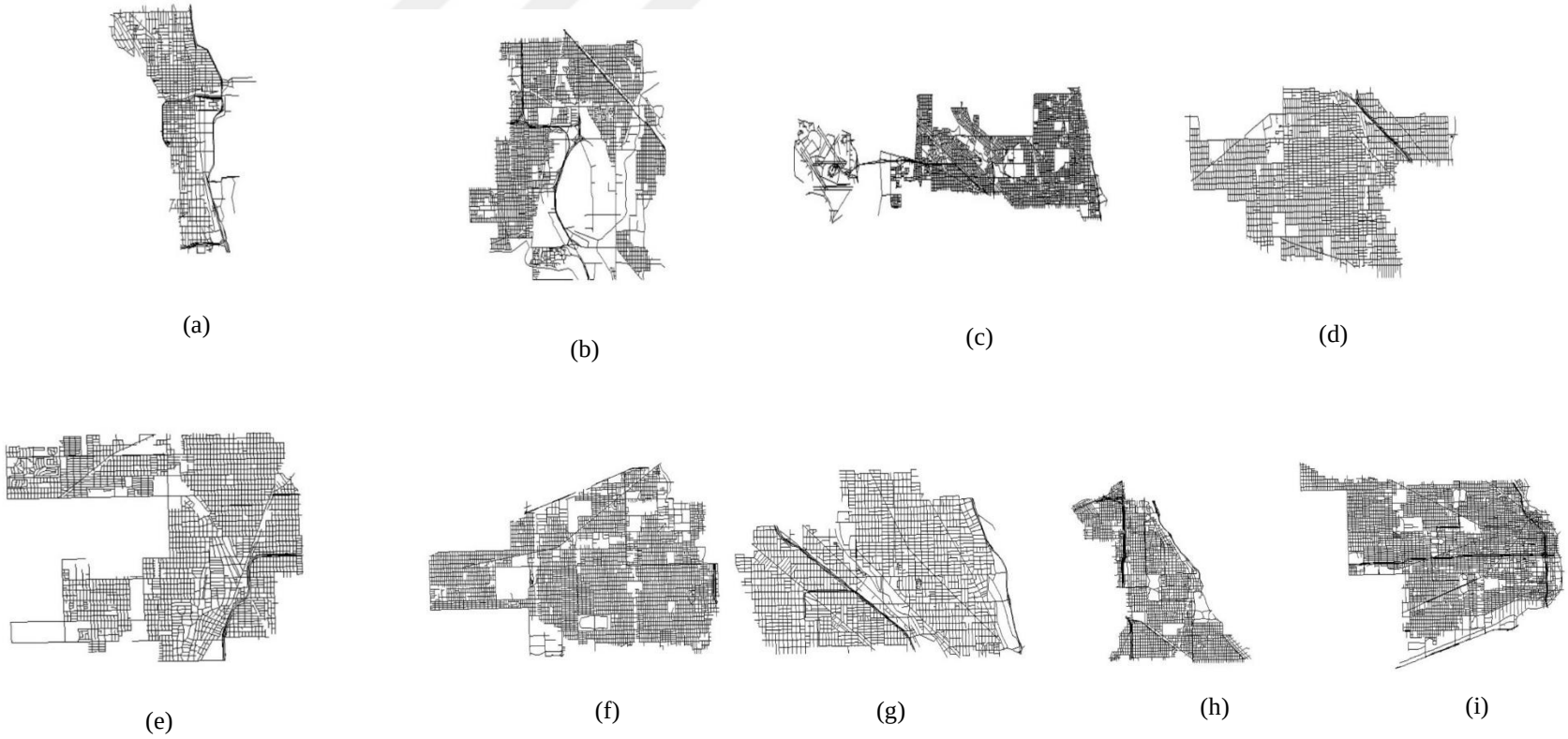


Figure 6. The side street networks (**a**) CS, (**b**) FSES, (**c**) FNS, (**d**) NWS, (**e**) FSWS, (**f**) SWS, (**g**) NS, (**h**) SS, and (**i**) WS.

3.4. Developing distance-aware risk signal functions (DRSFs)

This subsection summarizes the methodology of DRSFs that capture spatiotemporal fluctuations in crime risk around urban features across crime types. The datasets in this part are different from the ones we used in the second part. Therefore, this section starts with describing the first part datasets and continues with the introduction of the main techniques that we used to develop our DRSFs.

3.4.1. Data

3.4.1.1. Chicago Crime dataset

In this study, we used a sample from Chicago Crime dataset[1] that records all the crime incidents since 2001. Each record in this dataset defines a crime incident with 22 attributes including crime ID, date-time, x and y coordinates, the primary description of the crime, the neighborhood, and the community area. Our sample consists of all theft (n = 64024) and robbery (n = 9685) incidents in 2018. To examine the spatiotemporal behavior of crime risk around urban features, we created spatiotemporal models by aggregating selected crime incidents at sides and three temporal levels. In the first level, we mapped the crime incidents to the relevant side street networks without any temporal dimension (default). In the second level, based on the hour of occurrence, the incidents were distributed across the shifts during a day, the first shift (FS) from 00:00 to 07:59, the second shift (SS) from 08:00 to 15:59, and the third shift (TS) from 16:00 to 23:59. These shifts were defined based on the working watches of Chicago police officers (Payroll and Timekeeping—Attendance, 1996). In the last level, we grouped the incidents as weekend (WE) and weekday (WD). The WD included all the incidents that occurred between third shift of Friday and the first shift of Monday(excluded). All the

[1] Retrieved from https://data.cityofchicago.org/

others were assigned to WE group. These divisions result in a total of 54 spatiotemporal models (9 default + 9 × 3 intraday levels + 9 × 2 weekly levels). Crime distribution across these models is displayed in Figure 7.



Figure 7. Crime distribution across the spatiotemporal model

3.4.1.2. Location of urban features

Geolocating the urban features on the subgraphs that we created in the first section required a careful procedure. The procedure starts with retrieving the Chicago business licenses dataset. This dataset contains the GPS locations of the applicant businesses. It also provides a definition of the business activity of license applicant. A problem with using this dataset is the multiple license applications submitted by the same business. As a solution, we randomly selected one record from these multiple applications. Lastly, using activity definitions, of the unique instances, we obtained the locations of four

different urban features: fast-food restaurants, pubs, grocery stores, and gas stations. The locations of bus stops were obtained from the CTA bus stops shapefile. The resulting dataset contains fast-food restaurants (n = 402), grocery stores (n = 1330), gas stations (n = 350), pubs (n = 810), and bus stops (n = 10900). We selected these urban features based on a volume of empirical support derived from Chicago studies that showcased their significant influences on crime risk (e.g., Bernasco & Block, 2009;Bernasco & Block, 2011; Bernasco, Ruiter, & Block, 2017; Barnum, Caplan, Kennedy, & Piza, 2017; Kennedy, Caplan, Piza, & Buccine-Schraeder, 2016). Figure 8 displays the spatial distribution of these urban features across the sides of Chicago.

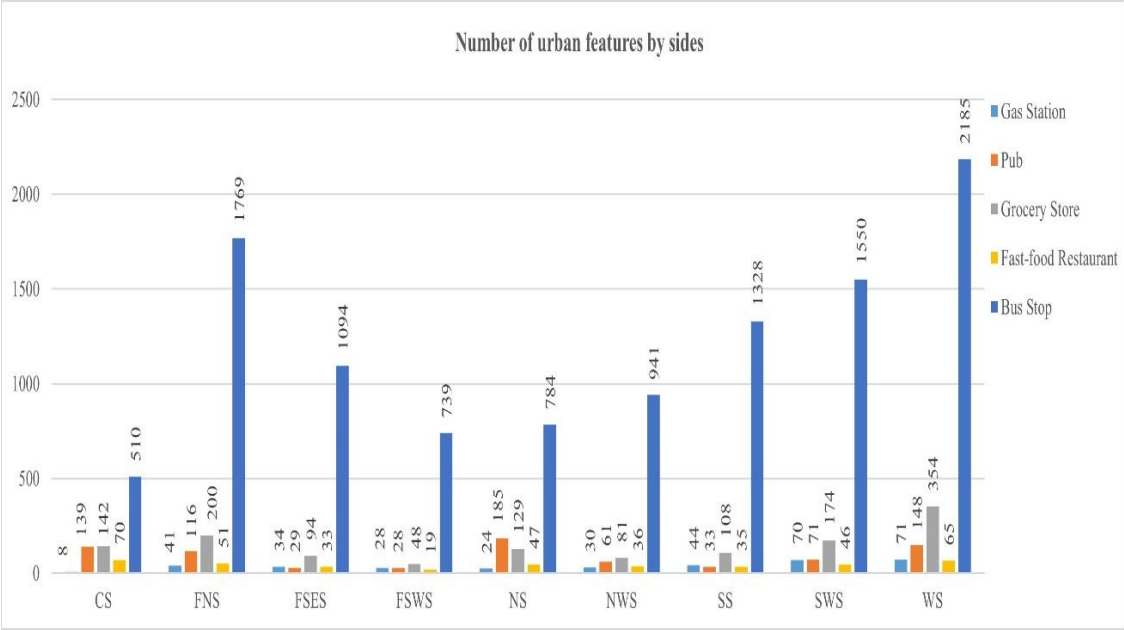


Figure 8. The distribution of the selected urban features across the sides of Chicago

3.4.1.3. Concentrated Disadvantage

Concentrated disadvantage (CD) is a popular indicator of social disorganization in a neighborhood (e.g., Jones & Pridemore, 2019; Nobles, Ward, & Tillyer, 2016). Past research has repeatedly confirmed its linkage to the spatial influence of urban features on crime in various settings (e.g., Stucky & Ottensmann, 2009). CD of a neighborhood is not directly measurable. Rather it is often measured by observable indicators. Therefore, we used four disadvantage indicators in this study: the percentage under 15 and above 64 years of age, the percentage of unemployed residents, the percentage of households below the poverty line, and the median income. Next, we performed a Principal Component Analysis (PCA) to examine whether these indicators could be loaded to a single factor: CD. Then, since the calculated CDs were not available at the side level, we used a weighted apportioning method (Kim, 2018) to impute the side-level CD data. This method calculates a population-weighted mean CD value from the values of the included community areas as the side CD. Table 4 summarizes the results for CD.

Table 4. The PCA Results of CD

	Mean	SD	Eigen Value	Factor Loadings	Cronbach's Alpha
Concentrated Disadvantage (CD)	*0.000*	*1*	*3.105*		*0.901*
% under 15 above 64 years	38.074	6.545		0.753	
% unemployed	5.973	3.075		0.907	
% less than poverty	29.936	13.964		0.925	
Inverted median income	51038.798	23754.78		0.926	

3.4.2 Network K function

Network K function (Okabe & Yamada, 2001) is a network variant of Ripley's K function that summarizes and analyses a point pattern on a homogenous infinite plane with the Euclidean distance. Differently, the network K function handles the point patterns residing on a finite irregular network with the shortest path distance. There are two types of Network K function: auto and bivariate. While the former examines the spatial association within the elements of the same point pattern, the latter tests whether the spatial distribution of a point pattern influences the distribution of another. Given the aim of the study, we selected the latter to examine how the selected urban features influence different types of spatial crime distribution in near areas across space and time. An advantage of using this function is its ability to quantify the degree of crime clustering around an urban feature at a given distance. Theoretically, it can be formulated as:

$$K^{ba}(t) = \frac{1}{\rho_a} E(\text{the number of points A within network distance t of a point } b_i \text{ in B}) \quad (1)$$

where E(.) denotes the expected value with respect to bi, b_i, …b_n ($b_i \in$ B). bi is obtained from a binomial point process, and $\boldsymbol{\rho_a}$ is the density of points a, $\boldsymbol{\rho_a} = (\frac{\mathbf{n_a}}{|\mathbf{L_T}|})$. Here, $|\mathbf{L_T}|$ is the total length of the street segments in a network. For observed point processes of different types, the observed network cross K function of A (i.e., crime) relative to B (i.e., urban feature) can be formulated as:

$$\hat{K}^{ba}(t) = \frac{|L_T|}{n_a n_b} \sum_{i=1}^{n_b} \left(\text{the number of points of A on } L_{b_i}(t)\right) \quad (2)$$

That $\widehat{\boldsymbol{K}}^{\boldsymbol{ba}}(\mathbf{t}) > \mathbf{K}^{\mathbf{ba}}(\mathbf{t})$indicates a cluster of a around b. The opposite shows that a is dispersed around b. The comparison of $\boldsymbol{K}^{\boldsymbol{ba}}$and $\widehat{\boldsymbol{K}}^{\boldsymbol{ba}}$ is only possible with the calculation of the expected value in (1). This value can be calculated by an analytical evaluation method testing the complete spatial randomness (CSR) hypothesis. This hypothesis assumes independent and identical distributions for different types of point patterns based on a binomial distribution over a network space. In this study, we adopted a Monte Carlo simulation approach to test CSR. This simulation generates 39 complete spatial random point patterns for crimes and urban features located on a network, $\mathbf{L_T}$. For each pattern, it calculates the crime counts on a network distance t from an urban feature. The minimum and maximum values at distance t amongst these simulated patterns are the critical upper and lower values at the $\alpha = 0.05$ significance level (Baddeley et al., 2014). A value above the upper value indicates a significant clustering, below the lower value indicates a significant dispersion, all other values indicate insignificant clusters. We performed this analysis for each of 54 spatiotemporal models using the spatstat package (Baddeley & Turner, 2005) on R. Thus, we were able to exclude all the insignificant spatial crime clusters around the selected urban features before moving on to the DRSFs.

3.4.2 DRSFs, RSIS and RSSS

Segmented regression is a special form of regression analysis that is developed to model the dynamic relationships between an independent (x) and a dependent (y) variable across different intervals of x values. This analysis identifies the change points beyond which the coefficients in the regression equation cannot be preserved. This equation is formulated as:

$$E[y|x] = \beta_0 + \beta_1 x + \delta_1 (x - \tau_1)^+ + \cdots . + \delta_k (x - \tau_k)^+ \quad (3)$$

where τ_k denotes the unknown changepoints whereby $(\mathbf{x_i} - \boldsymbol{\tau_k})^+ = (\mathbf{x_i} - \boldsymbol{\tau_k}$if $(\mathbf{x_i} - \boldsymbol{\tau_k})$>0. $\boldsymbol{\beta_0}$, $\boldsymbol{\beta_1}$, $\boldsymbol{\delta_1}$… $\boldsymbol{\delta_k}$are coefficients obtained from a method of permutation test (Kim et al., 2000).

In this study, each observation (xi, yi) consisted of a network distance and a corresponding cumulative crime density. The linear K cross function that we used in network K analysis outputs expected number of events at 513 equal bandwidths, marking equal distances within a bounding radius (i.e., maximum shortest path distance between any two points in a linear network). The bounding coordinates, radius, and distance chunks corresponding to the distance between these 513 points are reported in Table 3. To estimate the segmented regression models, we used an adequate number of these observations that cover at least 400 m in a side network. For example, it uses 10 observations for the West side (40.4 × 10 = 404 m) with a chunk distance of 40.4 m. We used Joinpoint Regression program to estimate segmented regression equations that are the working DRSFs. An observation here consists of a distance and cumulative crime density pair that correspond to x and y respectively. A sample DRSF is displayed in Figure 9. The accuracies of the DRSFs are measured by RMSE and MAE.

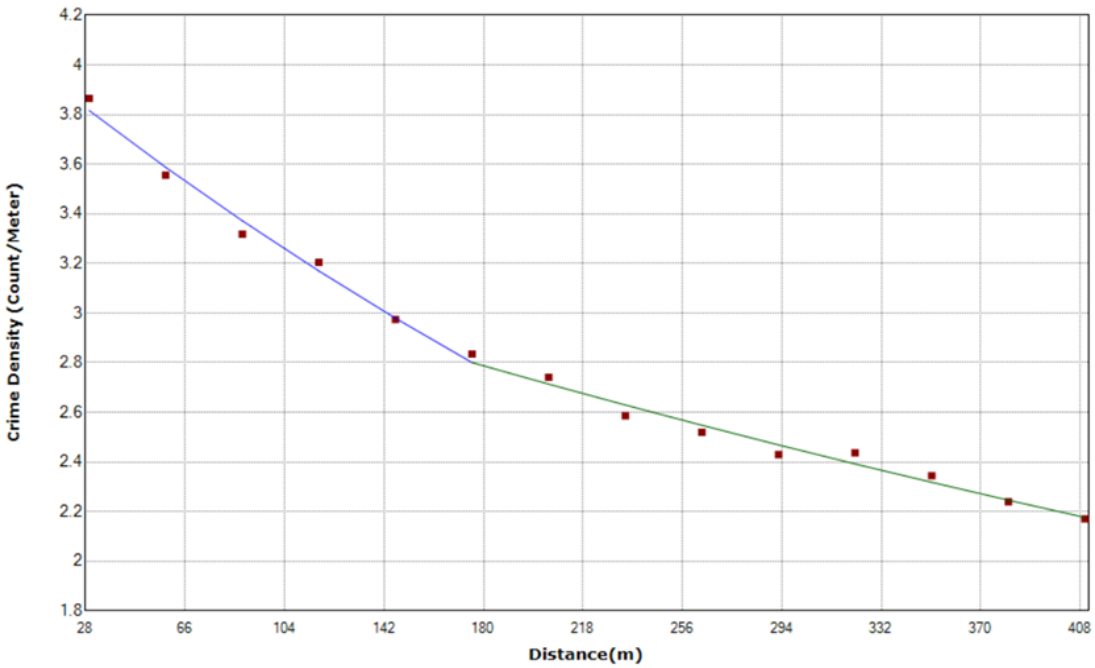


Figure 9. A sample DRSF.

In the next step, we devised two novel risk indicators (i.e., RSIS and RSSS) that allow a comparative analysis between DRSFs across spatiotemporal models. The first indicator denotes the estimated cumulative crime density at a maximum distance from an urban feature (i.e., spatial extent). In this study, we determined this maximum distance as 402.4 m (i.e., a quarter-mile). This value is borrowed from the transportation research that attempts to demarcate the maximum distance people would be willing to walk until the nearest public transport stations (Nelessen, 1994). Crime research interpreted this value as a boundary for the spatial interactions between urban features and crime (Groff, 2011). Accordingly, many researchers anchor the spatial extent at this distance (Weisburd, Groff, & Yang, 2012; Hart & Miethe, 2015; Caplan, 2011; He et al., 2020). RSSS can be calculated by the following formula:

$$\mathrm{RSSS} = \frac{\mathrm{RSIS} - \mathrm{E[y|Origin]}}{\mathrm{E[y|Origin]}} * 100 \ (4)$$

Here, $\mathrm{E[y|Origin]}$ quantifies the magnitude of the local effect of an urban feature on crime with the estimated crime density at 5.5 feet (1.7 m), $\mathrm{E[y|1.7]}$, Geocoding crime incidents at 1.7 m is a common practice amongst law enforcement agents due to identification purposes (Ratcliffe,2012). RSSS can be interpreted as an indicator of how much of the local effect could be retained throughout a spatial extent (i.e., spatial diffusion effect). Its values lie within $(-\infty, \infty)$ interval. Negative values confirm the distance decay, and positive values showed an increasing crime density throughout the spatial extent. We imputed zeros into RSSS and RSIS values of urban features in case of insignificant crime clusters since the observed clustering is not significantly different from the one that would be obtained if the crime is distributed randomly.

3.4.3 Analytical procedure

We begin our analysis by creating 54 different samples on the side street networks for each spatiotemporal model that we described above. Next, we test the significance of crime clusters in these models by using a bivariate network K analysis testing CSR with a 39 steps Monte Carlo simulation. Then, we fit separate segmented regression equations for each significant crime cluster. These equations model the spatiotemporal influence of an urban feature on crime and embody our DRSFs. Next, we characterize the DRSFs in each model through novel indicators: RSIS and RSSS. Then we compared these values across spatiotemporal models for each crime type by using nonparametric statistical tests (i.e., Wilcoxon signed-rank and Kruskal-Wallis test). In the last step, we examined the linkage between the spatial influence of an urban feature and the CD of the surrounding neighborhood across the spatiotemporal models through a correlation analysis.

3.5 Developing DL-based predictive crime hotspot prediction models,

The next section of the study aims to develop predictive crime hotspot mapping methods enhanced by theory-based auxiliary event surfaces. First, we introduce the unique datasets that are used throughout this part. Second, we describe how we created the proposed event surfaces. Third, we provide an overview of the selected graph DL algorithms. Lastly, we mention the details of the parameter tuning phase and experimental setups.

3.5.1. Study setting

We implemented our crime hotspot prediction models on the Center Side (CS) of Chicago. There are several reasons for this selection. First, CS is the epicenter of the city with the main business and entertainment districts as well as historical sites. Maintaining public safety in the CS is, therefore, not only increases the overall public safety but also ensures uninterrupted urban functioning of the city. Second, although this side covers only 2.5% of the total area, it accommodates nearly 6% of the total population. The population density in the CS (11049.69 person/km^2) is much higher than the average value for Chicago (4664.303 person/km^2). These numbers point to the large benefits derived from increasing public safety by only focusing on a relatively smaller area. Lastly, in the 2014-2019 period, the 5-year crime count per 1000 person was 466.6542 on this side, much higher than the city average (374.868). For the selected crime types on the CS, these values were 18.10006 for robbery and 300.4349 for theft. While the robbery value remained slightly below the city average (19.58536), this value is much higher for the theft (113.7040). Thus, any crime hotspot prediction method may yield a leveraged benefit to decreasing overall crime levels across the city. Lastly, we were severely limited by the available computational resources. We ran our experiments on

Google Colab Pro that provides its users with a GPU of 15 GB, and 147.15 GB RAM. From the nine sides of Chicago, the CS is the only side on which we managed to implement our crime hotspot prediction algorithms for it consists of a relatively smaller number of segments.

3.5.2. Creating theory-based auxiliary event surfaces

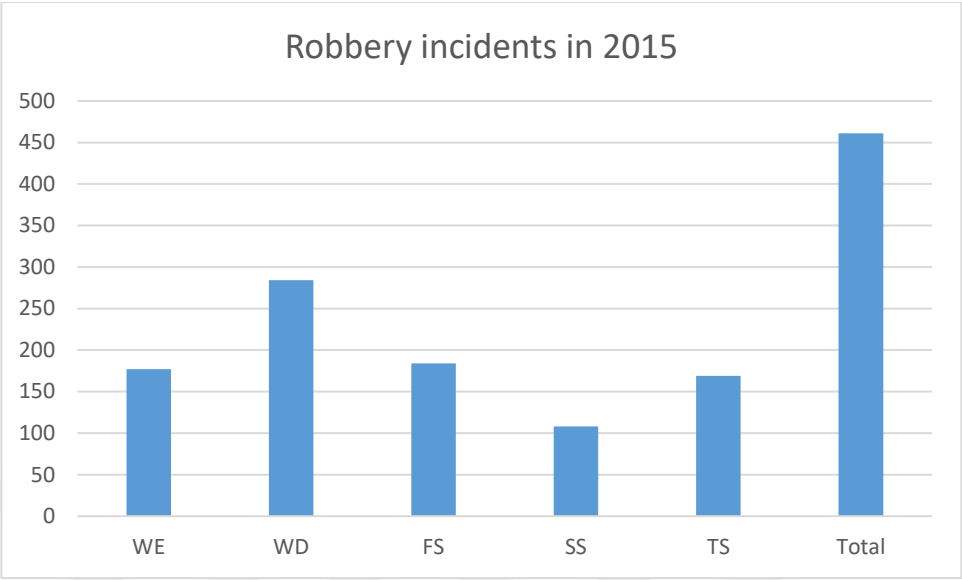
This section describes how we created theory-based auxiliary event surfaces that enhance the predictive crime hotspot mapping. These surfaces keep the records of the event intensity in street segments during a period in tabular format. The auxiliary event surfaces include environmental risk surface, feature counts, other crime risk surface, park event surface, and 311 event surfaces.

3.5.2.1. Environmental risk surface

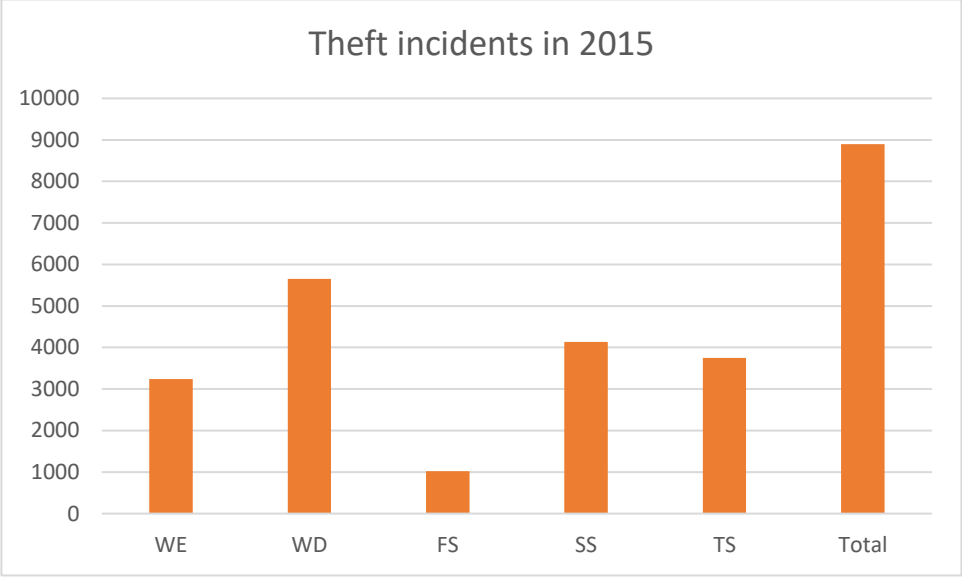
Environmental risk surface consists of crime specific SSRS values of segments in each time step. We obtained these values by quantifying the dynamic local effects of the selected urban features segments through a new risk indicator (LRS). LRS values are calculated by the value of the related DRSFs at 1.7 meters. We used a sample of 2015 robbery and theft incidents in the CS in fitting DRSFs. Different from the first part, we excluded the default model from the analysis since it does not have any predictive value to our model given dynamically changing environmental risk in time. The resulting crime distribution across intraday and weekly models is displayed in Figure 10. In the last step, we calculated crime specific SSRS of a street segment across temporal models by multiplying LRSS values of urban features residing in a street segment. Calculating SSRS is illustrated in Figure 11.

3.5.2.2. 311 calls

311 is a public reporting platform where citizens can file a request for non-emergency city services. Upon a request, the system dispatches responsible workforces to the place of request and provides continuous updates about its latest status until the request is completed. Previous research has shown that incorporating 311 calls improved the predictive accuracy of spatiotemporal crime prediction models (e.g., Duan et al., 2017). A typical City of Chicago 311 service request form includes information about the time of the request, address, service type, and GPS coordinates. Each service request has a timestamp. We obtained the street segment information by geocoding the GPS coordinates of the request to the CS street network. The timestamp and street segment information then help us create spatiotemporal datasets that include the daily and shiftly 311 requests for each street segment in the CS between 2016 and 2018.

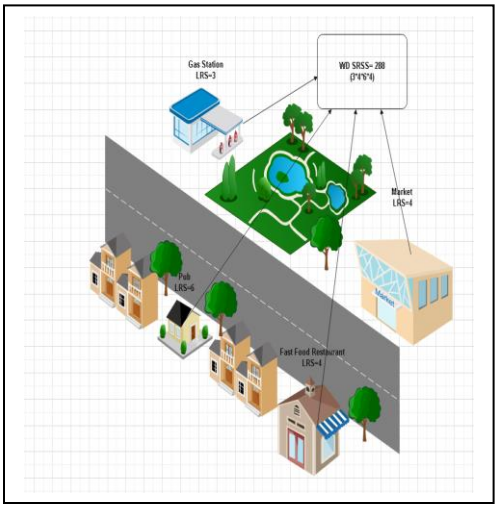


(a)

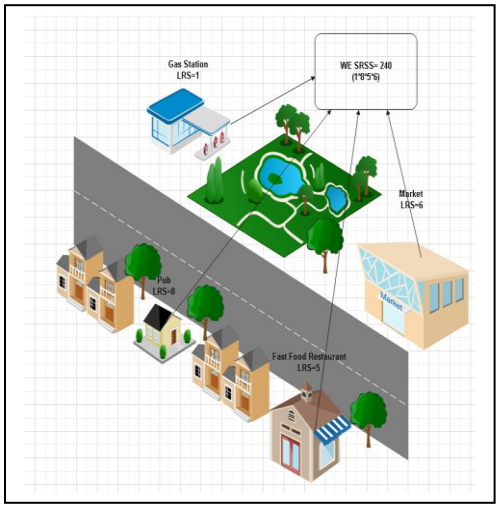


(b)

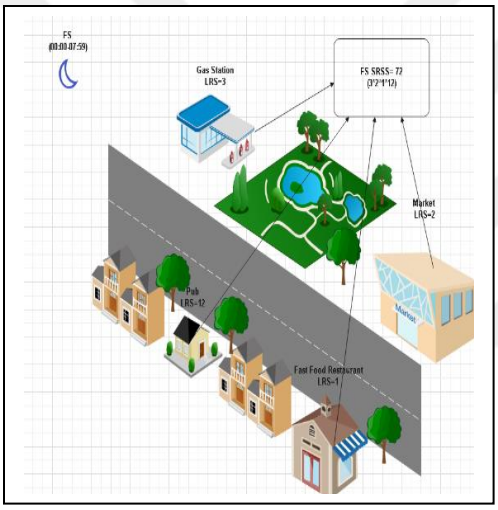
Figure 10. Temporal crime distribution a) Robbery, b) Theft



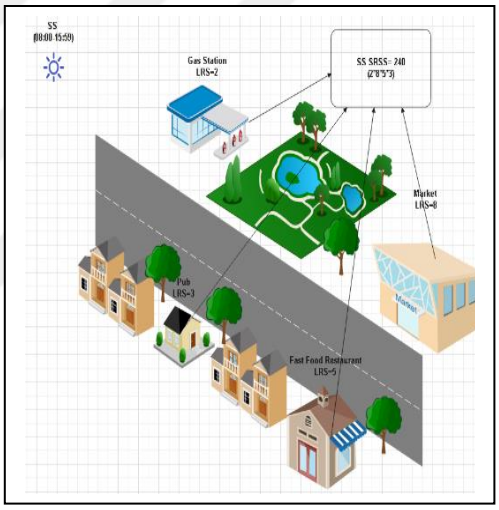
(a)



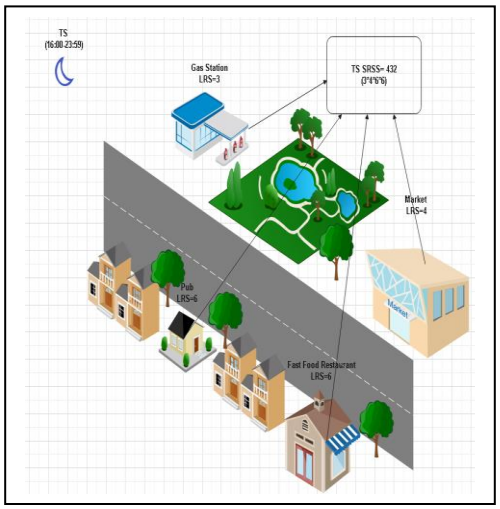
(b)



(c)



(d)



(e)

Figure 11. Calculating SSRS (a) WD (b) WE (c) FS (d) SS (e) TS

Based on a large volume of empirical research showcasing their significant impacts on crime occurrence (e.g., Wheeler, 2018; Chalfin, Kaplan, & LaForest, 2020), we selected five different types of 311 requests: Graffiti Removals, Potholes, Street Lights Out, Abandoned buildings, and Ordinance Violations. Next, we created separate transaction tables that tally the requests filed from each segment in each time step. Due to the extreme data sparsity in many tables, we decided to aggregate all the selected types in a single dataset, Total 311 dataset. In this dataset, each row represents a time step (i.e., day or shift), each column represents a street segment, and each field thus shows the total count of 311 requests in a segment at a time step. We aggregated this table at two temporal levels, creating two different 311 transaction tables to be used in predictive crime hotspot mapping methods.

3.5.2.3. Park events

Public events that draw large groups of people in limited areas act as temporary crime generators. These large groups escalate crime risk near parks for a short period. This is because a large crowd poses a fertile ground for motivated offenders who feed on the other people's distraction in these crowds (Cohen & Felson, 1979; Brantingham & Brantingham, 1981). Previous research has provided sufficient empirical evidence for these pop-up crime generators (e.g., Ristea, Kurland, Resch, Leitner, & Langford, 2018). We, therefore, decided to create an auxiliary event surface that represents the number of park events on street segments in a period. Similar to the 311 transaction tables, we created a park event transaction table. There is one major difference in the way we incorporate this external factor. Evidently, parks accommodate multiple segments. Meaning that in each time step, all the included segments should take up values that equal the number of ongoing events. This formulation is predicated upon an assumption

that the crime risk in a park event is positively related to the number of participants and events. Another important point is the duration of an event. If a park event spans the whole weekend, then the number of ongoing events in relevant segments should be represented by their intensity. They, therefore, should be smoothed into continuous values. The reason is that the elevated crime risk in areas near parks should be highlighted during the park events. In addition, this elevated crime risk tends to decay due to the decreasing number of attendants as the park events progress. Capturing this fluctuating risk may serve to predictive crime hotspot maps that by providing a leading indicator that informs predictions with the upcoming crime events. As a result, we created a park event surface where each row represents the smoothed number of events in park-related segments in a time step.

3.5.3. Representing the predictive hotspot mapping as a graph signal processing task

In this study, we followed the procedure proposed by Zhang & Cheng (2020) who reformulate the predictive crime hotspot mapping as a task of Graph Signal Processing (GSP). GSP is an umbrella term that encapsulates a set of related tasks for processing data on irregular graphs (Ortega, Frossard, Kovačević, Moura, & Vandergheynst, 2018). In that sense, we used this term to refer to the task of crime hotspot prediction on street networks. As illustrated in Figure 12, we achieved this reformulation in three main steps. Each step concerns a different component of GSP. First, we extracted a segment graph from the street network: G= (V, E, W)

Here, V (i.e., vertices), denotes the street segments, and the second E (i.e., edges) denotes the connection between two adjacent segments, and the last, W (i.e., weight), denotes the weight of this edge b. Next, signal (i.e., S) refers to the crime counts on

each segment in a period. Processing (i.e., P) addresses the task of spatiotemporal crime prediction.

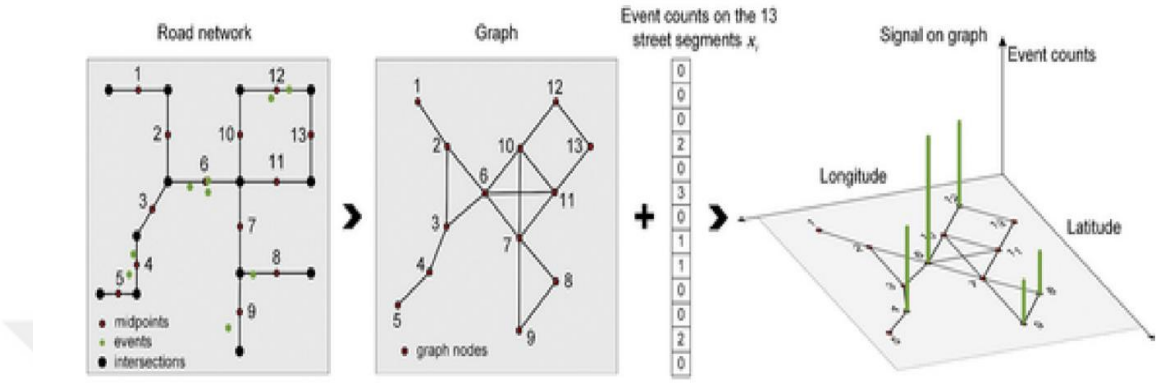


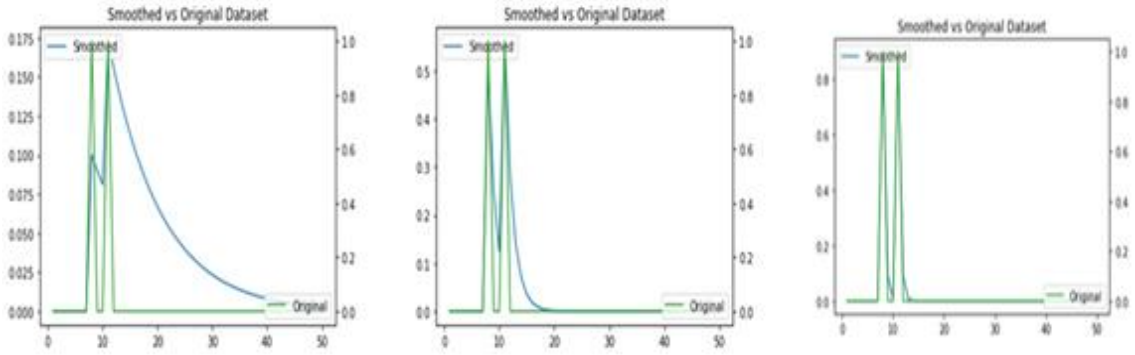
Figure 12. Reformulating crime prediction on a street network as a task of GSP (Zhang & Cheng, 2020)

Creating a graph requires a conversion of street segments into nodes on the suggested graph. In this graph, each street segment is represented by its midpoint. Edge stores the connection information between adjacent segments. At a time step, crime count represents signals. We modeled the spatial dependence between graph signals with a Gaussian kernel weighting function. Based on the first law of geography (Tobler, 1970) that assumes greater levels of relationships between near things than distinct ones, these weights are assigned inversely to street network distance between the neighboring segments. The following describes the Gaussian kernel weighting function that quantifies the weight of a particular edge like, e_{ij}.

$$w_{ij} = e^{\left(-\frac{[dist(i,j)]^2}{2\alpha^2}\right)} (5)$$

Here, w_{ij} represents the weight between the segments i and j, and $dist(i,j)$ is the network distance between two adjacent cells, α is the bandwidth.

After creating the graph, Zhang & Cheng (2020) continues with a Simple Exponential Smoothing (SES) operation that extends crime counts into periods instead of representing them as sudden spikes on a crime count- time plot. The explanation for this representation comes from the near-repeat phenomenon. This phenomenon posits that a crime occurrence elevates the crime risk in near areas for a short period before the risk totally dissipates (Pease, 1998). Here determining the smoothing coefficient is of particular importance because this value adjusts the temporal extent of elevated crime risk in an area: the extent increases as the coefficient decreases. Evidently, a smoothing operation transforms crime counts into what can be interpreted as crime risk values. Using a continuous crime risk, rather than crime count, as a graph signal serves better to the graph learning algorithms. Figure 13 illustrates the impact of the smoothing coefficient on the temporal extent of crime risk.



(a) (b) (c)

Figure 13. The impact of smoothing coefficient on crime count representation: (a) 0.1 (b) 0.5 (c) 0.9

These preprocessing steps enact the usage of a wide range of graph learning algorithms that produce network-based crime hotspot prediction. In this study, we have selected two popular graph learning algorithms, the Graph Wavenet(Wu et al., 2019) and

Spatiotemporal Graph Convolutional Network (i.e., STGCN) (Yu, Yin, & Zhu, 2018), to predict crime hotspots in the next time step.

3.5.2 Graph learning algorithms

3.5.2.1. Graph wavenet algorithm

This study adapts the Graph Wavenet algorithm (Wu et al., 2019) into the context of predictive crime hotspot mapping on street networks. This algorithm brings two main innovations to spatiotemporal graph modeling: self-adaptive adjacency matrix and stacked dilated 1D convolution component. The former can identify the hidden spatial dependencies by using node embeddings. This approach is plausible due to the circumstances where either two nodes are interdependent, but a connection does not exist, or a connection exists but there is not an interdependency between these nodes (Shleifer, McCreery, & Chitters, 2019). Such circumstances exist in the criminal context as well given the observed spatial lag between crime locations. More clearly, a crime in one street segment may trigger the crime in a second even a third-degree neighbor. This spatial lag can partly be explained by the intensity zones (Angel, 1968) that refer to the elevated crime risk in areas a few blocks away from a criminogenic place. The latter innovation helps capture long temporal dependencies between crime occurrences. Because 1D convolution layers with receptive fields that grow exponentially in parallel with increasing layers overcome the problem of learning sparse sequences in recurrent neural networks. The ability to capture such dependencies is crucial given the sparsity of a typical crime datasets that consist of mostly zero values.

Graph Wavenet operates on a graph that can be formulated as:

$$G = (V, E, W) (6)$$

is a graph where V denotes its nodes (i.e. street segments) , E signifies its edges, and W is the set of weights. An adjacency matrix derived from G is A ϵ R^{NXN}. N is the number of vertices. If $e_{i,j}$ exists, then $A_{i,j}$ is equal to $W_{i,j}$. The dynamic feature matrix (i.e., Graph Signals) of G is represented by $X^{(t)}$ϵ R^{NXD}, D is the number of input features. Given a graph, G, and time step S, the problem is to forecast the graph signals for each segment in the next time step by learning a function, *f*.

$$\left[X^{(t-S):t}, G\right] \xrightarrow{f} X^{(t+1)} \quad (7)$$

Here $X^{(t-S):t}$ϵ R^{NxDxS}, and $X^{(t+1)}$ϵ R^{NxM}. M is the output dimension

Graph Wavenet algorithm is built upon two main components: Graph Convolution (GCN)layer and Temporal Convolution (TCN) layer. The former models spatial dependence whereas the latter handles temporal dependence. The next two sections will detail these components.

The main function of GCN is to extract node features residing on a graph structure. This study used Chebyshev Spectral Filtering (Defferrard, Bresson, & Vandergheynst, 2016) to smooth a node's signal by aggregating and transforming its neighbors' information based on spatial dependence. This technique supports multidimensional inputs. Kipf & Welling (2016) formulated GCN as follows:

$$Z = \tilde{A}XW \quad (8)$$

Where X^{NXD} represents the input signal, Z^{NXM} is the output, $\tilde{A}$ ϵ R^{NXN} is the normalized adjacency matrix. Based on the diffusion convolution layer (Li, Yu, Shahabi, & Liu, 2017) that models the diffusion of graph signals with K finite steps, the layer can be reformulated as:

$$Z = \sum_{k=0}^{K} \widetilde{P^k} XW_k \text{ (9)}$$

P^k here is the power series of the transition matrix that is equal to P =A/rowsum(A)

The algorithm proposes a self-adaptive adjacency matrix ($\tilde{A}_{adp}$) that helps discover hidden spatial dependency in the graph. To do this, two node-embedding dictionaries with learnable parameters, $E_1, E_2 \epsilon\ R^{NXc}$, were initialized:

$$\tilde{A}_{adp} = \text{SoftMax (ReLU}(E_1 E_2^T) \text{ (10)}$$

E_1 is the source node embedding, and E_2 is the target node embedding. Multiplying them gives us the degree of their spatial dependence. The ReLU here eliminates weak connections, and SoftMax normalizes the remaining weights. The predefined spatial dependencies and self-learned hidden graph dependencies then create the ultimate GCN:

$$Z = \sum_{k=0}^{K} \widetilde{\tilde{A}_{adp}^{k=0}} XW_k \text{ (11)}$$

This model can be interpreted as an aggregation of transformed feature information from different orders, K.

Temporal dependency in nodes' features is captured by dilated causal convolution networks (Yu & Koltun, 2016). These networks overcome the gradient explosion problems by handling long-range sequences in a non-recursive manner through their exponentially growing receptive fields. The dilated causal convolution keeps causal order by means of padding zeros to the inputs to restrict the predictions to historical information. Dilated convolutions iterate over inputs by skipping some values with a certain step. A dilated convolution can be formulated as follows:

$$x * f(t) = \sum_{s=0}^{K-1} f(s)x(t - dxs) \text{ (12)}$$

x ϵ R^T is 1D sequence input, f ϵ R^k is a filter. d here is the dilation factor that determines the skipping distance. Stacking these dilated causal convolutions with the dilation factor is what exponentially expands the receptive field.

In Recurrent Neural Networks, gating mechanism is an important control structure that determines how much of historical information is forwarded to the next time steps. Dauphin, Fan, Auli, & Grangier (2016) have showcased its applicability to temporal convolution networks. Given an input of X ϵ R^{NxDxS}, the gate is:

$$h = g(\Theta_1 * \chi + b) \odot \sigma(\Theta_2 * \chi + c) \text{ (13)}$$

Θ_1, Θ_2, b, and c are all model parameters. $\odot$ is element-wise product, g(.) is the output activation function, σ(.) is the sigmoid function that determines the ratio of information to the next layer. This gated mechanism serves to model temporal dependencies.

Graph Wavenet consists of stacked spatiotemporal layers and an output layer. A spatiotemporal layer sits atop two main structures described above: GCN and Gated TCN. The latter consists of two separate temporal layers (TCN-a and TCN-b). Stacking spatiotemporal layers help process spatial data at various temporal levels. The former receives short-term information at a lower level whereas the top layers deal with longer sequences. Inputs to a GCN is a three-dimensional tensor, [N, C, L], N is the number of nodes, C is the hidden dimension, and L is the sequence length.

Different from a regular Mean Absolute Error (MAE) used in the original version of Graph Wavenet (Wu, Pan, Long, Jiang, & Zhang, 2019), we used a squared quantile loss

function that intentionally explodes the cost of missing a nonzero observation (i.e., crime occurrence). This weighted loss function is given below:

$$L(\hat{y}_{i-}, \Theta) = \frac{1}{ND} \sum_{j=1}^{j=N} \sum_{k=1}^{k=D} w_i * (\hat{y}_{i-} y_i)^2 \quad (14)$$

Here, $\hat{y}_i$ and $y_i$are predicted and actual values respectively. w_i is an asymmetric weight changing based on the value of y_i. Inspired from quantile loss functions, we formulated this weight as:

$$w_i = \begin{cases} q \; if \; y_i > 0 \\ (q-1) if \; y_i = 0 \end{cases} (15)$$

Here q ϵ [0,1] is the quantile. For larger values, it severely penalizes underpredictions which corresponds to missing a crime event. This function helps the selected algorithms focus on predicting rare crime events by sharpening the contrast between the costs of missing zero and nonzero observations.

Figure 14 illustrates Graph Wavenet framework.

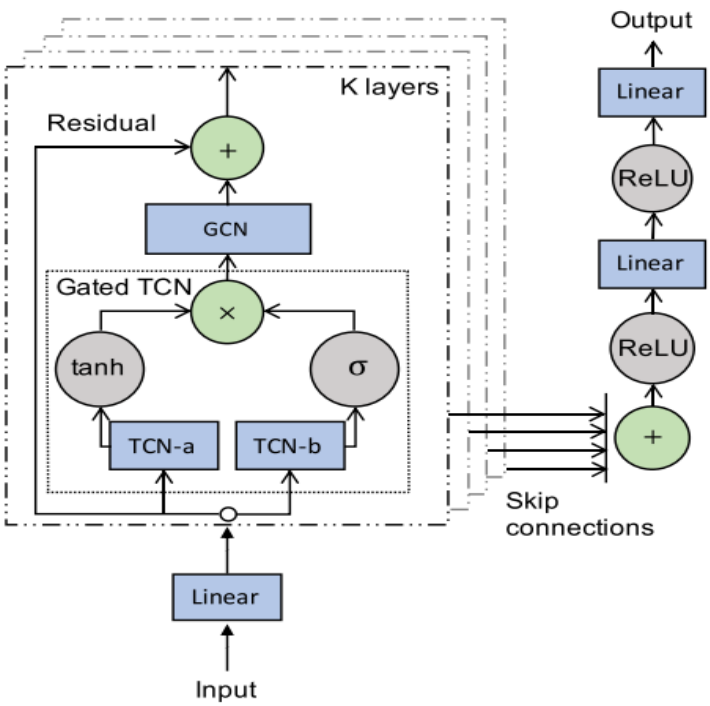


Figure 14. Graph Wavenet framework (Source: Shleiffer et al.,2019)

3.5.2.2. Spatiotemporal graph convolutional neural network (STGCN)

STGCN is another DL algorithm that models spatiotemporal dependence using only convolutional units rather than recurrent units (Yu, Yin, & Zhu, 2018). The suggested design is made up of spatiotemporal convolutional blocks, each of which contains one graph convolution layer and two gated CNN layers (Defferrard, Bresson, & Vandergheynst, 2016). These blocks are capable of extracting spatial connections while capturing temporal dependencies. Using entirely convolutional layers also enables for faster training durations and fewer parameters.

From graph-structured data, the graph convolution layer extracts spatial dependency. In graph convolution layers, there are two techniques to kernel approximation, which modifies the depth of the spatial dependence.

The first is the approximation of Chebyshev polynomials, which can be written as follows:

$$\Theta(\Lambda) = \sum_{k=0}^{K-1} \theta_k \Lambda^k (16)$$

K is the kernel size that adjusts the order of the spatial dependence, θ ϵ R^k represents polynomial coefficients. The graph convolution is formulated as:

$$\Theta * \zeta\chi = \Theta(\mathrm{L})\chi^{\exists} \approx \sum_{k=0}^{K-1} \theta_k T_k(\widetilde{L})\, \chi (17)$$

$T_k(\widetilde{L})$ ϵ R^{nxn} is the Chebyshev polynomial order of k.

The second is the 1st order approximation, which uses the first order approximation of graph laplacian to stack numerous localised graph convolutional layers. It may design deeper network structures without explicit parameterization defined by the polynomials. It can be formulated as:

$$\Theta * \zeta\chi = \theta(I_n + D^{-\frac{1}{2}} W D^{-\frac{1}{2}})\, \chi\ (18)$$

Here W ϵ R^{nxn} represents the weighted adjacency matrix, D ϵ R^{nxn} is the diagonal degree matrix, and θ is the kernel parameter.

A generalized graph convolutional layer can then be generalized as:

$$y_j = \sum_{i=1}^{C_i} \Theta_{i,j}(L)\, \chi_i \epsilon R^n, 1 \le j \le C_O\ (19)$$

C_i x C_O vectors of Chebyshev coefficients $\Theta_{i,j} \epsilon R^k$. The input for 2D can be denoted as $\Theta * \zeta\chi$. But it can be extended to 3D since each input consists of M frame that can be represented by the values of an external variable at each node. That means χ can be reformulated as $\chi \epsilon r^{MxnxC_i}$.

Gated CNN layer involves a 1-D causal convolution of width K, kernel complemented with Gated Linear Unit (GLU) as a non-linearity. The input to temporal convolution for each node is M-length sequence with C_i channels as $Y \epsilon R^{MxC_i}$. The convolution kernel $\Gamma \epsilon R^{K_t x C_i x 2C_0}$ maps the multichannel input into a single output. The temporal gated convolution can be described as:

$$\Gamma * TY = P \odot \sigma(Q) \epsilon R^{(M-K_t+1)xC_O} (20)$$

P and Q are input gates of GLU, $\odot$ signifies the Hadamard product. σ gate determines which P input serves to discover compositional structures and dynamic variances in time series. Similar to graph convolution layer, it can be extended to 3D convolutions.

Joining both structures is the Spatiotemporal Convolutional block that process the graph-structured time-series data. The sandwich structure (i.e., gated CNN layers on both ends and a bridging graph convolution layer in the middle) allows to apply bottleneck strategy. This strategy enacts scale compression and feature squeezing by increasing or decreasing the size of the channels. The blocks accept and produce 3D tensors. An output, υ^{l+1}, for an input υ^l is formulated as:

$$\upsilon^{l+1} = \Gamma \frac{l}{1} * TReLU\left(\Theta^l * \zeta\left(\Gamma_0^l * T\upsilon^l\right)\right) (21)$$

Here $\Gamma \frac{l}{1}$ and Γ_0^l are upper and lower temporal kernel respectively, Θ^l is the spectral kernel of graph convolution, ReLU is the rectified linear unit. Lastly the predictions are:

$$\hat{\upsilon} = Zw + b (22)$$

Where Z ϵR^{nxc} is the final output from the model, w is the weight, and b is the bias. The following illustrates the STGCN framework.

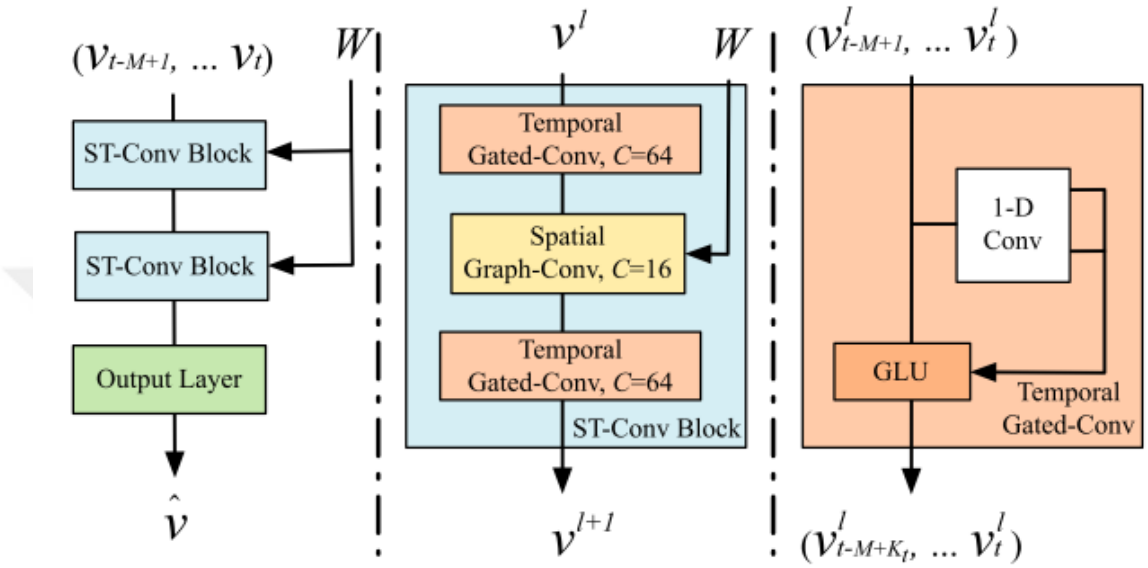


Figure 15. STGCN framework (Yu, Yin, & Zhu, 2018)

3.5.3. Baseline

In this study, we used LSTM to highlight the performance of the proposed graph learning algorithms. LSTM is a conventional DL technique that was widely used in crime prediction domain (e.g., Zhang, Liu, Xiao, & & Ji, 2020). Long Short-Term Memory (Hochreiter & Schmidhuber, 1997), or shortly LSTM, is a Recurrent Neural Network (RNN) variant that is developed to overcome the notorious problem of long-term dependencies. This problem refers to the RNN's growing inefficiency to learn temporal dependencies between the increasing number of data points. Past research has offered LSTM-based techniques to bring solutions to a multitude of real-world problems such as speech recognition (e.g., Graves, Jaitly, & Mohamed, 2013), time series

forecasting (e.g., Sagheer & Kotb, 2019), trajectory prediction (e.g., Alahi et al., 2016), and video classification (e.g., Ogawa, Sasaka, Maeda, & Haseyama, 2018).

Unlike standard RNNs, which use a primitive repeating module, LSTM uses a sophisticated repeating module with four primary components that interact with one another. Figure 16 depicts the internal construction of an LSTM module.

The cell state, ct, is the key distinction between an LSTM and a regular RNN. This factor essentially controls how much of information from the past will be carried forward into the future. This is accomplished through the use of gates, which determine the quantity of data to be kept/passed.

The first gate in LSTM is the forget gate. This is a sigmoid layer that generates a number between 0 and 1 by looking at the input (i.e., x_t) and hidden state (i.e., h_{t-1}). Zero keeps the state as is, and 1 completely changes it. Next, LSTM decides what information should be stored in c_t in two main steps. In the first step, input gate layer (i_t) selects the values to be updated through a sigmoid layer. In the second step, a vector with candidate values ($\hat{c}_t$)is created by a tanh function. Then, LSTM combines these two to update the state. To actually do that, LSTM multiplies the old cell state, c_{t-1}, with forget gate f_t adds this to the multiplication of candidate state and input gate ($\hat{c}_t * i_t$). In the last section, LSTM determines which parts of the c_t through another sigmoid layer and passes the resulting cell state to a tanh function, and multiply this state with output of the previous sigmoid layer. The following equations outline the LSTM working principle:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f \ (23)$$

$$i_t = \sigma(W_i * \ [h_{t-1}, x_t] + b_i \ (24)$$

$$\hat{c}_t = \tanh \ (W_c * [h_{t-1}, x_t] + b_c \ (25)$$

$$C_t = f_t * C_{t-1} + i_t * \hat{c}_t \ (26)$$

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o (27)$$

$$h_t = o_t * \tanh(C_t) \ (28)$$

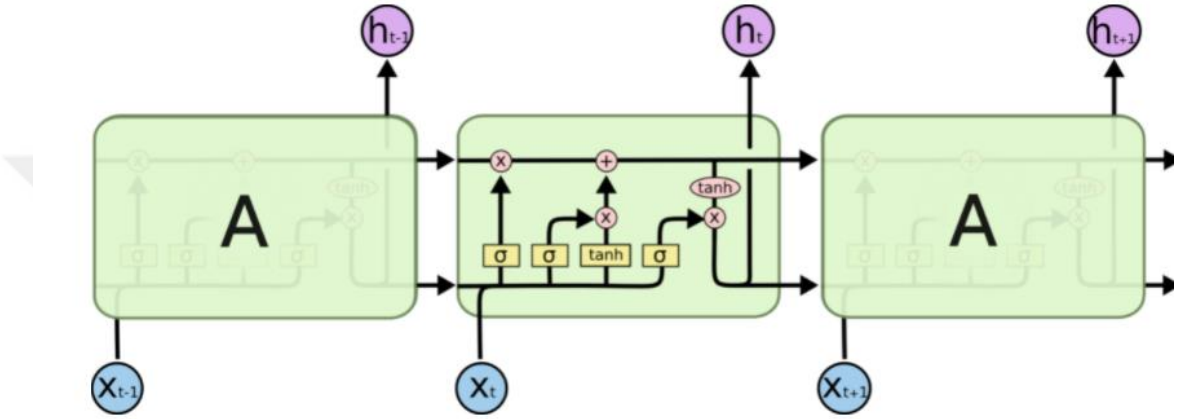


Figure 16. Internal structure of LSTM (Olah, 2015)

3.5.3 Performance measurement

We evaluated the performance of the selected algorithms with Mean Hit Rate (MHR) (Zhang & Cheng,2020). This measure calculates the mean percentage of crime incidents occurred in the predicted hotspots. At each time step, the segments are ranked by their predicted risk values in descending order. Next, we select segments from the top in a way that their total length does not exceed the specified coverage area. In this study, we used a coverage area of 20%. Meaning that we selected segments from the top in each time step whose total length will cover at most 20% of the total segment length. For example, an MHR of 0.7 indicates that 70% of crime incidents on average occurred in the predicted hotspots during the test days.

3.5.4 Experimental setup

In this study, we train our crime hotspot prediction models with a dataset that includes all the robbery (n=1555) and theft (n=22596) incidents reported on the CS between 1st January 2016 and 1st January 2018. We split the dataset into 70% training, 10% validation, and 20% test. For fine-tuning, we adopted a two-step strategy that performed a grid search on the parameters related to the data representation and the model. In the first step, we experimented with different train windows, smoothing coefficients, and spatial bandwidth values. In the second step, we experimented with 10 different learning rates that are considered as the most crucial parameters amongst others (Bengio,2012). We performed both steps on each selected crime type (i.e., robbery, theft) at each temporal level. Table 5 illustrates the experimental setup.

Table 5. Experimental Setup

	Daily	Shift
Data representation		
Train Window	7,10,14,30,42	3,21,30,42
Smoothing coefficients	0.05, 0.1, 0.5, 0.9	
Spatial Bandwidth	0.1, 0.2, 0.3	
Model		
Learning rate	0.001,0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.09,0.1	

In the representation parameter tuning, we first created separate datasets for each pair of train window size and smoothing coefficient in daily and shift models. Next, we run our models with a default learning rate (e.g., 0.01) for each dataset with three separate adjacency matrices each weighted by different spatial bandwidths. This strategy

produces a total of 60 daily models (i.e., 5 train window size x 4 smoothing coefficient x 3 spatial bandwidth), and 48 shift models (i.e., 4 train window size x 4 smoothing coefficient x 3 spatial bandwidth). We then calculate MHR for each model and select the parameters from the model with the highest MHR as the optimal data representation. In the second step, we create a single dataset with optimal parameters for data representation. We then perform model-related parameter tuning with 11 different learning rates that are displayed in Table 5. Lastly, the learning rate of the highest achieving models is selected as the best model.

Table 6. The Chicago Crime Dataset

			Time Steps		Crime Counts	
	#Nodes	*# Edges*	*Daily*	*Shift*	*Robbery*	*Theft*
Center Side	1680	2459	731	2192	1555	22596

CHAPTER 4

RESULTS

This chapter presents the results of the current study. We begin this chapter with the results of a descriptive analysis that provides an overview of crime in Chicago. Next, we report the results of DRSFs, and following spatiotemporal influence analysis. We conclude this chapter with the comparative analysis of the results of the developed predictive crime hotspot mapping models.

4.1. Descriptive analysis

This section summarizes the results of a descriptive analysis using the Chicago crime dataset that includes all incidents between 2015 and 2019(excluded). In the first step, we performed a frequency analysis that shows yearly, shift, and weekly total crime distribution across Chicago and its sides. In the second step, we did the same analysis for the selected crime types to establish an initial understanding of how the robbery and theft displace across the streets of Chicago. Lastly, we performed a micro-level analysis of street segments that illustrates the degree of crime concentration across Chicago and its sides.

4.1.1. Total crime distribution

Crime in Chicago has shown a slightly decreasing trend after it hits the top in 2016. One interesting result is that the total crime count does not exhibit much variation during the

analysis period. Figure 17 illustrates the yearly total crime counts in Chicago.

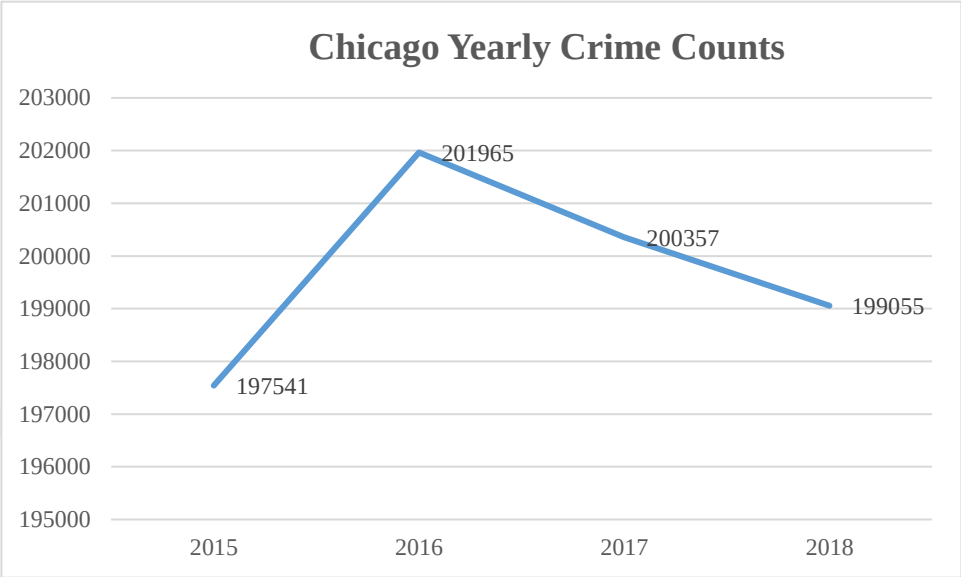
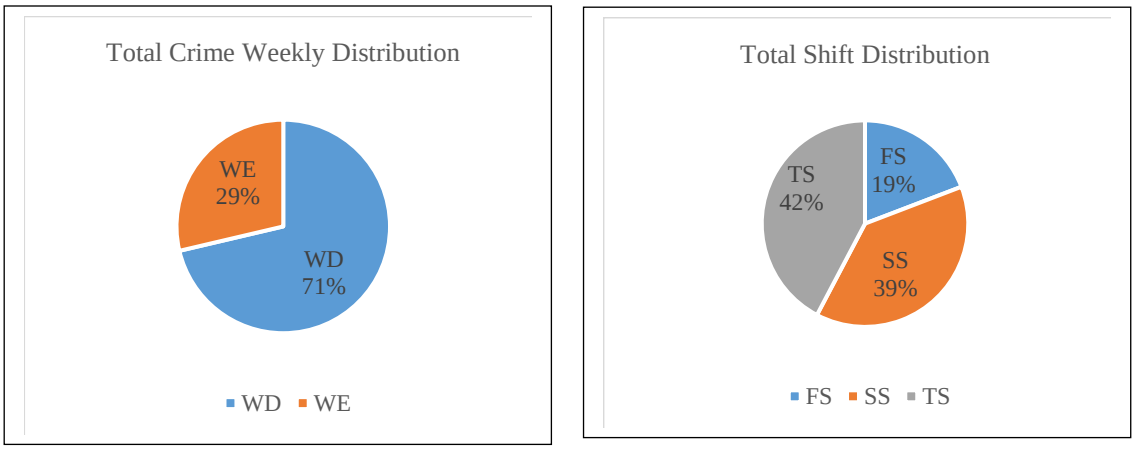


Figure 17. Total crime incidents in Chicago between 2015 and 2019

At the weekly level, Figure 18a shows that crime events concentrate in weekdays between 2015 and 2018. On the other hand, Figure 18b indicated an intraday crime concentration during the TS (i.e., 16:00-23:59). We also observed a similar level of crime concentration during SS (i.e., 08:00-15:59).

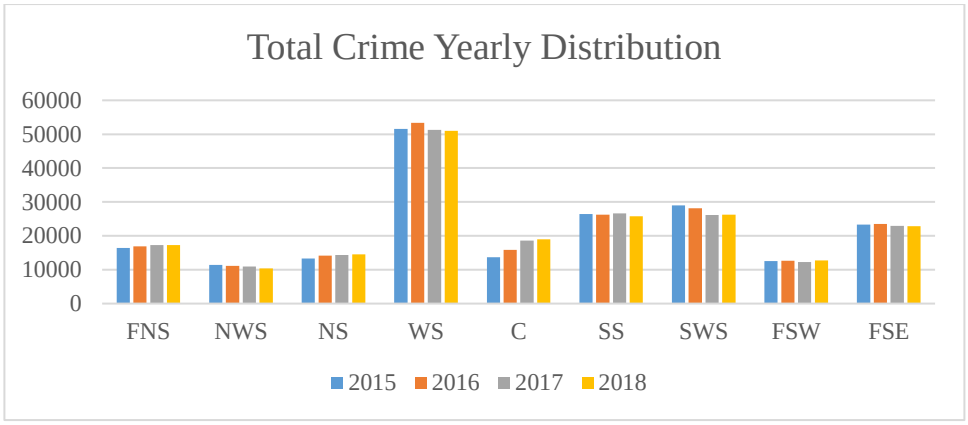


(a) (b)

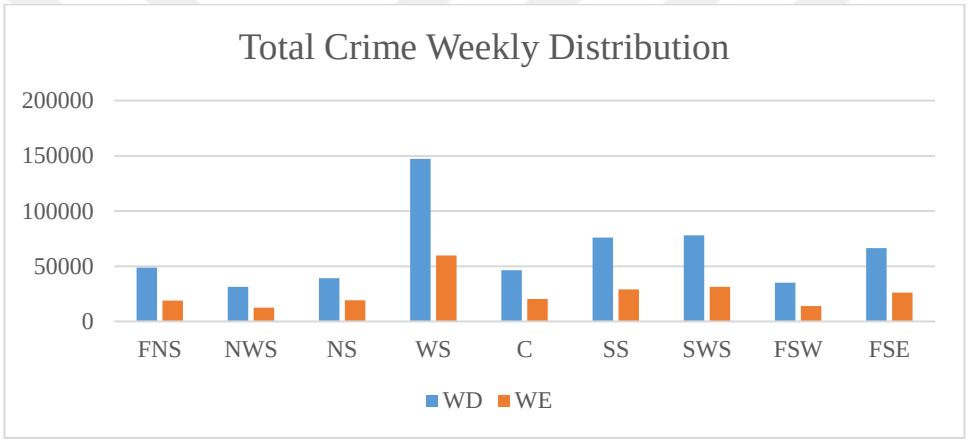
Figure 18. Crime distribution across (a) a week (b) day

Figure 19 revealed a much higher crime concentration on WS with an average crime count of 51796.5. We observed a consistently increasing trend for the CS during the analysis period. Likewise, the numbers indicated a 38.8% increase in crime counts between 2015 and 2018.

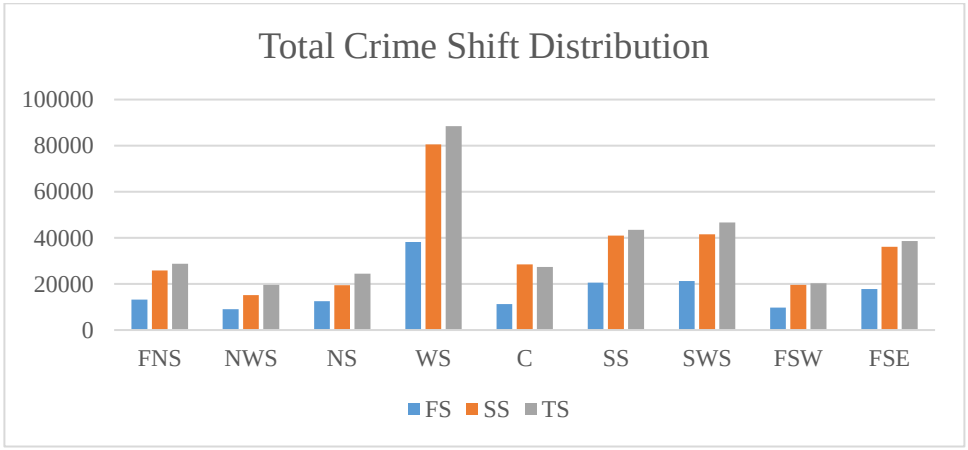
Similar to the citywide level, the crime concentration occurs during the weekdays. Although we calculate similar WE/WD ratios near 0.4 across the sides, this ratio is 0.49 in NS, indicating a more balanced crime distribution. The crime concentration occurs in the TS in all sides of Chicago. The distinguishing characteristic between the intraday crime distributions of the sides is the degree of increase between the shifts. The most dramatic difference occurred in WS between the FS and the others. We also observed more similar patterns in SS and SWS.



(a)



(b)



(c)

Figure 19. Total crime counts across the sides of Chicago (a) between the years 2015 and 2019 (b) across week (c) day

4.1.2. Selected crime counts

The robbery and theft incidents demonstrate different trends in the analysis period. While the former showed a slight bell-shaped trajectory that reaches its climax at 2016, the latter showed an ever-increasing trend that culminates in a 12% increase in incidents between 2015 and 2018. Figure 20 illustrates these trends.

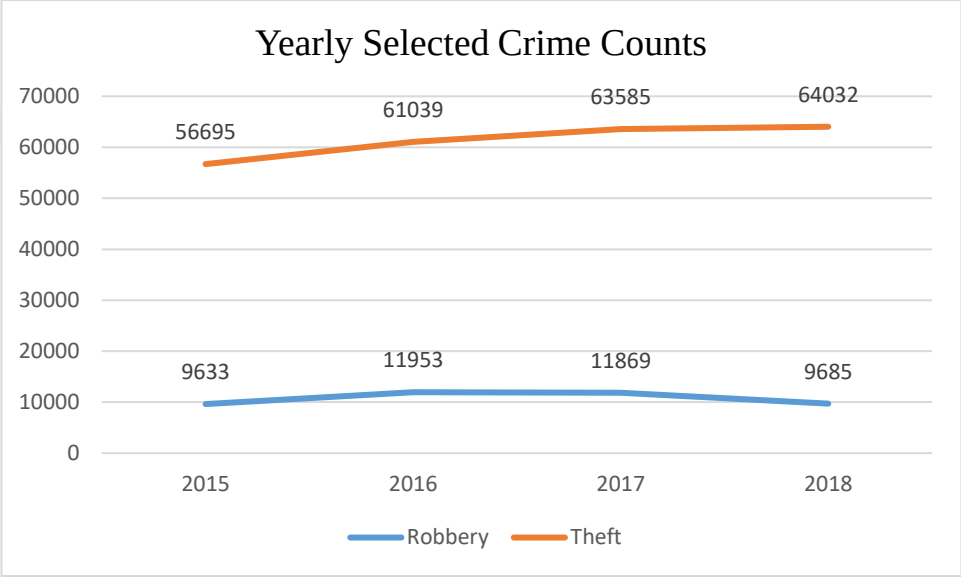


Figure 20. Yearly trends of the selected crime types

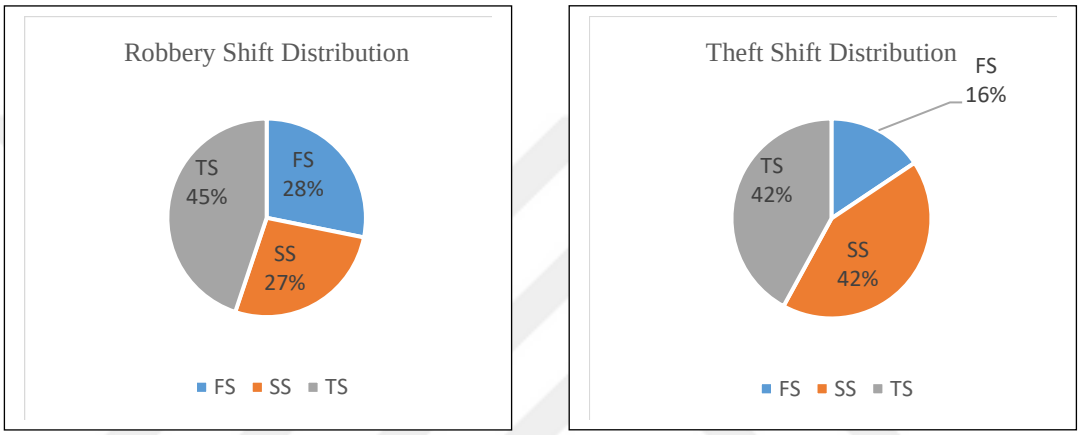
Figure 21 did not display a notable difference between robbery and theft incidents in terms of weekly distributions. The crime concentrations are similarly on weekdays.



(a) (b)

Figure 21. Weekly crime distribution (a) robbery (b) theft

The shift distribution in Figure 22 for the selected crime types showed an important difference in the crime counts that occurred during the FS, and SS. Although both types have a concentration in the TS, the degree of crime concentration is much balanced between FS and SS in robbery than it is in theft.

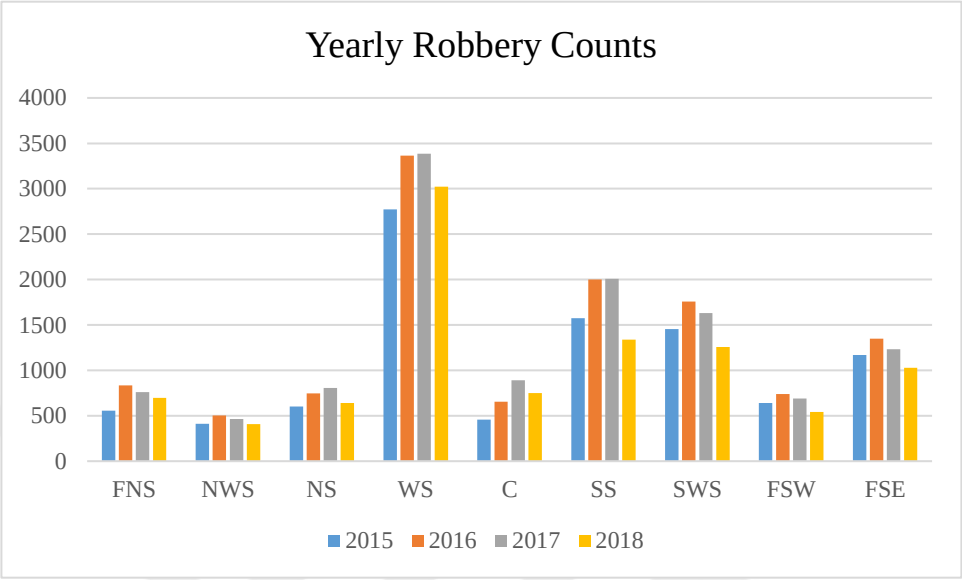


(a) (b)

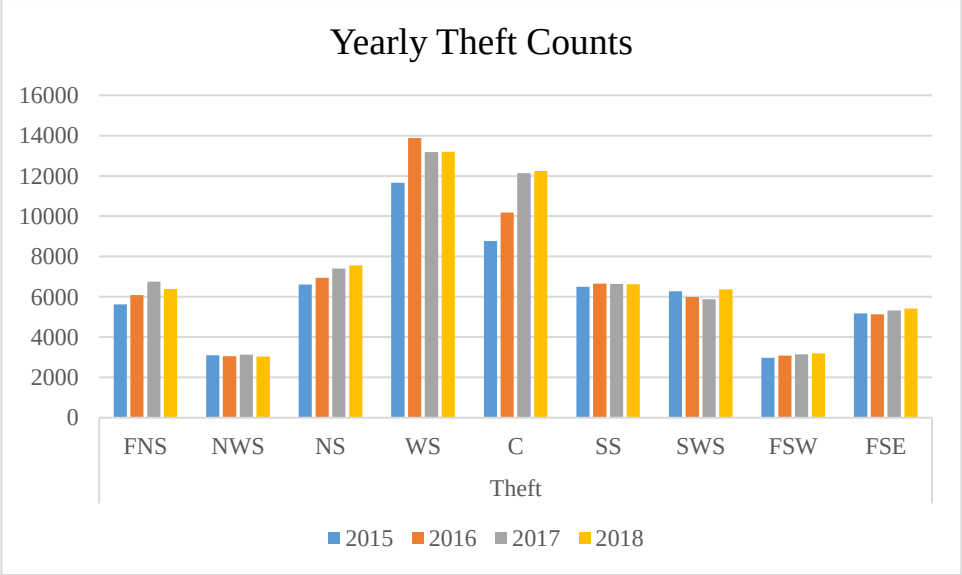
Figure 22. Shift crime distribution (a) robbery (b) theft

An interesting pattern in Figure 23 is the evident increase in the years in between the study period. WS and SS stood as the sides that are most exposed to robbery in that period. We observed a slightly increasing theft trend in nearly all sides of Chicago. In

theft incidents, CS replaced SS as the side having the second most crime concentration.



(a)

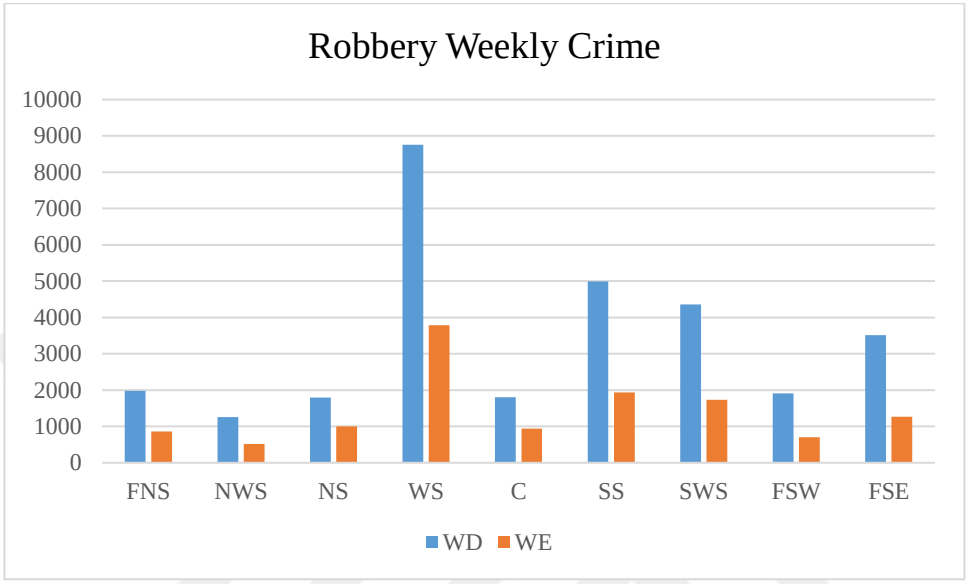


(b)

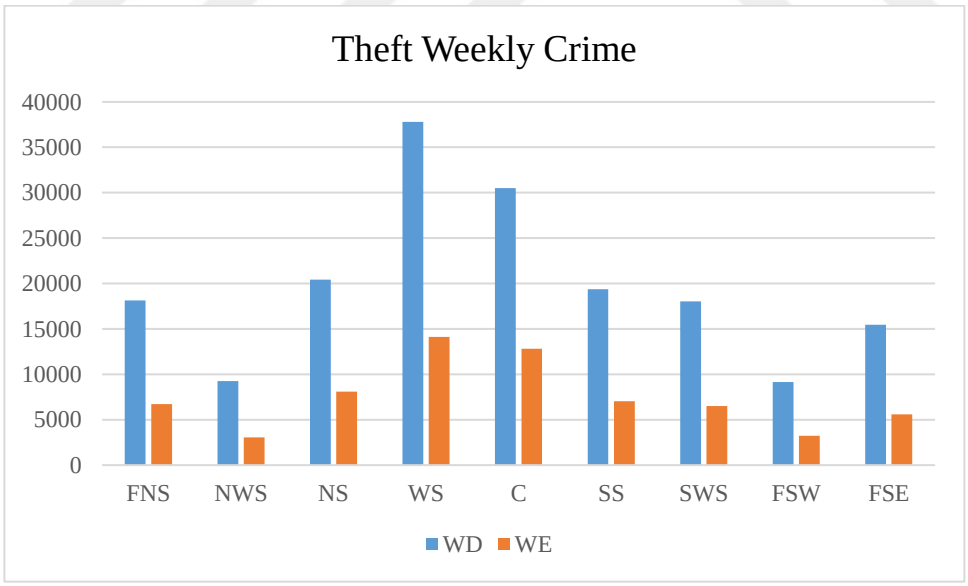
Figure 23. Yearly (a) robbery (b) theft counts

Figure 24 suggested that NS and CS, amongst others, stood out as the sides having more homogenous weekly robbery concentrations with the WE/WD ratios of 0.55 and 0.52

respectively. For theft, the CS is ranked at the top with a value of 0.42. The values for the other sides lie in a narrow interval between 0.33 and 0.39.



(a)

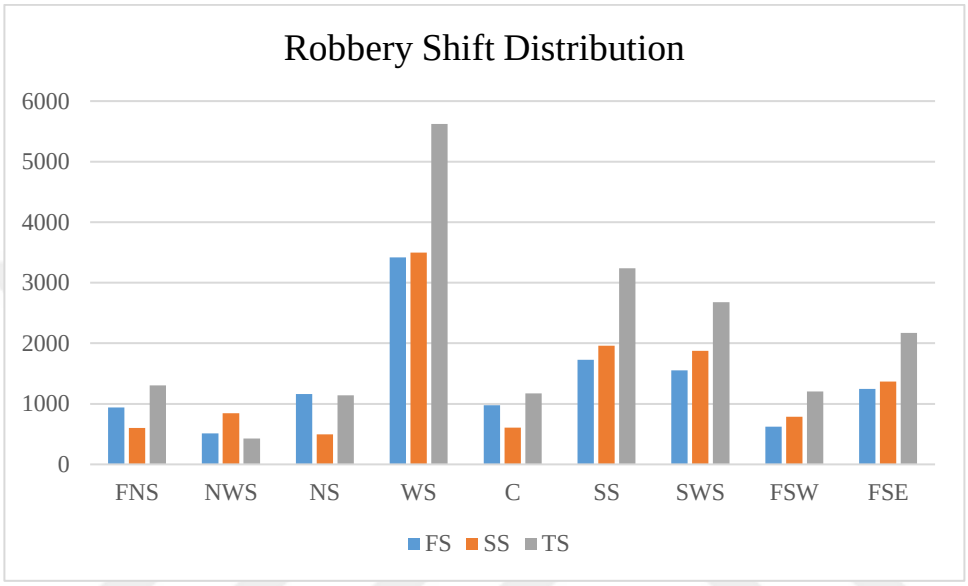


(b)

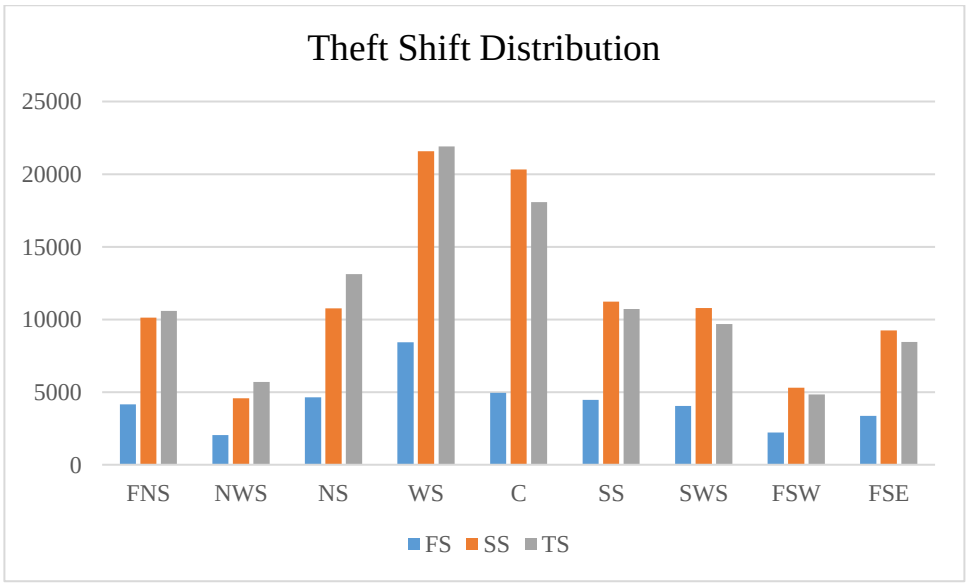
Figure 24. Weekly (a) robbery (b) theft distribution across the sides

Figure 25 illustrates a more balanced robbery distribution between the FS and SS. For all the sides, TS is the period having the largest robbery concentration. For theft incidents,

an interesting pattern occurred in the form of a switch between SS and TS. While the concentration is consistently higher in the TS on the northern sides, it shifts towards the SS on the southern sides.



(a)



(b)

Figure 25**.** Shift (a) robbery (b) theft distribution across the sides of Chicago

4.1.3. Characterizing crime concentration at street segment level

We characterized the crime concentration across and its sides through two measures: percentage of crime-free segments and percentage of segments accommodating 50% of the crime incidents. These measures serve to highlight the degree of heterogeneity in the spatial crime distribution and were thus widely used in the past research (e.g., Weisburd, 2015; Steenbeek & Weisburd, 2016; Levin, Rosenfeld, & Deckard, 2017). Figure 26 exemplifies a sample crime distribution across street networks.



Figure 26. A sample crime distribution across the street networks

We calculated these measures for each year in our analysis period. As a result, we illustrate street segments crime variability in the citywide and side-level crime concentration over the years. Figure 27 shows the number of street segments across the sides.

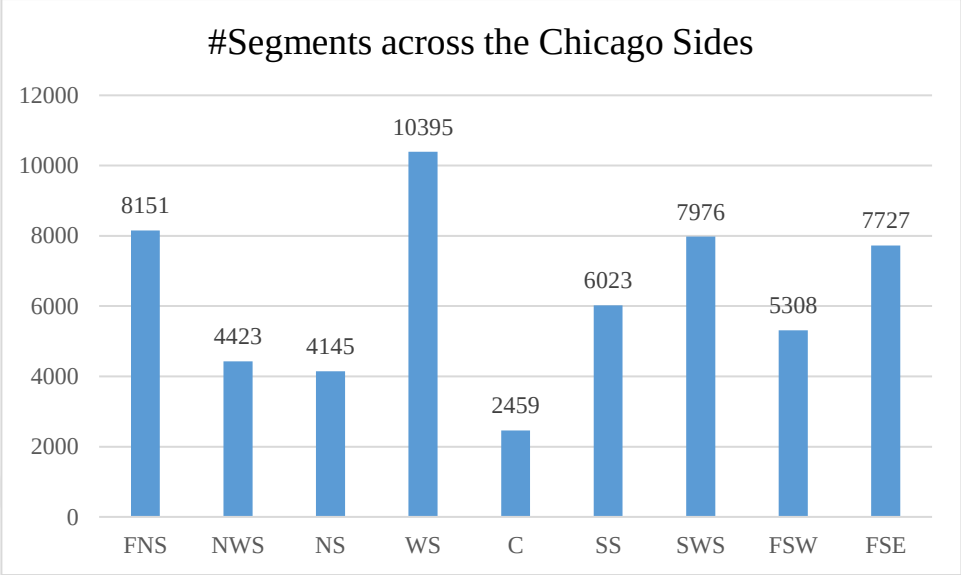


Figure 27. Number of street segments across the sides of Chicago

For the first measure, larger values indicate greater crime concentration. For instance, a value of 50 % means that one of each two units is crime-free. Accordingly, we observed the greatest crime concentration on the FNS with an average value of 52.55% over the four years period. NS, on the other hand, was found to have the most homogenous crime distribution across the street segments with an average of 31.38%. One interesting result here is the homogeneous crime distribution in the CS, and the concentrating crime distributions in the SWS and the FSW.

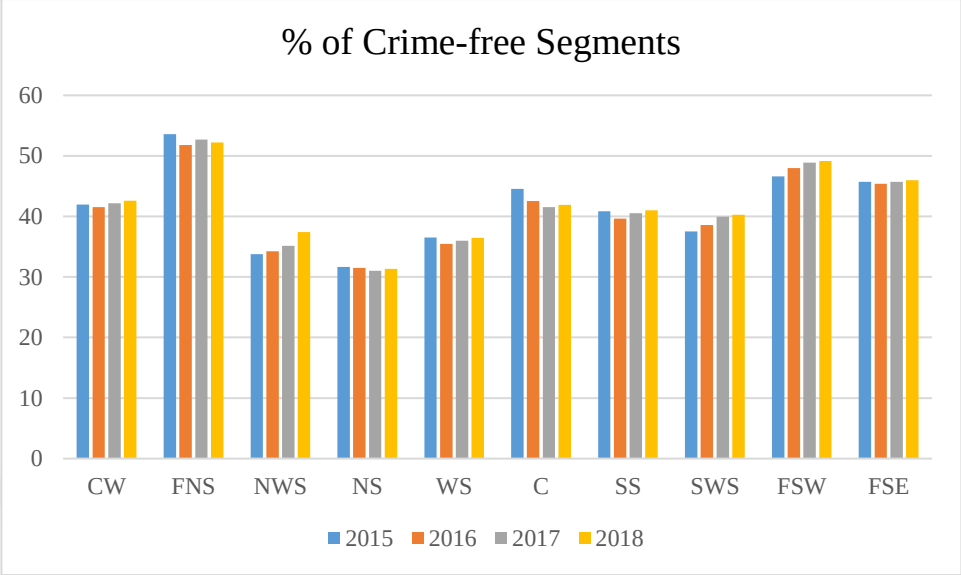


Figure 28. % of crime-free segments across the sides of Chicago between 2015 and 2019

Unlike the first measure, the smaller values of the second measure suggest greater crime concentration in areas. For the second measure, we found the CS to be the side having the least percentage of segments (4.92%) accounting for half of the crime incidents. By contrast, NWS is found to have the highest average value with an average of 12.48%. An interesting pattern that we observed was that the values of the CS in the second measure showed a decrease during the 2015-2019 period. Along with the first measure, this paints an interesting picture where crime incidents appear to be mitigated across the street network, and the number of segments accounting for half of the incidents decreases. One possible interpretation would be an increasing volume of crime incidents in the top crime hotspots along with a crime diffusion to crime-free segments from other crime hotspots below.

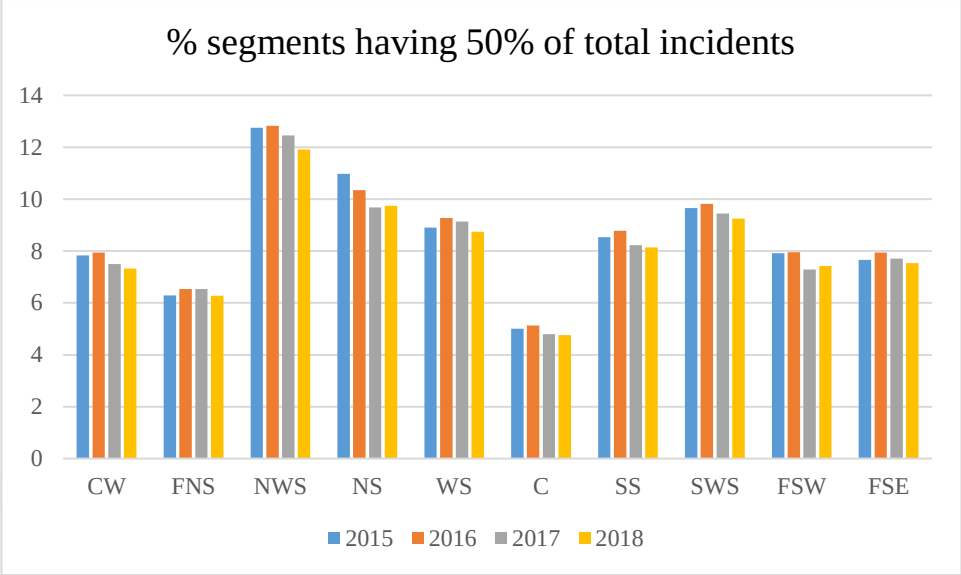


Figure 29. Yearly segments % having 50% of total incidents

4.2. DRSFs

This section summarizes the results of the DRSFs. First, we report the results for the Network K analysis that shows insignificant spatiotemporal crime clusters around the selected urban features. Then we continue our analysis with a spatiotemporal influence analysis that compares the calculated crime-specific RSIS and RSSS values across spatiotemporal models. In the last section, we conclude this part with the results of the correlation analysis that examines the relationships between these indicators and Concentrated Disadvantage (CD).

4.2.1. Bivariate network K analysis

The results (see Appendix A) indicated significant theft and robbery clusters around selected urban features across 509 out of 540 spatiotemporal models. The insignificant crime clusters highlight unstable spatial influences of gas stations, grocery stores, and pubs changing across space and time. Figure 30 illustrates an example of the unstable

spatial influence of pubs on the robbery in FSWS.

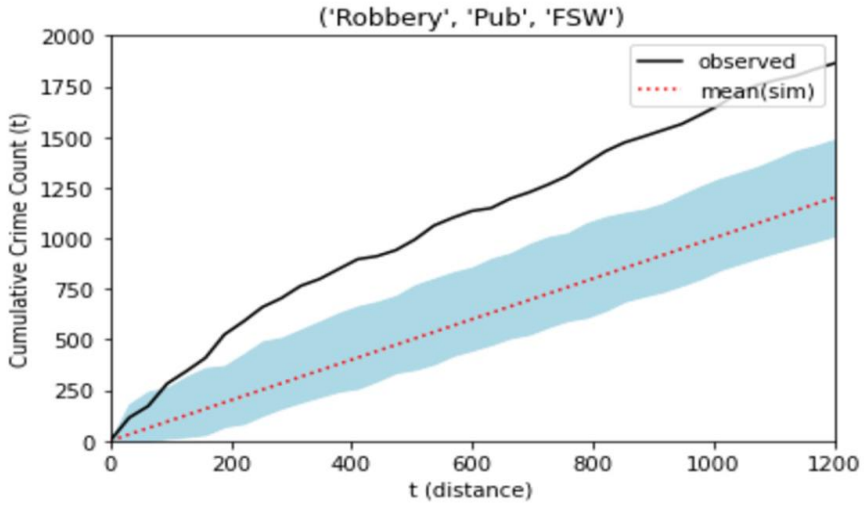
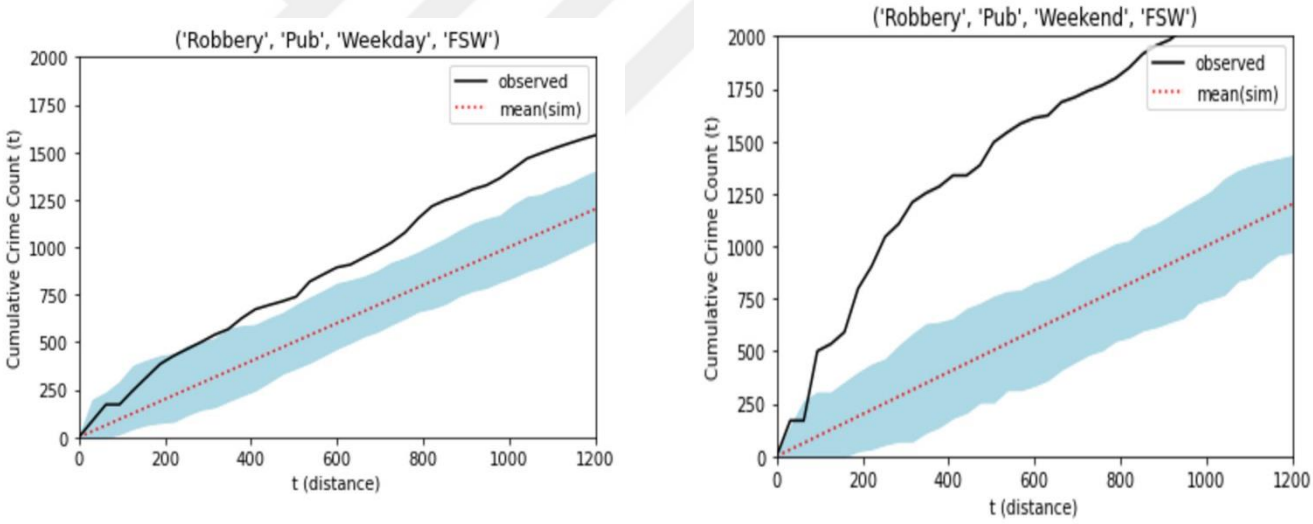


Figure 30**.** Network K result of the default model for robbery and pubs in FSW



(**a**) (**b**)

Figure 31. Network K result of the (a) weekday (b) weekend model for robbery and pubs in FSW

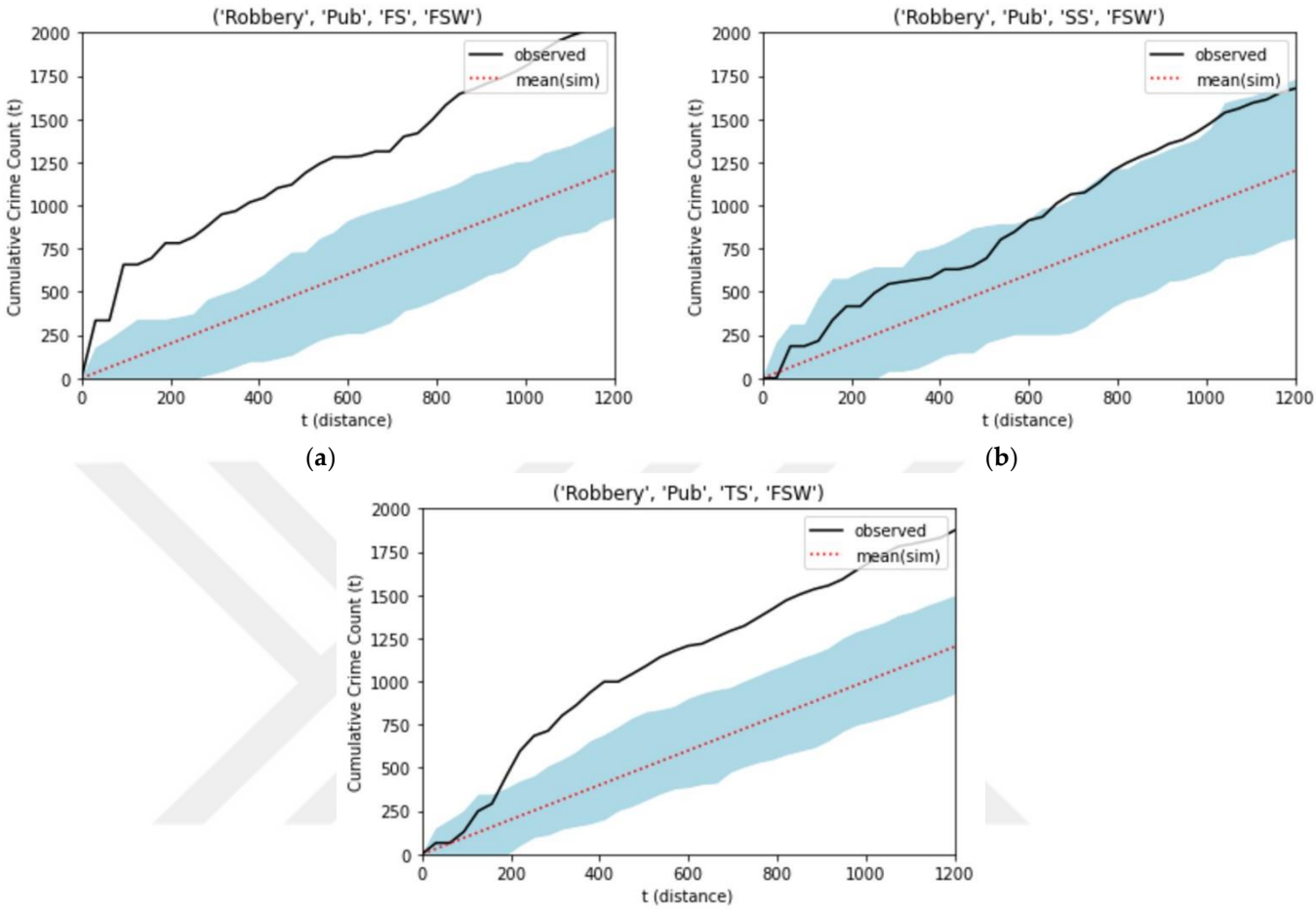


Figure 32**.** Network K result of the (a) FS (b) SS (c) TS model for robbery and pubs in FSW

The continuous black lines in these figures show the cumulative crime (i.e., robbery) counts within a 1200 m bandwidth around urban features (i.e., pub). Figure 32 indicates that the observed cumulative robbery count exceeds the expected count under the CSR hypothesis after nearly 100 m. Here, the boundaries of the insignificance area for each distance are drawn by the gray band running through the middle of the figure. An observed count falling between the maximum and minimum values in this band signifies insignificant spatial association. These values on the extremes are obtained from a Monte Carlo simulation with 39 steps. The dashed line passing through the band is the

mean cumulative crime density values at each distance generated by 39 simulations for each point. Shortly, a continuous black line above the gray band shows significant clusters, within the gray band shows insignificant clusters, and below the gray band shows significant dispersions. For instance, Figure 32 can be interpreted as a significant robbery cluster around pubs.

In the default model, we found no notable robbery and theft clusters within a quarter mile of gas stations on the CS.

The weekly model also revealed that gas stations had a different spatial influence on robbery and theft. While its impact on the robbery was nonexistent on both weekdays and weekends in the NS, it was only present on weekends in the CS. In weekly models, there were additional inconsistencies between pub and robbery. On the SS and SWS, pubs did not draw robbery on weekends, and they did not attract robbery on the FSWS and WS on weekdays.

Intraday models revealed many nuanced spatial influences. Grocery stores on the NWS, for example, and pubs on the SS and SWS did not significantly draw robberies during the first shift. Similarly, no significant number of robberies occurred near gas stations on the CS, FNS, or NS during the second shift.

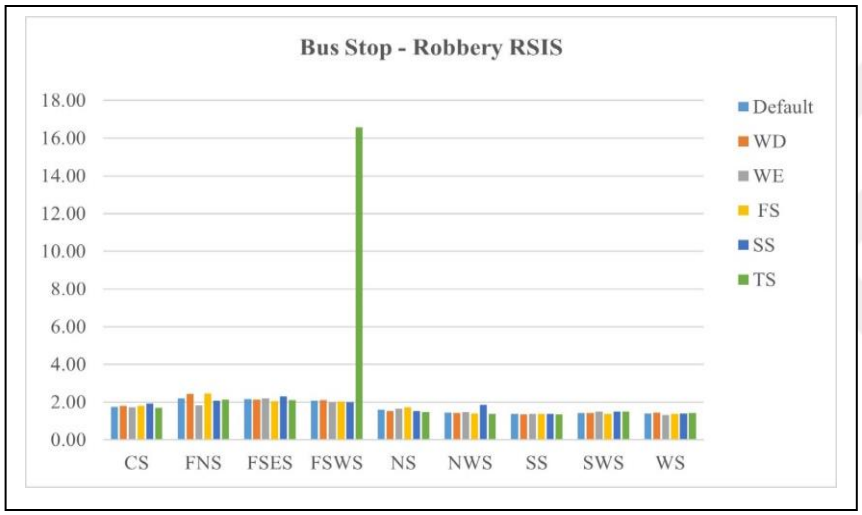
Intraday models identified many complex spatial influences. During the first shift, for example, we did not find any robbery clusters near grocery stores on the NWS or pubs on the SS and SWS. During the second shift, there were no significant robbery clusters around gas stations on the CS, FNS, or NS.

4.2.2. Spatiotemporal influence analysis

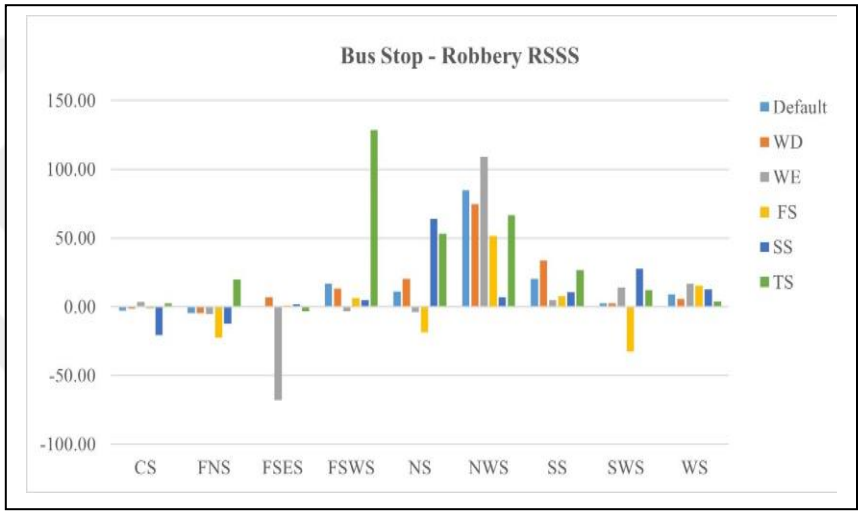
The mean MAE and RMSE values of the fitted risk functions (n = 509) were determined to be 0.119 and 0.13675, respectively, as overall performance indicators. Significant

disparities in RSIS and RSSS were discovered in spatiotemporal influence analysis across different crime categories (Appendix B).

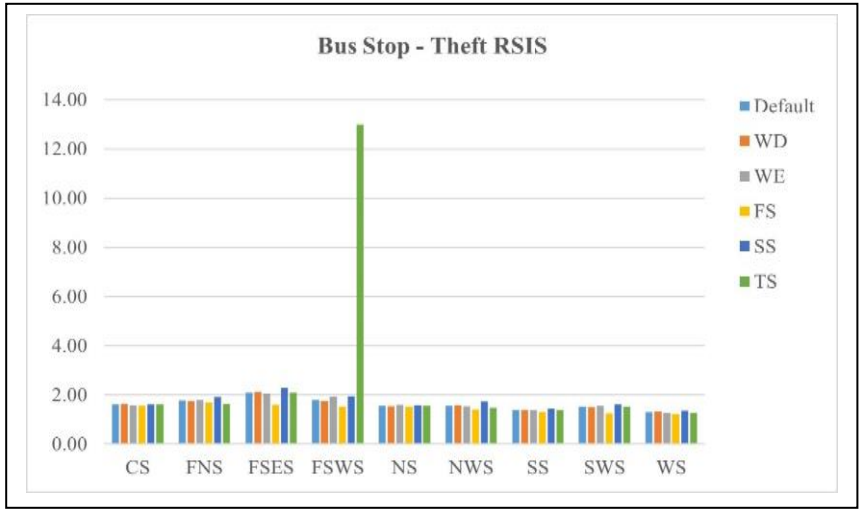
The RSIS and RSSS values of the bus stop in robbery and theft incidents are displayed in Figure 33 across spatiotemporal models. The Wilcoxon signed-rank test revealed that in the default model, theft RSIS of bus stations grew significantly in the second shift (p-value <0.01) and reduced significantly in the first shift (p-value <0.05). During the first shift, it also suggested a greater robbery RSIS (p-value <0.05). In the first shift, the theft RSSS was significantly greater than that of robbery. The RSIS of theft and robbery was lower at fast-food restaurants and grocery stores in general, according to the Kruskal-Wallis test. In the case of RSSS values, the relationship was flipped, with the majority of bus stops having positive RSSS values. The theft and robbery RSISs of bus stops in the FNS, FSES, and FSWS were found to be significantly higher. In both the CS and NS, bus stations were found to have the lowest values in both crime types for RSSS.



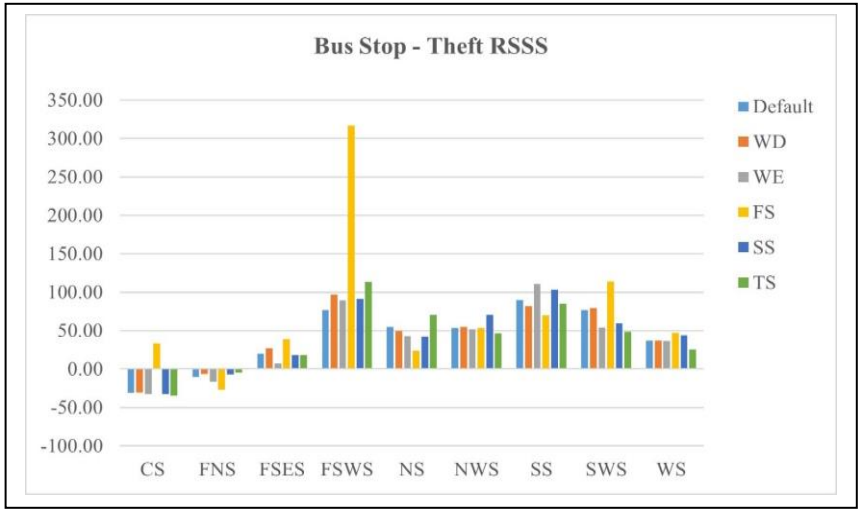
(a)



(b)



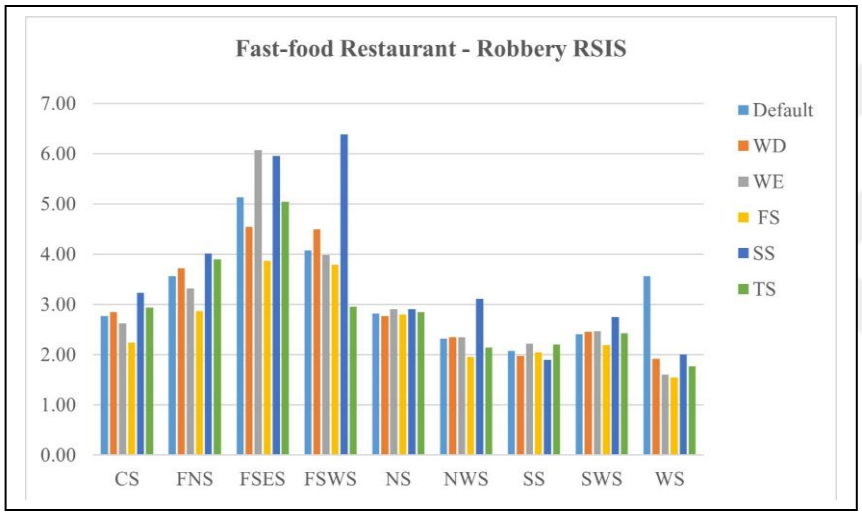
(c)



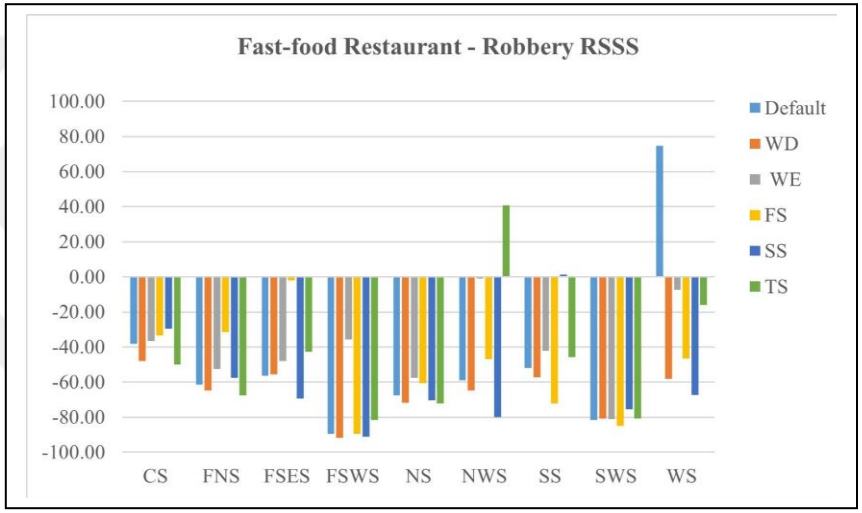
(d)

Figure 33. Bus stops RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models.

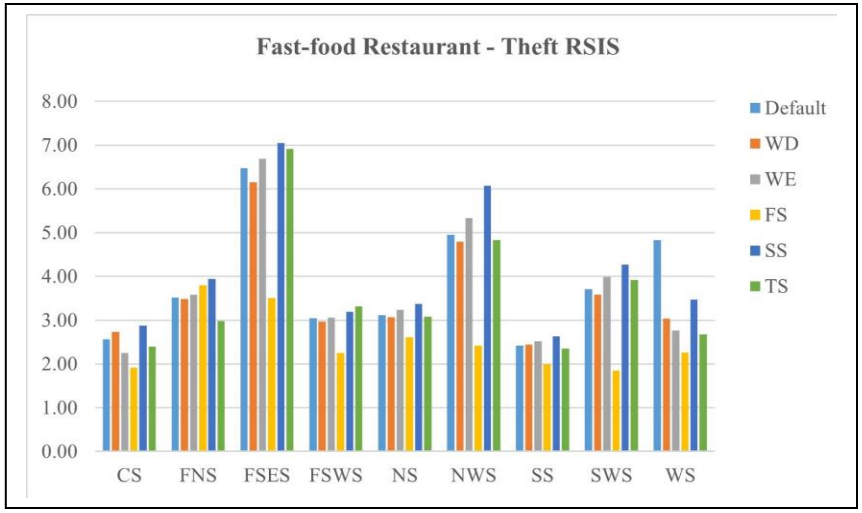
Figure 34 depicts the spatial-temporal theft and robbery RSIS and RSSS of fast-food restaurants. The key results revealed that the default model's RSIS for both theft and robbery is significantly higher than the first shift models (p-value< 0.01). A Kruskal-Wallis test for intraday differences found that theft intensity near fast-food restaurants was significantly lower in the first shift than in the other shifts (p-value< 0.01). For robbery, we found significant RSSS differences between the default and weekday models (p-value<0.05) as well as between the weekend and weekday models (p-value<0.01). The results also revealed that during the second and third shifts, the robbery intensity of fast-food restaurants was much higher than that of pubs. Except for the first shift model, all temporal models showed that fast-food restaurants had greater theft intensity values than gas stations and pubs. The tests found that fast-food restaurants in the FNS, FSES, and FSWS had significantly higher robbery RSISs. Significantly larger values were generally reported on the FSES for theft incidents. In the case of RSSS, the CS and FSES sides were shown to have higher values.



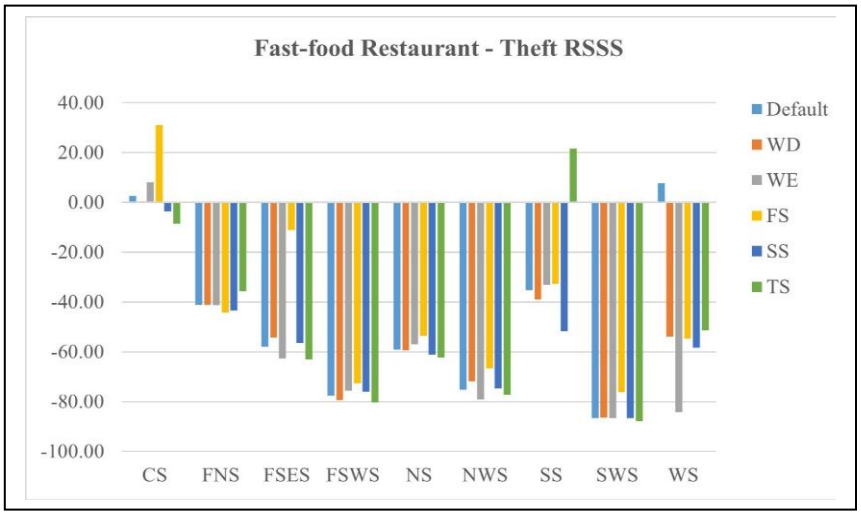
(a)



(b)



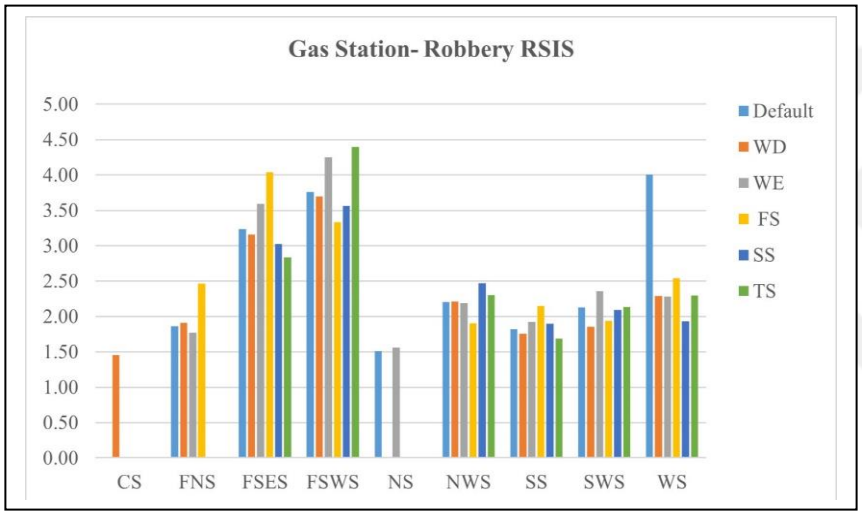
(c)



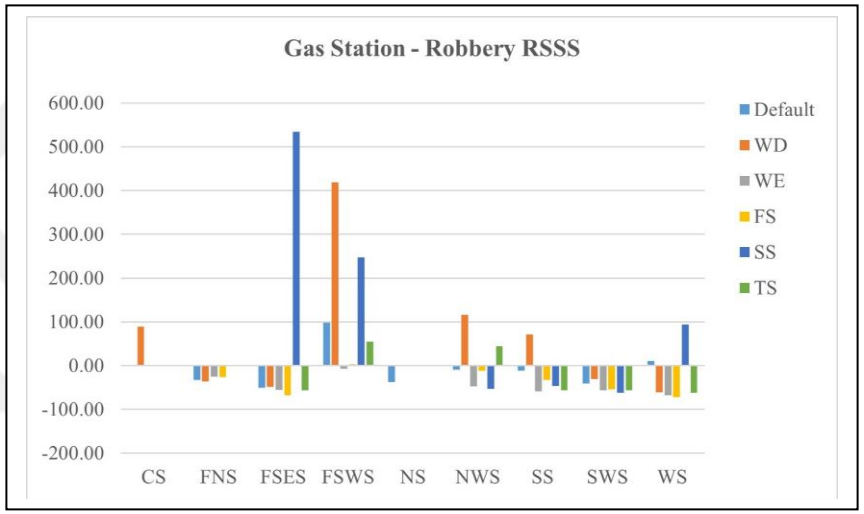
(d)

Figure 34. Fast-food restaurant RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models.

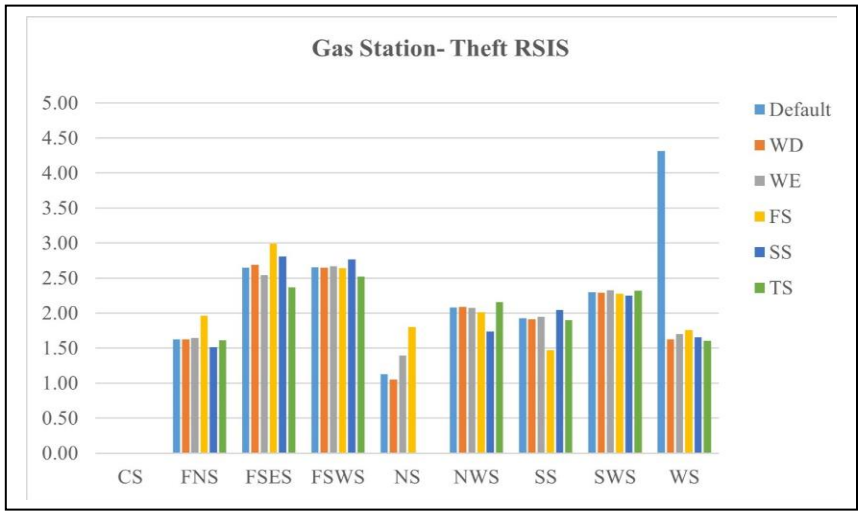
Figure 35 shows the results of the analysis for the gas station. When compared to theft values, the Wilcoxon signed-rank test showed that robbery RSIS was significantly higher on weekends (p-value < 0.05). We also discovered that robbery RSSSs were greater on weekdays. Across all temporal models, a comparative analysis using other features revealed significant theft intensity variations between gas stations and fast-food restaurants. This difference in the robbery was significant for RSSS reported across the default, weekday, and second shift models. Further tests revealed a significant increase in robbery intensity around gas stations on the FSES and FSWS sides, as well as a significant increase in theft on the FSES, FSWS, and SWS sides. The SWS was consistently found to have the smallest values for theft RSSS.



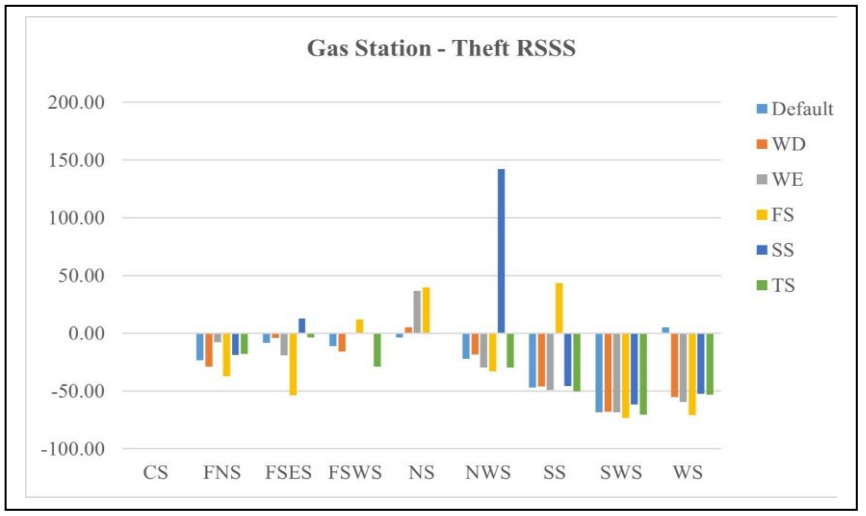
(a)



(b)



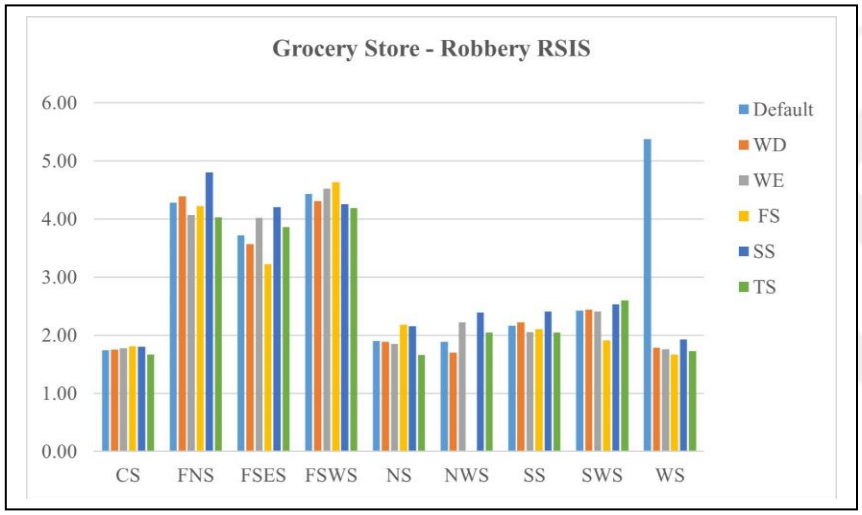
(c)



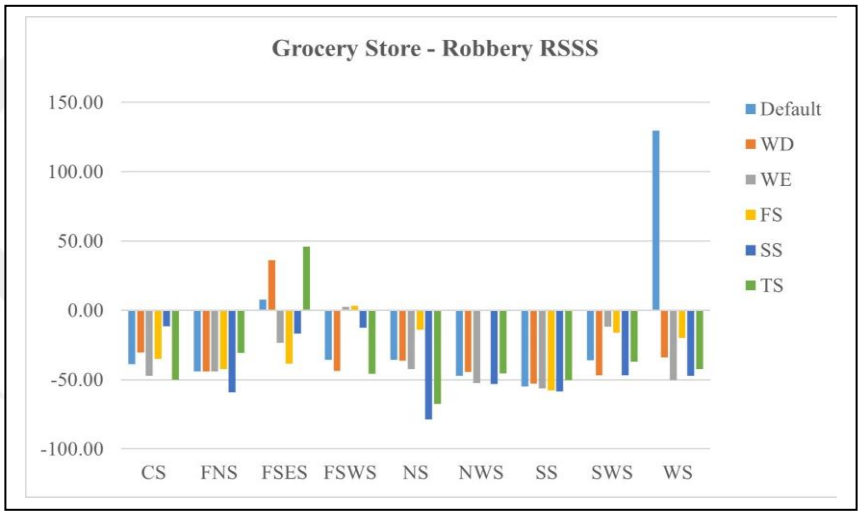
(d)

Figure 35. Gas station RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models.

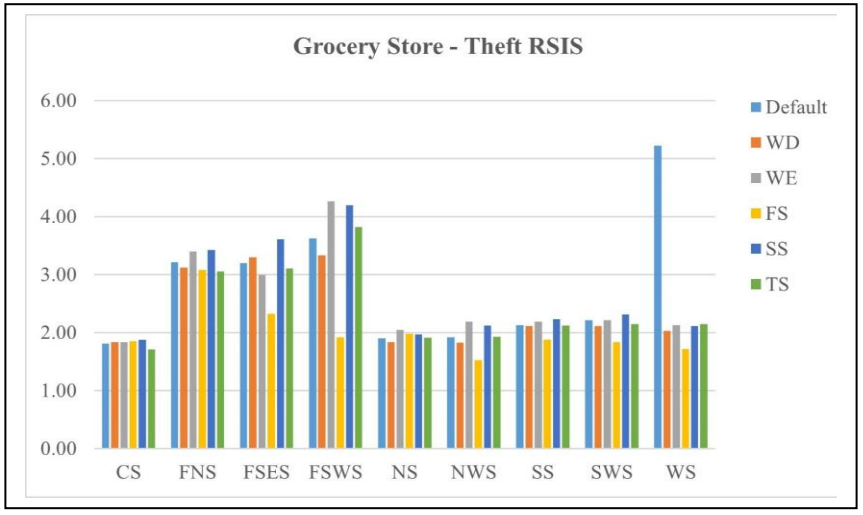
The results of grocery stores across spatiotemporal models are shown in Figure 36. The results showed that theft RSIS of grocery stores was significantly greater in the default model than in the first shift (p-value 0.05). During the second shift, robbery RSISs were also higher than their theft equals (p-value 0.05). In the default (p-value <0.01), weekday (p-value <0.01), and third shift (p-value< 0.05) models, robbery RSSS values were significantly higher than theft values. In second shift models, theft and robbery intensity were much greater around grocery stores than around pubs, according to a comparative analysis with other features. On the FNS, FSES, and FSWS, the side difference tests revealed significantly higher theft and robbery RSISs. Grocery stores were found to have the highest values on the FSES for theft RSSS.



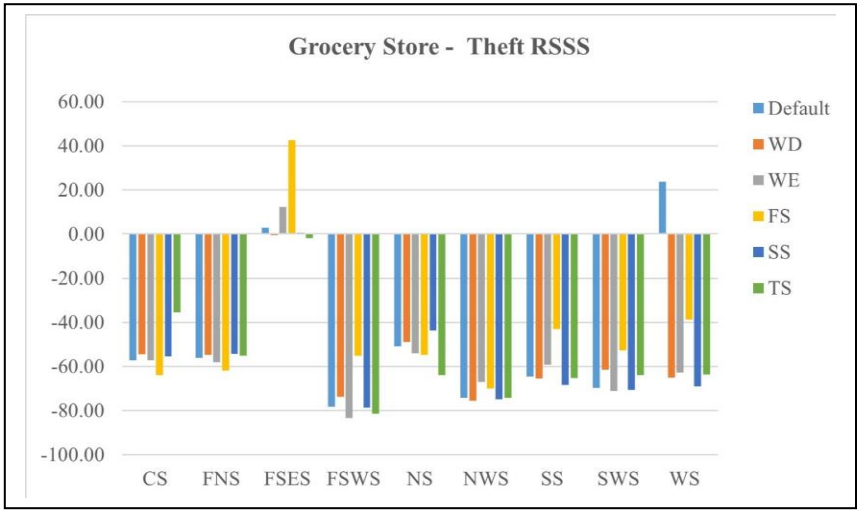
(a)



(b)



(**c**)



(**d**)

Figure 36. Grocery store RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models.

Figure 37 shows the results of a pub's spatiotemporal influence analysis. The theft RSISs of pubs in the default and weekday models were significantly higher than the weekday (p-value< 0.05) model and significantly lower than the weekend (p-value<0.05) model. In the first shift, we also discovered lower robbery RSSS values than in the default model. Also in the second shift, there were greater robbery RSSSs. The results also showed that robbery and theft RSSSs in pubs were much lower than those in fast-food restaurants throughout weekdays, second shift, and third shift. The Kruskal-Wallis test found that robbery RSISs in pubs on the SW and SWS were much lower, as were theft RSISs on the SS.

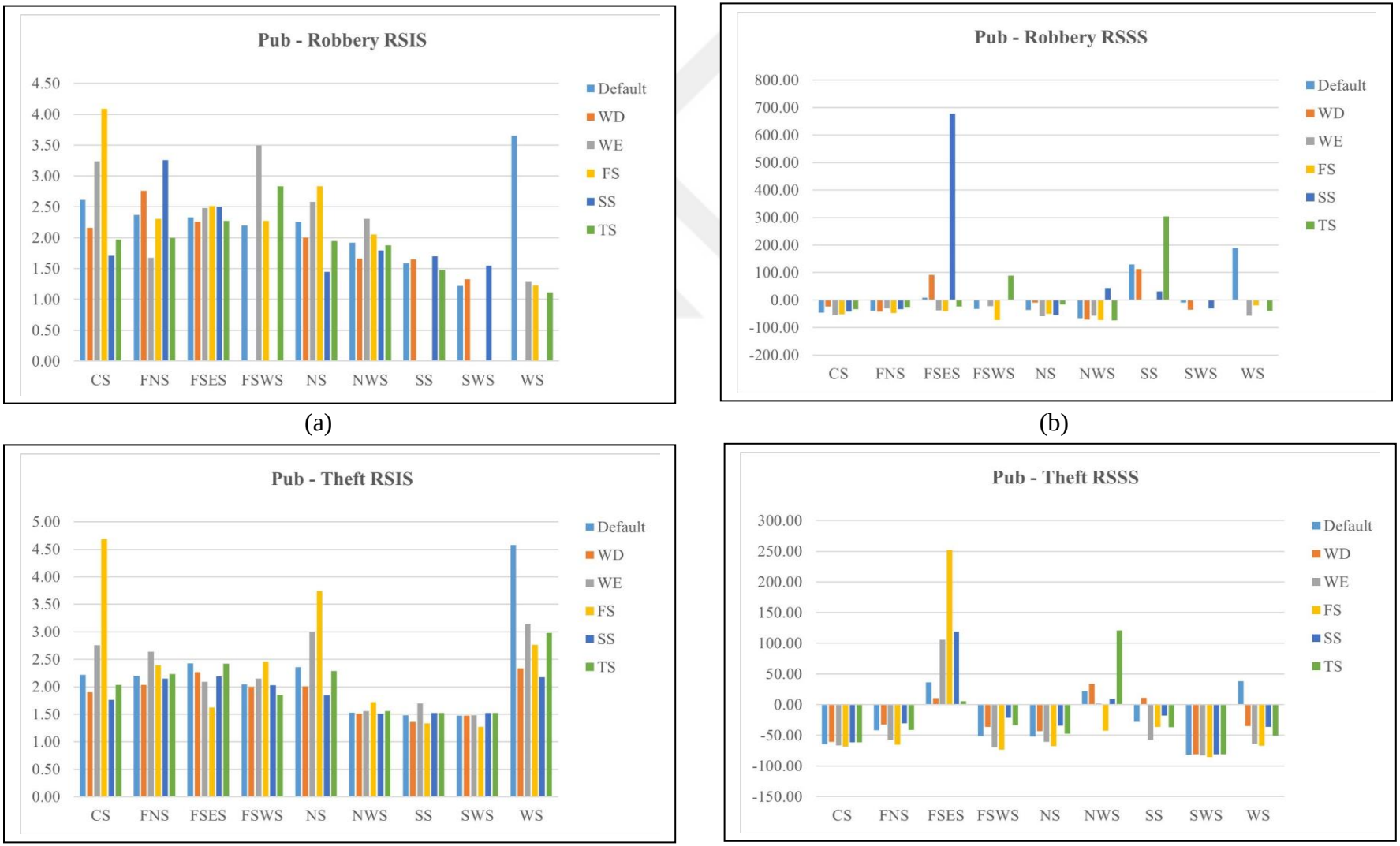


Figure 37. Pub RSIS on (**a**) robbery and (**c**) theft, and their RSSS on (**b**) robbery and (**d**) theft across spatiotemporal models.

4.2.3. Correlation analysis with concentrated disadvantage (CD)

The key findings discovered that some RSIS and RSSS values of urban features for various crime types were moderately correlated with each other as well as the concentrated disadvantage of the sides. In all temporal models, we found strongly correlated theft and robbery RSISs around bus stops. In the case of gas stations, the correlation analysis revealed an increase in crime intensity in the disadvantaged neighborhoods (Figures 38 and 39). In the second shift models, there was a strong correlation between CD and the intensity of theft ($\rho_{\text{Theft, CD}} = 0.92$) and robbery ($\rho_{\text{Robbery, CD}} = 0.76$).

For readers' convenience, the cell labels follow the naming convention of: urban feature risk score_ temporal model. For example, GS_TI_Default refers to theft RSIS of gas stations in the default model). CD, on the other hand, had a negative correlation with robbery RSSS during the first shift ($\rho_{\text{Robbery, CD}} =0.69$), as well as theft and robbery RSSS during the weekend ($\rho_{\text{Robbery, CD}} =0.79$ and $\rho_{\text{Theft, CD}} =0.68$). In the first and second shift models, the only significant correlations with CD were reported in the theft ($\rho_{\text{Theft, CD}} = 0.89$) and robbery ($\rho_{\text{Robbery, CD}} = 0.72$) intensities, respectively.

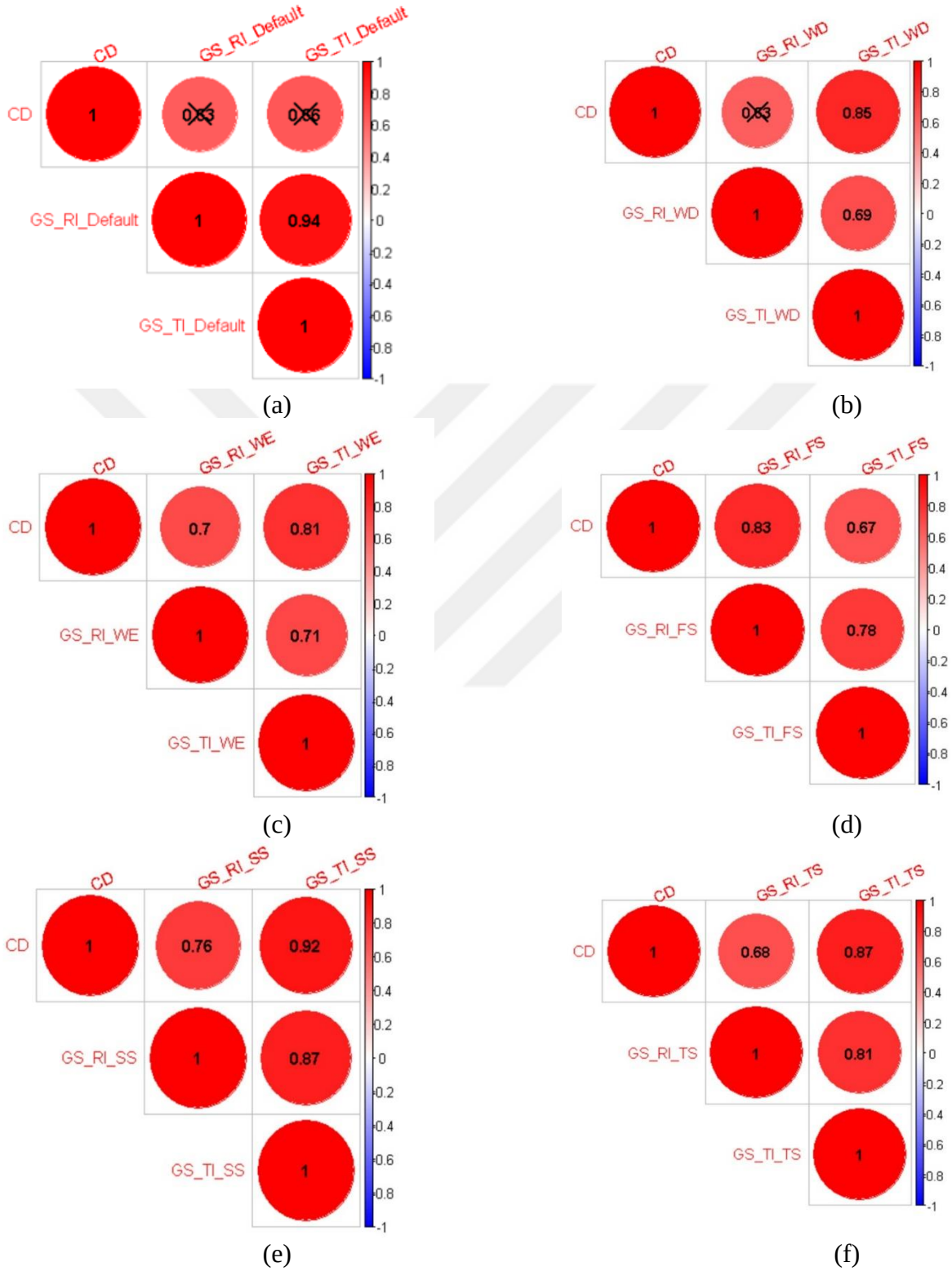


Figure 38. Correlograms between concentrated disadvantage (CD) and the RSIS values of gas stations in (a) default, (b) weekday, (c) weekend, (d) FS, (e) SS, and (f) TS models.

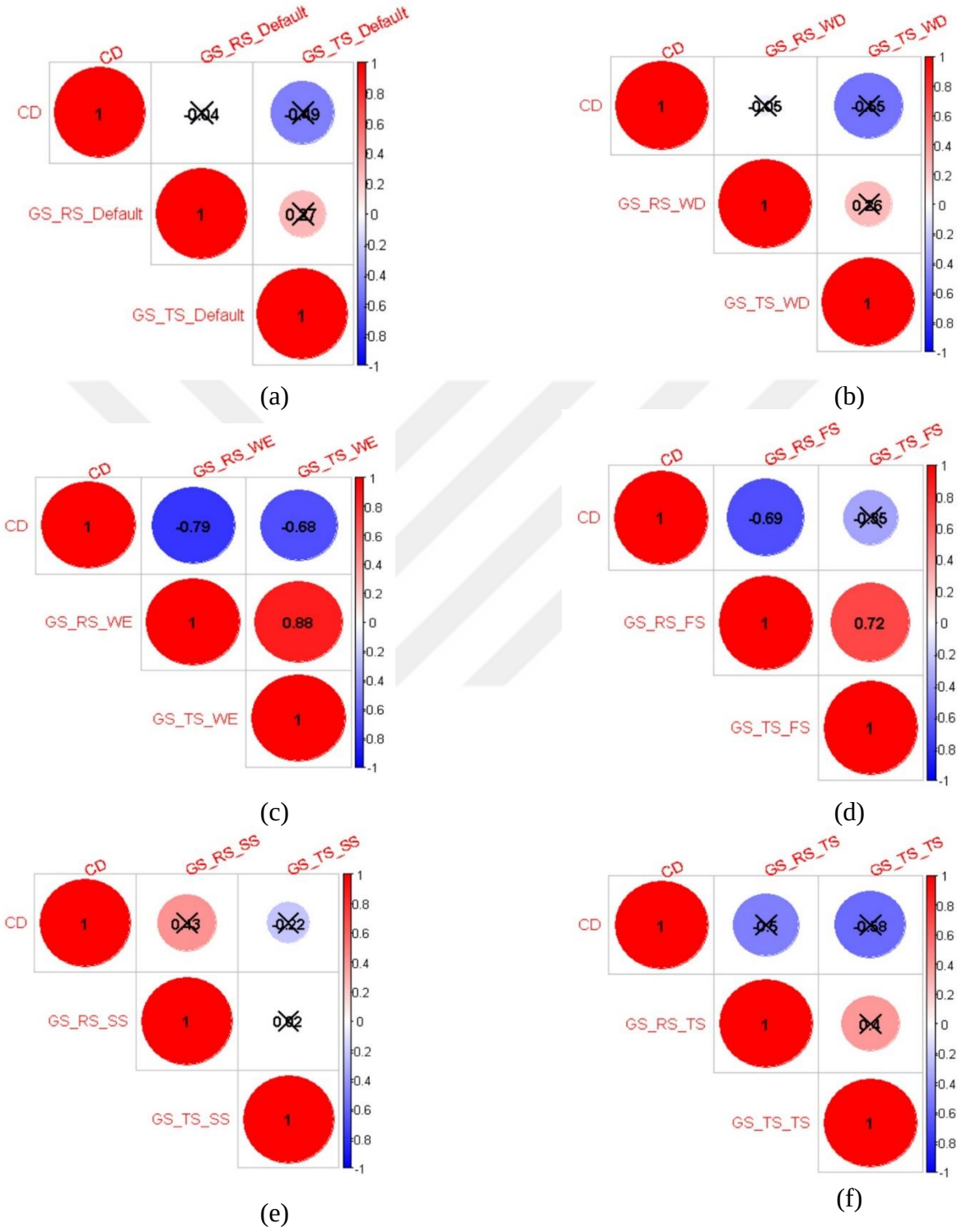


Figure 39. Correlograms between concentrated disadvantage (CD) and the RSSS values of gas stations in (**a**) default, (**b**) weekday, (**c**) weekend, (**d**) FS, (**e**) SS, and (**f**) TS models.

4.3. Predictive crime hotspot mapping algorithms

This section will report the results of the developed predictive crime hotspot mapping algorithms. The subsections were divided based on the theory-based auxiliary event surfaces that enhance the performance of the retrospective models. We begin each subsection by reporting the optimal hyperparameters regarding data representation and the model. Next, we present the MHRs of the DL algorithms using these hyperparameters in daily and shift models. Lastly, a comparative analysis highlights the relative performance improvement upon the autoregressive crime hotspot prediction models achieved by the incorporation of each theory-based auxiliary event surfaces.

4.3.1. Autoregressive model

As displayed in Table 7., the autoregressive models highlighted the differences between models in both data representation and model. For the robbery, Graphwavenet outperformed the others with an MHR of 0.562 in daily models whereas STGCN achieved the best predictive performance with 0.603 in shift models. In general, graph-based algorithms achieved nearly twice MHR compared to LSTM that produced traditional sequence-to-sequence predictions without considering spatial dependence between the street segments.

Table 7. The Daily and Shift Results of the Autoregressive Robbery Models

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	42	0.05	0.1	0.05	0.562	42	0.05	0.1	0.05	0.495
STGCN	14	0.05	0.3	0.04	0.528	42	0.9	0.3	0.05	0.584
LSTM	42	0.5	-	0.01	0.232	3	0.1	-	0.02	0.228

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

The performance of the selected algorithms in theft models is much higher than their performances in robbery models (see Table 8). From the daily prediction models, STGCN is the best one with an MHR of 0.776. For the shift models, Graphwavenet is the best performer that identifies 59.6% of the theft incidents on average in the next shift. One interesting point is the reversed values of optimal training windows between the theft and robbery models. While optimal TWs are 42, 14, 42 for GraphWavenet, STGCN, and LSTM respectively for the robbery models, these values change to 14,42,14 for the theft models.

Table 8. The Daily and Shift Results of the Autoregressive Robbery Models

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	14	0.05	0.2	0.03	0.764	21	0.05	0.3	0.1	0.596
STGCN	42	0.05	0.1	0.02	0.776	30	0.3	0.3	0.05	0.561
LSTM	14	0.05	-	0.04	0.245	3	0.9	-	0.04	0.245

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

4.3.2. SSRS

As displayed in Table 9, the robbery results indicated that Graph Wavenet is the best performer in daily models with an MHR of 0.580. For the shift models, however, STGCN outperformed the others with an MHR of 0.599. The results also indicate a shorter TW for GraphWavenet with a value of 10 in daily models compared to others both having TWs of 30. The optimal SBW for both graph learning algorithms is found at 0.3. For shift models, TWs are quite narrow for GraphWavenet and LSTM with values of 3.

Table 9. The Daily and Shift Results of the Robbery Models with SSRS

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	10	0.1	0.3	0.05	0.580	3	0.5	0.3	0.07	0.513
STGCN	30	0.1	0.3	0.02	0.554	30	0.05	0.3	0.08	0.599
LSTM	30	0.9	-	0.05	0.239	3	0.9	-	0.03	0.267

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

For daily theft predictions with SSRS (see Table 10), GraphWavenet and STGCN achieved comparable results with MHRs of 0.80, and 0.79 respectively, nearly three and a half times higher than that of what LSTM achieved. The ratio is nearly two and a half times for shift predictions with MHRs 0.605, 0.608, and 0.240 for GraphWavenet, STGCN, and LSTM respectively. One last important point here is that 0.05 is found as the optimal SC in all daily models, indicating a larger temporal extent.

Table 10. The Daily and Shift Results of the Theft Models with SSRS

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	30	0.05	0.2	0.02	0.800	3	0.5	0.3	0.02	0.605
STGCN	14	0.05	0.3	0.06	0.799	30	0.3	0.3	0.04	0.608
LSTM	14	0.05	-	0.05	0.242	21	0.05	-	0.07	0.240

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

4.3.3. Feature counts

The results of the robbery models (see Table 11) with feature counts marked Graph Wavenet as the best performing algorithms in both daily and shift models with MHRs of

0.514 and 0.496 respectively. The second-best model is STGCN with extremely reduced performances in both levels, 0.374 and 0.309 respectively. One important result is the reliance of STGCN on larger training windows in robbery shift models.

Table 11. The Daily and Shift Results of the Robbery Models with Feature Counts

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	10	0.9	0.3	0.08	0.514	3	0.1	0.2	0.01	0.496
STGCN	30	0.05	0.3	0.07	0.374	42	0.05	0.2	0.08	0.309
LSTM	30	0.5	-	0.04	0.211	3	0.5	-	0.04	0.233

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

Similar to robbery models, GraphWavenet was found superior to the other two selected algorithms with MHRs of 0.686 across daily models. For the shift model, STGCN achieved an MHR of 0.607, which is even higher than the one obtained in the autoregressive model. An important result found that the optimal TW and SC values across GraphWavenet and LSTM are the same (TW=3, and SC=0.5). Differently, STGCN needed the greatest number of timesteps. The results are displayed in Table 12.

Table 12. The Daily and Shift Results of the Theft Models with Feature Counts

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	42	0.5	0.2	0.01	0.686	3	0.5	0.2	0.09	0.505
STGCN	30	0.05	0.3	0.07	0.538	42	0.05	0.2	0.06	0.607
LSTM	10	0.05	-	0.04	0.232	3.	0.5	-	0.05	0.229

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

4.3.4. Other crime

Table 13 indicated that Graph Wavenet outperformed the other algorithms in robbery models by achieving a nearly two times higher daily and shift MHR with values of 0.607 and 0.584. We noted larger SBW values for graph learning algorithms in daily models, which indicates a much smaller spatial dependence between the neighboring segments.

Table 13. The Daily and Shift Results of the Robbery Models with Other Crime (I.E., Theft)

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	42	0.05	0.3	0.02	0.607	21	0.1	0.2	0.1	0.584
STGCN	30	0.05	0.3	0.08	0.360	42	0.9	0.1	0.05	0.280
LSTM	42	0.5	-	0.05	0.274	42	0.5	-	0.06	0.268

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

Table 14 shows that, for daily theft models, the Graph Wavenet algorithm achieved a much greater performance (MHR= 0.791) with a much shorter TW of 7 whereas STGCN has an MHR of 0.485 and a TW of 14, and LSTM has an MHR of 0.247 and a TW of 42. The theft shift models also showcased a sharp MHR difference in favor of the selected graph learning algorithms. While GraphWavenet achieved an MHR of 0.58, LSTM only predicted 22.9% of the hotspots in the next shift.

Table 14. The Daily and Shift Results of the Theft Models with Other Crime (i.e., Robbery)

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	7	0.1	0.1	0.05	0.791	3	0.5	0.2	0.001	0.580
STGCN	14	0.9	0.3	0.04	0.485	30	0.5	0.2	0.04	0.576
LSTM	42	0.1	-	0.05	0.247	3	.0.1	-	0.07	0.229

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

4.3.5. 311 calls

As displayed in Table 15, the best daily and shift robbery models were achieved by the Graphwavenet algorithm that achieved MHRs of 0.618 and 0.534 respectively. Using a week as a TW in the daily model and a day in shift models, this algorithm achieved the best performance with an LR of 0.03 and 0.001 respectively. The second-best algorithm is the other graph learning algorithm, STGCN. An interesting point here is that STGCN performed better in the shift model. Lastly, LSTM delivered a similar performance around 0.20 with a value of 0.24. Another notable point here is that robbery shift models with LSTM achieved a better performance in the shift models with a nearly 10% increase in the MHR.

Table 15. The Results of Daily and Shift Robbery Models with 311 Calls

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	7	0.05	0.1	0.03	0.618	3	0.05	0.1	0.001	0.534
STGCN	30	0.1	0.3	0.08	0.494	30	0.05	0.2	0.05	0.524
LSTM	10	0.05	-	0.04	0.240	42	0.5	-	0.005	0.266

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

For theft models (see Table 16), GraphWavenet achieved the best MHR with a value of 0.794 and STCGN is the best performer with an MHR of 0.626. GraphWavenet algorithm used the minimum TWs from the experimented values (i.e., 7 for daily, and 3 for shift models). STGCN used 30-time steps as the training window in both models. On the other hand, LSTM again delivered a performance not much different from what it delivered in other models with MHRs of 0.261 and 0.255 in daily and shift models respectively.

Table 16. The Results of Daily and Shift Theft Models with 311 Calls

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	7	0.05	0.1	0.001	0.794	3	0.05	0.1	0.03	0.573
STGCN	30	0.9	0.3	0.05	0.538	30	0.3	0.3	0.05	0.626
LSTM	30	0.1	-	0.04	0.261	3	0.05	-	0.04	0.255

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

4.3.6. Park events

For the robbery models (see Table 17), the results of park events largely indicated an inferior performance for the selected algorithms. The most dramatic decrease occurred in the robbery shift models that output MHRs of 0.398, and 0.257 much lower than these algorithms delivered in the previous models with different auxiliary event surfaces.

Table 17. The Results of Daily and Shift Robbery Models with Park Events

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	30	0.05	0.1	0.08	0.593	3	0.05	0.3	0.01	0.398
STGCN	30	0.05	0.3	0.07	0.404	30	0.05	0.2	0.09	0.392
LSTM	14	0.05	-	0.07	0.235	42	0.5	-	0.08	0.257

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

The most striking result in the theft shift models (see Table 18) is the MHR obtained from the Graph Wavenet algorithm. It is 0.745, much higher than the previous models. The algorithms needed longer TWs for the daily models, whereas these values were relatively shorter in the shift models that use a TW of three shifts. One last point worth noting here is the relatively smaller values of SCs. These values indicate the daily and shift theft models require the data to be represented with a larger temporal extent.

Table 18. The Results of Daily and Shift Theft Models with Park Events

	Daily					Shift				
	TW	*SC*	*SBW*	*LR*	*MHR*	*TW*	*SC*	*SBW*	*LR*	*MHR*
Graph Wavenet	30	0.05	0.1	0.001	0.796	3	0.1	0.3	0.004	0.745
STGCN	14	0.05	0.1	0.06	0.710	30	0.05	0.3	0.005	0.606
LSTM	42	0.1	-	0.02	0.247	3	0.05	-	0.006	0.256

Note: **TW**=Train Window, **SC**= Smoothing Coefficient, **SBW**= Spatial BandWidth, **LR**= Learning rate

4.3.7. Comparative performance analysis

The main results suggest that GraphWavenet is the best performing algorithm. Across 24 models (i.e.,6 daily and 6 shift models for each selected crime type), we found the

highest MHR for GraphWavenet in 17 models. The maximum reported MHR value of Graph Wavenet was reported at the daily theft model enhanced by SSRS with an MHR of 0.80. This means that it predicted the locations on average 80% of theft incidents the next day. For the theft shift models, we calculated an MHR of 0.745 in the model enhanced by the park events. We found the highest robbery MHR in the daily model enhanced by the 311 calls surface with a value of 0.618. The best robbery shift model is the one with the theft surface (MHR= 0.584). A comparative analysis with the autoregressive model found that the daily robbery model was best improved by the incorporation of 311 calls with an increase of 9.9%. For the shift models, it is the other crime that achieved an 18% improvement on the MHR of the autoregressive model. Our novel SSRS is found as the surface that most improves the performance of the autoregressive theft daily model with a 4.7% increase in MHR. For theft shift models, it is the incorporation of park events that improves the MHR of the autoregressive model most by 68%.

The other graph learning algorithm, STGCN, is the second-best algorithm in terms of MHR. The maximum MHR value of STGCN is reported at the daily theft model with SSRS with an MHR of 0.799. For the theft shift models, we calculated the highest MHR with a value of 0.626 in the model with 311 calls. We found the highest daily and shift robbery MHRs in the model with SSRS surface values of 0.554 and 0.599 respectively. A comparative analysis with the autoregressive model found that daily and shift robbery models were best improved by the incorporation of SRSS with increases of 4.8% and 2.5% respectively. One striking result indicated that all the auxiliary event surfaces, except for our SSRS, led to a decrease in the performance of the autoregressive models across all the model configurations except for the theft shift model. The model that

benefits most from our novel SSRS was the theft shift model. However, it is not SSRS that most improved the autoregressive theft shift model. The event surface that led to the greatest increase was 311 calls with an 11% increase in MHR. All the results are displayed in Table 19.

Table 19. The Results of Predictive Crime Hotspot Mapping Algorithms

	Autoregressive				SSRS				Feature Count				Other Crime				311 Calls				Park Events			
	Robbery		*Theft*		*Robbery*		*Theft*		*Robbery*		*Theft*		*Robbery*		*Theft*		*Robbery*		*Theft*		*Robbery*		*Theft*	
	DM	*SM*	*DM*	*SM*	*DM*	*SM*	*DM*	*SM*	*DM*	*SM*	*DM*	*SM*	*DM*	*SM*	*DM*	*SM*	*DM*	*SM*	*DM*	*SM*	*DM*	*SM*	*DM*	*SM*
GWN	0.562	0.492	0.764	0.596	0.58	0.513	0.8	0.605	0.514	0.496	0.686	0.505	0.607	0.584	0.791	0.58	0.618	0.534	0.794	0.573	0.593	0.398	0.796	0.745
STGCN	0.528	0.584	0.776	0.561	0.554	0.599	0.799	0.608	0.374	0.309	0.538	0.607	0.360	0.283	0.485	0.600	0.494	0.524	0.538	0.626	0.404	0.392	0.710	0.606
LSTM	0.232	0.228	0.245	0.245	0.239	0.267	0.242	0.24	0.211	0.233	0.232	0.229	0.274	0.268	0.247	0.229	0.24	0.266	0.261	0.255	0.235	0.257	0.247	0.256

Note: DM= Daily Model, SM= Shift Model, GWN=Graph Wavenet, STGCN= Spatiotemporal Graph Convolution Neural Network

CHAPTER 5

DISCUSSION

This chapter will heavily discuss the findings of the study in two parts. While the first part emphasizes how DRSFs contributed to the understanding of the spatial influence of urban features on different crime types, the second will elaborate on how and why the theory-injected predictive crime hotspot mapping methods are different from the previously developed methods. These parts will mostly touch on our theoretical and empirical interpretations of the results. The focus is placed on how (mis)aligned the current study is with the previous research that we overviewed in the literature review.

5.1. DRSF

The bivariate network K analysis revealed potential spatial influence dead zones. These zones ensue from specific spatiotemporal configurations in which urban features do not broadcast any crime risk signals in their environs. Furthermore, we discovered that crime clusters near urban features differed across the sides throughout a day or week. The default model, for example, discovered minor theft and robbery clusters around gas stations around the CS. However, a weekly model revealed that robberies occurred in a significant cluster on the same side on weekdays. Similarly, while all of the temporal models except the default model showed insignificant robbery clusters near gas stations on the NS, they all found significant theft clusters on the same side. Similar to what past research documented (Corcoran, Zahnow, Kimpton, Wickes, & Brunsdon, 2021; Hipp & Kim, 2019; Andresen & Malleson, 2015; Yue, Zhu, Ye, & Guo, 2017; Feng, Piza, Kennedy, & Caplan, 2019), we found distinct intraday and weekly crime patterns near urban features across different regions. Another significant finding was that theft clusters were more consistent around all urban features than robbery clusters, with the

exception of insignificant clusters around gas stations in a few spatiotemporal models. This underscores the situational crime opportunity by stressing offenders' differing reward definitions across different crime types (Clarke, 1995). Evident from more consistent spatiotemporal crime clusters around urban features, thieves appear to be less impacted by contextual circumstances than robbers. Robbers, on the other hand, use violence and intimidation against their victims, which may lead to a physical struggle between the individuals involved. Outsiders who can act as capable guardians may pay more attention to this struggle (Cohen & Felson, 1979), thus deterring the robbers from offending.

When compared to other urban features, such as fast-food restaurants, bus stops had much less spatial influence in nearly all temporal models, according to the RSIS findings (Song et al., 2019). On the contrary, numerous positive RSSS on all sides, except the CS and FNS, demonstrated that they actually had a stronger spatial diffusion effect throughout a spatial extent, especially in theft. This discovery stands in contrast with prior research that had empirically established the distance decay effect (Ratcliffe,2012; Groff, 2013; Groff, 2011; Xu & Griffiths,2017; McCord & Ratcliffe, 2009). As one possible explanation, the presence of other persons waiting for or getting off the bus who can function as informal security agents could make a bus stop a safe space for potential victims (Jacobs, 1961). Potential offenders are drawn to more deserted regions (i.e., alleyways and connector roads, etc.) as a result of the increasing number of capable guardians, generating a spatial spillover effect into neighboring areas (Angel, 1968). However, some urban features (e.g., schools and bars) were reported to have a differential spatial influence across regions, by a body of research (Wheeler, 2019; Breetzke & Edelstein, 2020). This provides supporting evidence to the MCOT (Wilcox,

Land, & Hunt, 2003) that delineates a tight relationship between social and physical context. Accordingly, the changing crime risk around urban features across the sides of Chicago may accrue from the differential exposure levels to the offenders in areas. Another intriguing finding was that all temporal models showed positive robbery RSSSs of pubs on the SS. Their theft RSSSs, on the other hand, significantly backed the distance decay effect on the same side. In conclusion, our findings demonstrated that risk intensity and strength varied considerably depending on location, time, and crime type.

The results of comparative RSSS analysis through the Wilcoxon signed-rank and Kruskal-Wallis test provided mixed support for spatiotemporal differences conditioned by crime type (de Melo, Pereira, Andresen, & Matias, 2018). The insignificant intraday differences were consistent with the findings of Bernasco, Ruiter, & Block (2017) who found only limited intraday crime level differences around urban features during weekdays and the time of day. We identified an intraday RSIS difference only for fast-food restaurants in this study. This difference occurred between the first shift than others, the former being significantly lower. Given that the majority of fast-food restaurants were closed between 00:00 to 07:59 a.m., this result was not unexpected. However, we found a lot of evidence for RSIS and RSSS temporal differences near urban features. For example, a Wilcoxon signed-rank test found that robberies around pubs were more intense on weekends than on weekdays. These findings suggest cyclic crime opportunity patterns that recur on a weekly basis around certain locations (Corcoran, Zahnow, Kimpton, Wickes, & Brunsdon, 2021; Hipp&Kim,2019; Andresen & Malleson,2015).

Spatiotemporal RSIS and RSSS study revealed that urban features have a greater spatial influence on FNS, FSWS, and FSES, which are located on Chicago's north and south borders. The correlation analysis indicated a strong positive link between theft and robbery RSISs of gas stations and concentrated disadvantage when looking at the various interactions between side-level variables and the degree of spatial influence. Theft and robbery RSISs of pubs and CD, on the other hand, were found to have an inverse relationship, although only in the first shift model. The increasing robbery and theft intensity near pubs is not surprising, given that the CS is the most affluent side of Chicago, IL, and is the city's core in terms of commerce, nightlife, and entertainment. This is because offenders are far more likely to encounter an intoxicated person who may be unable to protect themselves well in these situations (Roncek & Maier, 1991; Roncek & Pravatiner, 1989). The results provide mixed support to previous studies (Stucky & Ottensman, 2009; Browning et al., 2010) that found a negative relationship between concentrated disadvantage and spatial influence. The effect of concentrated disadvantage on spatial influence is mediated by the type of urban feature, according to these contradictory findings.

5.2. Enhancing DL-based predictive crime hotspot mapping with theory-based event surfaces

This study has provided empirical evidence to the superiority of graph-based deep learning (i.e., Graph Wavenet, STCGN) algorithms to a traditional counterpart in crime hotspot prediction. The results revealed graph learning algorithms attained performances as higher as nearly four times in some models. The key factor here is the ability of the graph-based algorithms to capture the spatial dependency between the street segments by incorporating the street network topology through adjacency matrices. Predicated

upon the spatial aspect of the near-repeat phenomenon that assigns an elevated crime risk area around a crime incident, these algorithms diffuse the existing crime risk in segments across the neighbors. We determined the optimal value by experimenting with three different Gaussian kernels having different spatial bandwidths.

Another reason behind the contrasting predictive performances between the graph-based algorithms and LSTM may be their approach to model temporal dependency. The traditional LSTM works with the recurrent units that suffer from prolonged training times and exploding/vanishing gradients with longer time series (Wu, Pan, Long, Jiang, & Zhang, 2019; Shleifer, McCreery, & Chitters, 2019). On the other hand, the selected graph-based algorithms exploit fully convolutional units that are much less computationally expensive, faster in training, and have lower memory requirements. In sum, this study is well-aligned with a large body of empirical research that reported the superiority of graph-based DL algorithms traditional DL algorithms across many domains (e.g., Geyer, 2017; Wang et al., 2018). One last point worth noting here is the performance differences between the graph learning algorithms. The results showed that Graph Wavenet was found as the best model in 17 out of 24 models. The most important difference between Graph Wavenet and STGCN is the self-adaptive adjacency matrices that eliminate the connection between the nodes with weak spatial dependence by using the ReLU function (Wu, Pan, Long, Jiang, & Zhang, 2019). Limiting the shared weights only to the nodes with strong connections might improve the predictive performance of the models by avoiding the impacts of the insignificant connections on the training weights.

The proposed two-step parameter tuning procedure underlines the importance of a data-driven approach in determining the optimal parameters for crime hotspot prediction

models. First and foremost, the crime data should be represented in a way that it can recognize the spatiotemporal patterns ingrained in the crime incidents. This is crucial because these represented datasets are then fed to the DL algorithms for producing crime hotspot predictions. An inaccurate representation may inflict the performance of the DL-based crime hotspot prediction algorithms. In that sense, training window (i.e., time lag), SC, and SBWs should reflect the behavior of crime risk across street networks as accurately as possible. The results to a large extent supported the differing spatiotemporal risk extent across the selected crime types and time. For instance, TWs in the theft daily models are 14, 42, and 14 for GraphWavenet, STCGCN, and LSTM respectively. On the other hand, the same TWs are 42,14, and 42 for the daily robbery models. A similar difference also occurs in other parameters related to data representations (i.e., smoothing coefficient and spatial bandwidth). These differential parameters can be ascribed to the unique spatiotemporal patterns observed in the crime types (Youstin, Nobles, Ward, & Cook, 2011). In contrast with the expectations, the learning rate was found to have only a weak impact on the predictive performance of the selected algorithms. While the parameters related to data representation have a substantial impact on the predictive performance of the algorithms, the learning rate could only lead to a weak improvement. As a result, this study has clearly shown the importance of data representation in increasing the performance of the crime hotspot prediction models.

The autoregressive prediction models delivered differential performances across the crime types. While the daily theft models achieved MHRs as high as 0.776, the highest reported MHR amongst the robbery models is 0.562. Likewise, 0.596 is the MHR of the best theft shift model whereas it is 0.584 for the best robbery shift model. The difference

between the MHR values particularly in the daily models may accrue from the number of crime incidents within the selected period. While there were 22596 theft incidents in the center side of Chicago between 2016 and 2018, the number of robbery incidents was much lower (n=1555), making it an even more sparse event dataset. Given the significant impact of sample sizes on train and test accuracies, we can argue that theft models exploiting larger datasets produce more accurate results (Ng, Minasny, Mendes, & Demattê, 2020). For comparative purposes, we do not have any direct baseline study that used the same sample as the current study. The only study that can give a sense of how well our models performed is the study of Zhang & Cheng (2020) (GLDNet) where they developed network-based DL models for predictive crime hotspot mapping. Included in the selected crime types is the theft where they reported an MHR of 0.635 at 20% coverage of the street network in another side of Chicago (i.e., SS). In a smaller land coverage, our autoregressive model achieved an MHR of 0.8. In sum, the selected DL algorithms can be argued to have delivered comparable results with the existing methods.

Achieving noticeably higher MHRs by incorporating our novel environmental risk surface (i.e., SSRS) is the most important contribution of the current study. The SSRS captures the dynamic environmental crime risk levels of different types during day and week by aggregating the changing local effects of residing urban features residing in the street segments. Including SRSS improved the performances of the DL algorithms that produce daily and shift autoregressive robbery models across the selected algorithms by 3.6% and 7.9% respectively on average. For the daily and shift theft models, these values are 2.1% and 14.5% respectively. To highlight the effectiveness of the SSRS, we used a static indicator that only counts the selected urban features on a street segment.

Except for the robbery shift model, including this indicator led to a decrease in the performance of autoregressive models up to 50%. This indicates that what determines the crime risk score in a street segment is the historical crime exposure level, rather than the mere existence of the selected features. This conclusion lends support to a dynamic interpretation of the CPT (Brantingham & Brantingham, 1981; Brantingham & Brantingham, 1995) that points to the daily and weekly fluctuations in the spatial influence of urban features on crime (Haberman&Ratcliffe,2015; Irvin-Erickson & La Vigne, 2015) On the contrary, assuming a static spatial influence of urban features on different crime types has a negative impact on the performance of the autoregressive hotspot prediction models. A possible reason may be that an unchanging environmental crime risk in a street segment may make the models blind to the displacements of different crime risks that recur on a daily or weekly basis. In their vulnerability and exposure framework, Kennedy, Caplan, Piza, & Buccine-Schraeder (2016) posited that the existence of a criminogenic urban feature is a source of crime vulnerability in an area. However, its mere existence does not suffice to cause higher crime rates. In fact, this vulnerability should be supplemented with an indicator of exposure that continuously updates the vulnerability based on the past crime levels. For example, although pubs are known to make a place more vulnerable to crime, there could be some pubs that have not been exposed to any crime incidents before. In that case, we can not assume a spatial influence of that pub on crime in near areas. By excluding the insignificant crime clusters around the urban features in the first step of developing DRSF, we ensured the selection of vulnerable urban features that have a significant spatial influence on different crime types during a week and day. We were thus able to reduce the false-positive crime risks that hurt the predictive performance of the selected

algorithms. In short, our novel SSRS achieved to improve upon the base model by capturing the dynamic environmental crime risk in street segments.

Including other crimes proved effective in improving the performances of the autoregressive models in most of the models. Having created an auxiliary theft event surface for robbery hotspot prediction and vice versa, we obtained greater MHRs in both daily and shift models. Its contribution is much more evident in shift prediction models. For instance, it improved the performance of GraphWavenet in the robbery and theft models by 18.6% and 31.5% respectively. A similar result was previously reported by Mohler (2014) who applied a marked point process to the violent crimes in Chicago. By including other crime types (e.g., robbery, assault), he managed to increase the accuracy of daily homicide predictions by up to 33%. The improved predictive accuracy can be explained by distinctive yet interacting spatiotemporal crime patterns of different types (Grubesic & Mack, 2008; He et al., 2020). For the current sample, the developed DL algorithms might have captured spatiotemporal associations where theft incidents trigger robbery in near areas or vice versa. All these spatiotemporal associations imply a joint opportunity field where different crime opportunities dynamically change in time. For example, pubs demonstrate a fertile ground for thieves in early evening hours when people rush into those places in large groups after a workday. After midnight, they become the places conducive to robbery when the number of intoxicated patrons who can not defend themselves properly increases in an area. Therefore, theft incidents that occurred in the second shift (i.e., 16:00-23:59) could usher robbery incidents in the following shift (i.e., 00:00-07:59). In sum, the predictive value of spatiotemporal interactions between different crime types manifested itself in the increased predictive accuracies in the developed models.

One striking result we'd like to note here is the contradicting contribution of incorporating park events to the shift models using graph-based algorithms. While it substantially improved the performance of the autoregressive theft shift model using Graph Wavenet with a 68% increase, we observed a significant drop in the performance of the robbery shift model with a 19% decrease in MHR. Although park events provided a similar contribution to the daily models across crime types (i.e., 5% and 4% respectively), the contrasting performances in the shift models invoke further explanation. One possible explanation here is that public events are makeshift crime attractors that gather large people groups within a limited space. The increased people density in these areas, like public transit stations in the rush hours, create an environment where people become pretty much distracted while trying to find their ways to their final destinations in the middle of a moving crowd(Song et al., 2019). Given the number of unattended belongings of the distracted people increase, these events may be a gold mine for the theft offenders. But, the same can not be said for the robbers. Since a typical robbery involves a physical struggle between a target and offender, the chances are much higher for intervention from formal or informal guardians (i.e., attendants, security) that deter the robber from offending. Additionally, these public events are the events where fortified security measures are mostly in effect. The increased presence of the law enforcement agents in these areas is thus another preventive mechanism that "hardens the target" (Clarke, 1995). This explanation may also serve to explain why including park events lead to a performance decrease in the robbery models. Accordingly, elevating the robbery risk in the street segments hosting park events might have resulted in inflated scores for these areas where robbers would not normally prefer to offend. Placing a risk on these areas seriously causes risk false positive predictions,

thus decreasing the accuracies. One important conclusion that can be drawn from the inconsistent impact of park events on crime risk predictions would be that one needs to be careful of selecting the theory-based indicators to be used as covariates in the crime hotspot prediction models. One covariate that proved to be extremely predictive of the risk of one crime type may remain completely irrelevant to another. Therefore, a crime-specific approach should be adopted to select the auxiliary event surfaces to enhance the crime hotspot prediction models.

The event surface of 311 calls achieved moderate improvements in the performances of daily and shift models. We noted its greatest contribution in theft shift models. There are several studies in the previous research that lends empirical support to its predictive value in crime prediction (e.g., Duan, Hu, Cheng, Zhu, & Gao, 2017). What we captured here is one of the urban pulses that function in various domains. The benefit of this pulse comes from its ability to send early warning signals about the looming crime risk in future time steps (Groff & La Vigne, 2002). In this study, we aggregated five different types of 311 calls in different domains under a single category due to the data sparsity we face in some of the categories. Also included in these types are the calls related to street lights complaints. There exists a large volume of research that has highlighted the predictive value of street lights in predicting future crime incidents in an area (e.g., Xu, Fu, Kennedy, Jiang, & Owusu-Agyemang, 2018; Chalfin, Kaplan, & LaForest, 2020; Bappee, Petry, Soares, & Matwin, 2021). Combining our results with the ones that are listed above, we can argue that the urban indicators that were obtained from a Broken Windows Theory (Wilson & Kelling,1982) perspective may serve to the predictive crime hotspot mapping in that they provide important physical cues regarding the overall crime risk level in an area. As a result, the indicators that capture various facets of urban

life may be of great predictive value for the methods that aim to predict the location of crime incidents in the next time step.

In general, this study has made an important contribution not only to crime hotspot prediction methods but also to the understanding of how differential factors play out to form these hotspots. For the former, it clearly shows the superiority of graph-based algorithms that incorporate the spatial dependence between the prediction units into the crime prediction models. Another important contribution to the former is to show how important it is to represent the datasets accurately. Through a multistep parameter tuning procedure, it showed that the performances of the selected algorithms can be greatly improved by feeding them with the datasets with the optimal data representation. In that sense, determining suitable values for the parameters that were emphasized in the procedure of Zhang & Cheng (2020) that reformulated crime hotspot prediction as a graph signal processing task is of paramount importance. For the latter, it showed that how these factors contribute to changes based on the crime type, the horizon of the crime hotspot predictions, and the selected algorithm. For instance, our novel SSRS attained an average improvement of 14% in theft shift models, but its contribution to daily predictions remained at 2.1%. Similarly, while park events largely contributed to the performances of theft predictions generated by graph-learning algorithms, its impact on the performance of the same group of algorithms was found to be negative. In short, this study developed various theory-driven network-based predictive crime hotspot mapping methods using different DL algorithms. The results are quite promising as evident from the improved predictive performances of the models.

CHAPTER 6

CONCLUSION

In this study, we developed a dynamic predictive policing system in two distinctive yet interacting parts. The first part focuses on developing a method that captures the dynamic environmental crime risk across street networks with novel DRSFs. For this, we first introduce a crime-specific spatiotemporal approach that conceptualizes urban features as base stations broadcasting signal-like spatial influence on crime. We tested this approach on a Chicago sample that includes all the theft (n=64024) and robbery (n=9685) incidents in 2018. The methodology starts with creating 54 spatiotemporal models where we examine the spatial influence of the selected urban features on robbery and theft incidents at various spatial (i.e., sides of Chicago) and temporal (default, intraday, and weekly models) configurations. Next, we fitted distance-aware risk signal functions with a segmented regression technique to characterize this spatial influence within a spatial extent after eliminating insignificant spatiotemporal crime clusters through a bivariate network K analysis. Two new scores, RSIS and RSSS, were developed from these functions to quantify the intensity and strength of the hypothesized spatial influence signals. The major findings of the spatiotemporal influence analysis showed that spatial influence demonstrates significant fluctuations across space, time, and crime types. For example, we discovered potential dead zones, where we observed no spatiotemporal influence of urban features on certain crime types. Another important result revealed that the type of urban feature and spatiotemporal context are key determinants of the distance decay effect. On the CS and FNS, for example, bus stops had a distance decay effect on robbery and theft across all temporal levels, but not on the other sides. According to our findings, the direction of the association between spatial

influence and concentrated disadvantage may be affected by the type of urban features. Also, we noted significant temporal fluctuations in the associations between spatial influence and CD.

The second part proposes a theory-driven predictive crime hotspot mapping method using graph-based DL algorithms. The novelty of this part lies in the incorporation of many theory-based auxiliary event surfaces into the DL algorithms that have not been applied to the predictive crime hotspot mapping domain before. These surfaces were found to improve upon the performance of the autoregressive models that solely use retrospective crime data to generate crime risk predictions. One of the developed event surfaces is the environmental crime risk surface that captures the dynamic environmental crime risk through a novel indicator, SSRS. This indicator is derived from the DRSFs that we developed in the first part. To test the contribution of our theory-based event surface, we created another static surface that represents the counts of the selected urban features in the street segments. The other auxiliary event surfaces include park events, other crimes, and 311 calls. We used a sample from the center side of Chicago that includes all the robbery (n=1555) and theft (n=22596) incidents between 2016 and 2018. Following the procedure described by Zhang & Cheng (2020) that translates predictive crime hotspot mapping into a graph signal processing task, we performed many data preprocessing steps: create a graph whose nodes are street segments, smooth the crime counts to convert sparse event time series into a format suitable for deep learning, weight the edges between the nodes (i.e., street segments) by using a gaussian kernel function. In addition to the procedure, we described another two-step procedure where we gradually optimized the parameters related to data representation and model learning. In the first step, we experimented with a set of values for each parameter. After ensuring

the data is represented in an optimal manner by experimenting with various training windows, smoothing coefficient, and spatial bandwidth values, we performed model parameter tuning by experimenting with a set of learning rate values. In both steps, we measure the performance of the models based on MHR that refers to how much of the crime incidents on average are captured in the predicted crime hotspots by the model throughout the test days. An important result indicated the differential parameters across crime types as well as the prediction horizons (i.e., daily, shift). This simply underscores the necessity of a data-driven approach in determining the optimal data representation. Another important result illustrated the efficiency of the proposed auxiliary surfaces in improving the performance of crime prediction models. Amongst these surfaces, SSRS, other crime, and park events were the surfaces that stood out due to achieving the highest predictive values for the autoregressive model. Another surface, park events, has clearly shown the necessity of a feature selection process driven by crime theories. This is because of the inconsistent performance that the shift models with park events delivered across the crime types. While it substantially improves the predictive performance of the theft shift models, it was found to have an adverse impact on the robbery shift models. In short, we developed a theory-driven predictive policing system in this study. The results we obtained were quite promising in terms of enhancing the existing methods that are heavily reliant upon historical data. In that sense, the developed system caters to law enforcement agencies that both strive to understand the spatiotemporal dynamics of a crime risk and to make accurate crime risk predictions accordingly for increasing public safety as well as achieving operational efficiency.

6.1. Research implications

The first part of the study has several research implications. For example, it offered solid empirical support for the requirement of spatial influence analysis with a crime-specific spatiotemporal approach, as implied by the concept of environmental backcloth (Brantingham & Brantingham, 1981; Brantingham & Brantingham, 1995). Likewise, an urban feature may be a significant crime attractor or generator in one context for one type, but it can be completely irrelevant for another type in another context. Furthermore, the results demonstrated that the type of urban features had the greatest impact on the distance decay effect. We discovered dynamic spatial diffusion effects for urban features on Chicago's four sides. This conclusion was consistent with Feng, Piza, Kennedy, & Caplan (2019) who similarly found changing spatial influence levels of urban features in New York City's five boroughs. These findings emphasize the need of using a dynamic spatiotemporal approach to operationalize spatial influence within a spatial extent. Lastly, our research found that CD affected spatial influence only for particular urban features in some intervals.

We can also articulate many research implications for the second part of the current study. First and foremost, it suggests that univariate crime hotspot prediction methods may be enhanced by the inclusion of theory-based auxiliary event surfaces that show significant interactions with the crime risk surface. One important point here emphasizes a careful selection as not all kinds of surfaces may interact with every crime type. Putting the datasets into the analysis just because of their availability rather than seeking a theoretical rationale may lead to deteriorated prediction performances. Therefore, selecting the surfaces should be established upon a theoretical foundation. Failure to do so may result in significant performance reductions in the prediction model due to the

inflated crime risks that cause false-positive predictions. Another research implication is the superiority of the selected graph-based deep learning algorithms, Graph Wavenet and STGCN, over their traditional counterpart, LSTM in predicting crime hotspots. The key difference between these methods is the fully convolutional units that the first group of algorithms utilizes to model temporal dependence. Using convolutional units, rather than recurrent units for that purpose yielded better results. Another important point places the emphasis on spatial dependence. The ability to model spatial dependence through the adjacency matrices is what puts the first group one step ahead of the second group. A further division between the first group algorithms concerns the way they incorporate spatial dependence into the algorithm. While Graph Wavenet used a self-adaptive adjacency matrix, STGCN used the weights as is. The result supported the former approach that eliminates the weak connections between the nodes. One implication would be to use lower-order neighbors in modeling spatial dependence as including higher-order neighbors distorts the training weights, hence the performance. As a result, this study emphasized the efficiency of convolution-based graph learning algorithms sparse time series such as crime. Fourth, the proposed representation tuning underlines the importance of a data-driven approach in crime hotspot prediction models. Given unique crime datasets as a result of unique interactions between crime and urban configurations, it is of crucial importance to ensure optimal data representation. Only with optimally represented data will it be possible to obtain models that produce the most accurate predictions. One last research implication is that this study has shown the applicability of many graph signal processing algorithms that were originally developed to forecast traffic speed across street networks into crime hotspot prediction context. After translating a sparse crime count dataset into a continuous crime risk dataset, the

researchers can harness any network-based DL algorithms to produce spatiotemporal crime predictions.

6.2. Practical implications

For law enforcement agents, the current study entails a wealth of information. First, depending on the situation, a preliminary spatiotemporal clustering method may help organize patrol routes by allowing the prioritization of the routes with the highest crime risk. Second, by modeling cumulative crime densities within a spatial extent, the DRSFs allow for the quantification and comparison of the spatial influence of urban features on different crime types within any spatial extent across selected spatiotemporal units. Third, when combined at the street segment level, RSIS and RSSS values can operationalize the whole environmental crime risk across a street network. Various spatiotemporal crime hotspot maps on street networks could be created using aggregated statistics. By alerting decision-makers about optimal patrol routes for any given period across the regions of a city, it enables for more targeted interventions and efficient use of resources. Furthermore, a more in-depth examination of these sites that are frequently ranked at the top by RSIS and RSSS may reveal further information about these criminogenic environments. These values may lead to safety improvements through environmental modifications such as streetlights (Xu, Fu, Kennedy, Jiang, & Owusu-Agyemang, 2018) or more CCTV cameras as possible crime prevention through environmental design (CPTED) (Jeffery, 1977) application (Corcoran, Zahnow, Kimpton, Wickes, & Brunsdon, 2021).

For the second part, including theory-based surfaces improved the performance of the predictive crime hotspot mapping algorithms up to 68%. This means that using a multilayered predictive crime hotspot mapping technique enhanced by auxiliary event

surfaces warrants a great increase in the operational efficiency by predicting the future crime hotspots more accurately. The increasing availability of the urban datasets brings new opportunities to further improve the univariate crime hotspot prediction methods. The police departments should therefore embark upon developing a predictive policing system fed by various urban datasets. Grounded in theory, these systems go beyond the predictive limitations of the existing retrospective techniques. Another practical implication would be to create crime-specific predictive crime hotspot maps on a daily or shift basis to be complemented with different sets of features. Monitoring the crime risk across street segments with as many datasets as possible can capture the crime risk from many aspects. Therefore, it better characterizes the crime risk across street networks. With a reference to the first part, crime hotspot maps should consider regional differences in terms of spatial influence. To know which types of urban features are exposed to which types of crime risk during which periods across the regions is of great practical value. Limiting predictive crime hotspot maps to the regional level may provide more manageable patrol routes than the ones obtained from the citywide maps that may be located too scattered across the city.

6.3. Limitations

This study is not without its limitations. First and foremost, the accuracy of risk signal functions was strongly dependent on the geographical and temporal precision of crime data, as lower precision levels could threaten the estimated functions' validity. Second, the modifiable areal unit problem (Wong, 2004) and the modifiable temporal unit problem (Cheng & Adepeju, 2014) both represent significant challenges to the analysis' validity. Third, the computational complexity of network K analysis, as well as the exponentially growing number of space x time x crime x urban feature configurations,

impeded the use of spatiotemporal units at higher resolution levels. Fourth, the list of urban features included in the spatiotemporal analysis might be expanded to include more features. For example, Fox, Trolard, Simmons, Meyers, & Vogel (2021)identified vacancy as a major contributor to violent crime on both the north and south sides of St. Louis, Missouri. Finally, the sample is limited to Chicago, IL, which confines the study's generalizability to other cities. As a result, we strongly encourage replication studies in different urban settings. The biggest challenge for the second part is the computational complexity of the selected graph-based deep learning algorithms. Due to exponentially growing complexity, we were unable to implement citywide crime prediction models. Furthermore, the growing complexity in parallel with the number of nodes also impeded all the models at the side level, except the center side. We, therefore, had to apply our predictive models to a sample obtained from the center side. A similar issue ensues when we attempt to apply various popular RNN-based graph learning algorithms (e.g., DCRNN) to our datasets. Each attempt produced an out-of-memory error due to intractable complexity. For this reason, we had to utilize convolution-based graph learning algorithms that are far less memory demanding than their RNN-based counterparts (Wu, Pan, Long, Jiang, & Zhang, 2019). Another limitation for the second part is the number of the selected crime types. Since an additional crime type could mean an additional twelve models each taking days to fine-tune, we decided to limit the crime types to theft and robbery. The number of model-related parameters we experimented is another limitation of our study. The exponentially growing number of model configurations in our analysis is the main reason why we did not choose to experiment with additional parameters such as dropout rate, batch size, etc. Lastly, the sample is only limited to the center side of Chicago. Therefore, the generalizability of

our dynamic predictive policing system to other regional or even city contexts is unknown.

6.4. Future research

Although we used appropriate spatial and temporal units to overcome data sparsity (Bernasco, Ruiter, & Block, 2017)and computational intractability of the K-function (He et al., 2020), a study with a finer spatiotemporal granularity level could still reveal interesting relationships and provide useful insights for future research. Examining different versions of the proposed risk signal function is another possible research direction. For example, rather than a linear function like the one used in this work, a non-linear function could better capture a curvilinear spatial influence within a spatial extent. Researchers can now assess the level of human activity along streets with better accuracy because of the increased availability of GPS data. An investigation based on this metric would also provide interesting details regarding the interaction of spatial influence and ambient population in terms of crime risk across streets. For the second part, the primary research direction would be an endeavor to examine further auxiliary risk surfaces that contribute to the performances of the predictive crime hotspot mapping algorithms. A secondary direction involves examining the adaptability of various graph-learning algorithms into the task of crime hotspot prediction. A third direction would be to apply the developed predictive policing algorithms to other crime types with different sets of auxiliary event surfaces. One last possible direction would be to use even shorter prediction horizons due to the large benefits of specificity of the hotspot predictions in a timely manner.

APPENDIX A.

THE SPATIOTEMPORAL NETWORK K RESULTS MATRIX

		Robbery						Theft					
	Sides	**Default**	**WD**	**WE**	**FS**	**SS**	**TS**	**Default**	**WD**	**WE**	**FS**	**SS**	**TS**
Bus Stop	***C***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FSE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FSW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	N	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	S	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	SW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	W	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fast-food Restaurant	***C***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FSE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FSW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	N	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	S	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	SW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	W	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gas Station	***C***	x	✓	x	x	x	x	x	x	x	x	x	x
	FN	✓	✓	✓	✓	x	x	✓	✓	✓	✓	✓	✓
	FSE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FSW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	N	✓	x	x	x	x	x	✓	x	✓	✓	x	x
	NW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	S	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	SW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	W	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grocery Store	***C***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FSE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FSW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	N	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NW	✓	✓	✓	x	✓	✓	✓	✓	✓	✓	✓	✓
	S	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	SW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	W	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pub	***C***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FSE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	FSW	✓	x	✓	✓	x	✓	✓	✓	✓	✓	✓	✓
	N	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	S	✓	✓	x	x	✓	✓	✓	✓	✓	✓	✓	✓
	SW	✓	✓	x	x	✓	x	✓	✓	✓	✓	✓	✓
	W	✓	x	✓	✓	x	✓	✓	✓	✓	✓	✓	✓

Note: WD= Weekday, **WE:**Weekend, **FS=** First Shift (00:00- 7:59), **SS=** Second Shift (08:00-15:59), **TS=** Third Shift (16:00-23:59), **✓ = s**ignificant relationship**, x =** non-significant relationship at α=0.05 level

APPENDIX B.

THE SPATIOTEMPORAL RSIS RESULTS MATRIX.

		Robbery						Theft					
	Sides	**Default**	**WD**	**WE**	**FS**	**SS**	**TS**	**Default**	**WD**	**WE**	**FS**	**SS**	**TS**
Bus Stop	***C***	1.74	1.81	1.73	1.80	1.94	1.69	1.62	1.63	1.58	1.56	1.62	1.62
	FN	2.21	2.43	1.83	2.45	2.07	2.12	1.77	1.76	1.78	1.69	1.92	1.64
	FSE	2.15	2.12	2.20	2.06	2.30	2.10	2.09	2.12	2.05	1.59	2.28	2.09
	FSW	2.07	2.10	2.01	2.03	1.99	16.56	1.79	1.75	1.93	1.51	1.94	13.00
	N	1.58	1.52	1.65	1.71	1.51	1.48	1.56	1.54	1.59	1.52	1.58	1.56
	NW	1.43	1.42	1.46	1.40	1.84	1.38	1.56	1.57	1.54	1.41	1.74	1.48
	S	1.38	1.35	1.37	1.37	1.38	1.34	1.39	1.39	1.39	1.30	1.43	1.38
	SW	1.42	1.42	1.48	1.38	1.49	1.49	1.51	1.49	1.56	1.25	1.62	1.51
	W	1.40	1.45	1.32	1.36	1.40	1.42	1.29	1.31	1.26	1.21	1.36	1.26
Fast-food Restaurant	***C***	2.77	2.85	2.62	2.24	3.23	2.94	2.56	2.73	2.26	1.91	2.87	2.39
	FN	3.56	3.72	3.32	2.87	4.01	3.89	3.52	3.48	3.59	3.80	3.94	2.98
	FSE	5.13	4.55	6.08	3.87	5.96	5.04	6.48	6.15	6.69	3.50	7.05	6.92
	FSW	4.07	4.49	3.99	3.79	6.38	2.96	3.05	2.97	3.06	2.25	3.19	3.31
	N	2.82	2.77	2.91	2.80	2.91	2.85	3.11	3.07	3.23	2.61	3.37	3.07
	NW	2.32	2.34	2.35	1.95	3.11	2.15	4.95	4.80	5.33	2.41	6.07	4.83
	S	2.08	1.97	2.22	2.04	1.90	2.20	2.42	2.44	2.52	1.99	2.63	2.36
	SW	2.41	2.45	2.46	2.19	2.75	2.43	3.71	3.59	3.98	1.84	4.27	3.92
	W	3.56	1.92	1.60	1.55	2.00	1.77	4.83	3.04	2.77	2.27	3.47	2.68
Gas Station	***C***	0.00	1.46	1.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	FN	1.86	1.91	1.77	2.46	0.00	0.00	1.63	1.62	1.65	1.96	1.51	1.61
	FSE	3.23	3.16	3.59	4.04	3.03	2.83	2.65	2.69	2.54	2.99	2.81	2.37
	FSW	3.76	3.70	4.25	3.33	3.56	4.40	2.66	2.65	2.67	2.64	2.76	2.52
	N	1.51	0.00	0.00	0.00	0.00	0.00	1.13	1.05	1.39	1.80	0.00	0.00
	NW	2.21	2.21	2.19	1.90	2.47	2.30	2.08	2.09	2.07	2.01	1.73	2.16
	S	1.82	1.76	1.92	2.15	1.89	1.69	1.92	1.91	1.95	1.47	2.04	1.90
	SW	2.12	1.85	2.36	1.94	2.09	2.14	2.30	2.29	2.33	2.28	2.25	2.32
	W	4.01	2.29	2.28	2.54	1.93	2.30	4.32	1.62	1.71	1.76	1.65	1.60
Grocery Store	***C***	1.74	1.75	0.00	1.81	1.80	1.67	1.81	1.84	1.84	1.85	1.88	1.71
	FN	4.28	4.39	4.07	4.22	4.80	4.03	3.22	3.12	3.40	3.08	3.43	3.05
	FSE	3.72	3.57	4.03	3.22	4.21	3.87	3.19	3.29	2.99	2.32	3.61	3.11
	FSW	4.43	4.31	4.53	4.64	4.26	4.19	3.62	3.33	4.27	1.92	4.19	3.82
	N	1.90	0.00	0.00	2.18	2.16	1.66	1.90	1.83	2.05	1.98	1.97	1.91
	NW	1.89	1.71	2.22	0.00	2.39	2.05	1.92	1.83	2.19	1.52	2.12	1.93
	S	2.17	2.22	2.05	2.11	2.41	2.05	2.13	2.11	2.19	1.88	2.23	2.12
	SW	2.42	2.44	2.41	1.91	2.54	2.60	2.21	2.11	2.21	1.84	2.31	2.15
	W	5.37	1.79	1.77	1.67	1.93	1.73	5.22	2.03	2.13	1.71	2.11	2.15
Pub	***C***	2.61	2.16	3.23	4.09	1.71	1.97	2.22	1.90	2.76	4.69	1.76	2.03
	FN	2.37	2.76	1.67	2.30	3.26	2.00	2.20	2.04	2.63	2.39	2.15	2.23
	FSE	2.33	2.26	2.48	2.51	2.50	2.27	2.43	2.26	2.09	1.62	2.19	2.42
	FSW	2.20	0.00	3.50	2.27	0.00	2.83	2.04	2.00	2.15	2.45	2.03	1.86
	N	2.25	2.00	2.58	2.83	1.44	1.94	2.36	2.00	2.99	3.75	1.84	2.29
	NW	1.92	1.66	2.30	2.05	1.79	1.87	1.53	1.51	1.56	1.72	1.51	1.56
	S	1.58	1.65	0.00	0.00	1.70	1.48	1.48	1.36	1.70	1.34	1.53	1.52
	SW	1.22	1.33	0.00	0.00	1.55	0.00	1.47	1.47	1.48	1.27	1.52	1.52
	W	3.66	0.00	1.28	1.23	0.00	1.11	4.58	2.34	3.14	2.76	2.18	2.98

Note: WD= Weekday, **WE:**Weekend, **FS=** First Shift (00:00- 7:59), **SS=** Second Shift (08:00-15:59), **TS=** Third Shift (16:00-23:59)

APPENDIX C.

THE SPATIOTEMPORAL RSSS RESULTS MATRIX.

		Robbery						Theft					
	Sides	**Default**	**WD**	**WE**	**FS**	**SS**	**TS**	**Default**	**WD**	**WE**	**FS**	**SS**	**TS**
Bus Stop	**C**	-2.62	-1.22	3.41	-1.07	-20.70	2.41	-30.84	-30.4	-32.21	33.36	-32.39	-34.54
	FN	-4.66	-4.44	-5.26	-22.4	-12.34	20.02	-10.14	-6.55	-16.79	-26.61	-7.34	-4.87
	FSE	-0.46	7.13	-67.90	1.20	1.93	-3.09	20.17	27.04	7.40	38.63	18.16	18.05
	FSW	16.73	13.16	-2.97	6.22	5.05	128.43	76.50	96.62	89.63	316.88	91.26	113.40
	N	11.19	20.36	-3.95	-18.5	64.16	53.11	54.39	49.89	42.75	23.82	41.99	70.11
	NW	84.68	74.59	109.42	51.56	6.97	66.47	53.25	54.50	51.28	53.17	70.69	46.34
	S	20.46	33.65	5.01	7.55	10.43	26.56	90.14	81.51	110.47	69.86	102.87	85.03
	SW	2.64	2.33	14.15	-32.4	27.52	12.37	76.78	79.00	53.71	114.18	59.62	48.66
	W	9.24	5.47	16.69	15.41	12.57	3.71	36.81	37.29	36.17	47.36	43.92	25.46
Fast-food Restaurant	**C**	-38.22	-48.14	-36.47	-33.3	-29.65	-50.03	2.49	0.36	7.98	31.00	-3.79	-8.63
	FN	-61.38	-64.72	-52.40	-31.4	-57.54	-67.63	-41.16	-41.2	-41.31	-44.27	-43.51	-35.58
	FSE	-56.32	-55.54	-48.06	-2.09	-69.18	-42.66	-58.00	-54.3	-62.67	-11.16	-56.44	-63.04
	FSW	-89.52	-91.61	-35.55	-89.4	-91.24	-81.75	-77.55	-79.3	-75.71	-72.70	-76.12	-80.33
	N	-67.75	-71.90	-57.39	-60.6	-70.45	-72.05	-59.12	-59.3	-56.99	-53.63	-61.06	-62.41
	NW	-59.02	-64.90	-0.95	-46.8	-79.89	40.81	-75.37	-71.9	-79.16	-66.64	-74.75	-77.33
	S	-51.79	-57.37	-42.05	-72.0	1.23	-45.67	-35.22	-39.0	-33.20	-32.70	-51.69	21.65
	SW	-81.63	-80.83	-80.97	-85.0	-75.58	-80.77	-86.57	-86.5	-86.71	-76.33	-86.68	-87.87
	W	74.57	-58.19	-7.47	-46.5	-67.34	-16.09	7.60	-53.8	-84.38	-54.67	-58.38	-51.37
Gas Station	**C**	0.00	89.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	FN	-33.12	-36.37	-25.50	-26.6	0.00	0.00	-23.11	-28.7	-7.66	-36.94	-18.50	-17.84
	FSE	-51.37	-48.28	-55.81	-67.1	534.04	-56.54	-7.98	-3.75	-19.27	-53.44	12.69	-3.63
	FSW	98.39	418.20	-7.03	3.38	247.40	54.44	-10.97	-15.5	0.14	12.23	0.11	-28.61
	N	-37.71	0.00	0.00	0.00	0.00	0.00	-3.47	5.24	36.65	39.62	0.00	0.00
	NW	-9.27	116.58	-47.44	-11.1	-52.88	44.49	-22.21	-18.1	-29.51	-33.05	142.11	-29.71
	S	-12.06	70.96	-58.32	-33.0	-46.92	-56.72	-46.96	-45.7	-49.11	43.43	-45.57	-49.97
	SW	-41.10	-31.14	-56.32	-53.8	-62.11	-56.02	-68.11	-67.9	-68.41	-73.24	-61.46	-70.17
	W	10.53	-60.42	-67.64	-72.1	93.28	-61.87	5.16	-55.2	-59.49	-70.87	-52.09	-53.18
Grocery Store	**C**	-38.82	-30.43	-47.26	-34.9	-11.55	-50.02	-57.06	-54.5	-57.07	-64.00	-55.43	-35.39
	FN	-44.10	-44.09	-44.07	-42.2	-59.23	-30.88	-55.96	-54.7	-58.15	-61.85	-54.28	-55.08
	FSE	7.66	36.23	-23.35	-38.5	-16.73	46.00	2.97	-0.41	12.39	42.60	0.70	-1.86
	FSW	-35.46	-43.76	2.67	3.42	-12.62	-45.93	-78.20	-73.7	-83.50	-55.05	-78.63	-81.33
	N	-35.67	-36.30	-42.33	-13.9	-78.56	-67.46	-50.99	-48.9	-53.94	-54.75	-43.71	-63.92
	NW	-47.19	-44.30	-52.40	0.00	-53.02	-45.29	-74.30	-75.5	-66.95	-69.98	-74.77	-74.21
	S	-55.03	-52.92	-56.25	-57.6	-58.44	-50.46	-64.66	-65.5	-59.17	-43.02	-68.37	-65.18
	SW	-35.86	-46.74	-11.97	-15.9	-46.97	-37.08	-69.82	-61.3	-71.04	-52.61	-70.72	-63.83
	W	129.55	-33.88	-50.45	-20.0	-47.05	-42.43	23.75	-64.9	-62.82	-38.87	-69.00	-63.67
Pub	**C**	-45.66	-22.87	-53.52	-52.0	-41.13	-32.50	-64.51	-60.4	-66.64	-68.30	-61.48	-61.47
	FN	-38.69	-40.99	-30.96	-46.7	-32.97	-27.13	-41.93	-32.5	-57.66	-65.24	-30.36	-40.95
	FSE	8.73	91.32	-37.52	-40.0	677.99	-23.79	36.42	10.31	105.73	251.67	119.07	5.60
	FSW	-32.01	0.00	-22.26	-71.8	0.00	89.16	-51.35	-35.8	-69.66	-73.16	-21.81	-33.89
	N	-36.62	-9.79	-57.76	-50.4	-54.80	-16.92	-51.75	-43.1	-60.49	-67.54	-34.46	-47.46
	NW	-65.86	-71.38	-56.39	-71.8	43.83	-73.14	21.96	34.09	1.73	-42.16	8.97	120.68
	S	129.35	112.56	0.00	0.00	30.94	304.44	-27.92	11.37	-57.40	-35.98	-17.83	-36.72
	SW	-9.44	-35.20	0.00	0.00	-30.61	0.00	-81.58	-81.0	-82.51	-85.09	-80.62	-80.81
	W	188.85	0.00	-57.35	-18.9	0.00	-38.86	38.08	-35.2	-63.97	-66.79	-35.91	-50.76

Note: WD= Weekday, **WE:**Weekend, **FS=** First Shift (00:00- 7:59), **SS=** Second Shift (08:00-15:59), **TS=** Third Shift (16:00-23:59)

REFERENCES

About Chicago: Facts and Statistics. (2021). Retrieved October 15, 2021, from https://www.chicago.gov/city/en/about/facts.html

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social lstm: Human trajectory prediction in crowded spaces. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 961-971).

Andresen, M. A., & Malleson, N. (2015). Intra-week spatial-temporal patterns of crime. *Crime Science*, 1-11. doi:https://doi.org/10.1186/s40163-015-0024-7

Andresen, M. A., Curman, A. S., & Linning, S. J. (2017). The trajectories of crime at places: Understanding the patterns of disaggregated crime types. *Journal Of Quantitative Criminology, 33*(3), 427-449. doi:https://doi.org/10.1007/s10940-016-9301-1

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis, 27*(2), 93-115. doi:https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Baddeley, A., & Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software, 12*(1), 1-42. doi:10.18637/jss.v012.i06

Baddeley, A., Diggle, P. J., Hardegen, A., Lawrence, T., Milne, R. K., & Nair, G. (2014). On tests of spatial pattern based on simulation envelopes. *Ecological Monographs, 84*(3), 477-489. doi:https://doi.org/10.1890/13-2042.1

Bappee, F. K., Petry, L. M., Soares, A., & Matwin, S. (2021). Analyzing the Impact of Foursquare and Streetlight Data with Human Demographics on Future Crime Prediction. In R. Stahlbock, G. Weiss, M. Abou-Nasr, C. Yang, H. Arabnia, & D. Deligiannidis, *Advances in Data Science and Information Engineering* (pp. 435-449). Dordrecht: Springer.

Barnum, J. D., Caplan, J. M., Kennedy, L. W., & Piza, E. L. (2017). The crime kaleidoscope: A cross-jurisdictional analysis of place features and crime in three urban environments. *Applied Geography, 79*, 203-211. doi:https://doi.org/10.1016/j.apgeog.2016.12.011

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In G. B. Orr, *Neural networks: Tricks of the trad* (Vol. 7700, pp. 437-478). Berlin: Springer.

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society, 28*(7), 806-822. doi:https://doi.org/10.1080/10439463.2016.1253695

Bernasco, W. (2010). Modeling micro-level crime location choice: Application of the discrete choice framework to crime at places. *Journal of Quantitative Criminology, 26*(1), 113-138. doi:https://doi.org/10.1007/s10940-009-9086-6

Bernasco, W., & Block, R. (2009). Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology, 47*(1), 93-130. doi: https://doi.org/10.1111/j.1745-9125.2009.00140.x

Bernasco, W., & Block, R. (2011). Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency, 48*(1), 33-57. doi:https://doi.org/10.1177/0022427810384135

Bernasco, W., & Nieuwbeerta, P. (2005). How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *British Journal of Criminology, 45*(3), 296-315. doi:https://doi.org/10.1093/bjc/azh070

Bernasco, W., Ruiter, S., & Block, R. (2017). Do street robbery location choices vary over time of day or day of week? A test in Chicago. *Journal of Research in Crime and Delinquency, 54*(2), 244-275. doi:https://doi.org/10.1177/0022427816680681

Block, R. (1993). *Street Gang Crime in Chicago.* Washington, DC: National Institute of Justice.

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014). Once upon a crime: towards crime prediction from demographics and mobile data. *Proceedings of the 16th international conference on multimodal interaction* (pp. 427-434). Istanbul : ICMI.

Bowers, K. J., & Johnson, S. D. (2003). Measuring the geographical displacement and diffusion of benefit effects of crime prevention activity. *Journal of Quantitative Criminology, 19*(3), 275-301. doi:https://doi.org/10.1023/A:1024909009240

Bowers, K. J., Johnson, S. D., & Pease, K. (2004). Prospective hot-spotting: the future of crime mapping? *British Journal of Criminology, 44*(5), 641-658. doi:https://doi.org/10.1093/bjc/azh036

Braga, A., Papachristos, A., & Hureau, D. (2012). Hot spots policing effects on crime. *Campbell Systematic Reviews, 8*(1), 1-96. doi: https://doi.org/10.4073/csr.2012.8

Brantingham, P., & Brantingham, P. (1981). *Environmental Criminology.* Beverly Hills, CA: Sage Publications.

Brantingham, P., & Brantingham, P. (1995). Criminality of place. *European Journal on Criminal Policy and Research, 3*, 5-26. doi:https://doi.org/10.1007/BF02242925

Breetzke, G. D., & Edelstein, I. S. (2020). Do crime generators exist in a developing context? An exploratory study in the township of Khayelitsha, South Africa. *Security Journal*, 1-18. doi:https://doi.org/10.1057/s41284-020-00264-0

Browning, C. R., Byron, R. A., Calder, C. A., Krivo, L. J., Kwan, M. P., Lee, J. Y., & Peterson, R. D. (2010). Commercial density, residential concentration, and crime: Land use patterns and violence in neighborhood context. *Journal of Research in Crime and Delinquency, 47*(3), 329-357. doi:https://doi.org/10.1177/0022427810365906

Browning, M., & Arrigo, B. (2021). Stop and Risk: Policing, Data, and the Digital Age of Discrimination. *American Journal of Criminal Justice, 46*(2), 298-316. doi:https://doi.org/10.1007/s12103-020-09557-x

Caplan, J. M. (2011). Mapping the spatial influence of crime correlates: A comparison of operationalization schemes and implications for crime analysis and criminal justice practice. *Cityscape, 13*(3), 57-83. doi:http://www.jstor.org/stable/41426675.

Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting. *Justice quarterly, 28*(2), 360-381. doi:https://doi.org/10.1080/07418825.2010.486037

Cavadas, B., Branco, P., & Pereira, S. (2015). Crime prediction using regression and resources optimization. *Portuguese Conference on Artificial Intelligence* (pp. 513-524). Coimbra: Springer.

Cerdá, M., Tracy, M., Messner, S. F., Vlahov, D., Tardiff, K., & Galea, S. (2009). Misdemeanor policing, physical disorder, and gun-related homicide: a spatial

analytic test of" broken-windows" theory. *Epidemiology*, 533-541. doi:http://www.jstor.org/stable/25662699.

Chalfin, A., Kaplan, J., & LaForest, M. (2020). *Street light outages, public safety and crime displacement: Evidence from Chicago.* Pennsylvania State University. Retrieved from https://poseidon01.ssrn.com/delivery.php?ID=573003097124108026102019029101077102004031054052030066103124099126127084090025071007006034063101028059032067028103007126071100046016071077042067002092086078113072096068080057089024002087067083001086127071123081 0

Chen, X., Cho, Y., & Jang, S. Y. (2015). Crime prediction using Twitter sentiment and weather. *2015 Systems and Information Engineering Design Symposium* (pp. 63-68). Charlotesville,VA: IEEE. doi:10.1109/SIEDS.2015.7117012

Cheng, T., & Adepeju, M. (2014). Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PloS One, 94*(5), e100465. doi:10.1371/journal.pone.0100465

Clarke, R. V. (1995). Situational crime prevention. *Crime and Justice, 19*, 91-150. doi:https://doi.org/10.1086/449230

Cockrell, J. (2021, March 1). *Law and order and data*. Retrieved September 27, 2021, from Chicago Booth: https://review.chicagobooth.edu/economics/2021/article/law-and-order-and-data

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44*(4), 588-608. doi:https://doi.org/10.2307/2094589

Corcoran, J. J., Wilson, I. D., & Ware, J. A. (2003). Predicting the geo-temporal variations of crime and disorder. *International Journal of Forecasting, 19*(4), 623-634. doi:https://doi.org/10.1016/S0169-2070(03)00095-5

Corcoran, J., Zahnow, R., Kimpton, A., Wickes, R., & Brunsdon, C. (2021). The temporality of place: Constructing a temporal typology of crime in commercial precincts. *Environment and Planning B: Urban Analytics and City Science, 48*(1), 9-24. doi:https://doi.org/10.1177/2399808319846904

Cornish, D. B., & Clarke, R. V. (1987). Understanding crime displacement: An application of rational choice theory. *Criminology*, *25*(4), 933-948. doi: https://doi.org/10.1111/j.1745-9125.1987.tb00826.x

Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. (2020). A framework for estimating crime location choice based on awareness space. *Crime Science, 9*(1), 1-14. doi:https://doi.org/10.1186/s40163-020-00132-7

Dauphin, Y. N., Fan, A., Auli, M., & Grangier, D. (2016). Language modeling with gated convolutional networks. *International Conference on Machine Learning* (pp. 933-941). Anaheim,CA: PMLR. Retrieved from https://proceedings.mlr.press/v70/dauphin17a.html.

de Melo, S. N., Pereira, D. V., Andresen, M. A., & Matias, L. F. (2018). Spatial/temporal variations of crime: A routine activity theory perspective. *International Journal of Offender Therapy and Comparative Criminology, 62*(7), 1967-1991. doi:https://doi.org/10.1177/0306624X17703654

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *NIPS* (pp. 1-9).

Barcelona: 30th Conference on Neural Information Processing Systems (. Retrieved from https://arxiv.org/pdf/1606.09375v3.pdf

Demeau, E., & Parent, G. (2018). Impacts of crime attractors and generators on criminality in Montreal. *Canadian Journal of Criminology and Criminal Justice, 60*(3), 387-412. doi:https://doi.org/10.3138/cjccj.2017-0028.r1

Deryol, R., Wilcox, P., Logan, M., & Wooldredge, J. (2016). Crime places in context: An illustration of the multilevel nature of hot spot development. *Journal of Quantitative Criminology, 32*(2), 305-325. doi:https://doi.org/10.1007/s10940-015-9278-1

Diener, M. (2015). *Python Geospatial Analysis Cookbook.* Birmingham: Packt Publishing Ltd.

Drawve, G., Thomas, S. A., & Walker, J. T. (2016). Bringing the physical environment back into neighborhood research: The utility of RTM for developing an aggregate neighborhood risk of crime measure. *Journal of Criminal Justice, 44*, 21-29. doi:https://doi.org/10.1016/j.jcrimjus.2015.12.002

Duan, L., Hu, T., Cheng, E., Zhu, J., & Gao, C. (2017). Deep Convolutional Neural Networks for Spatiotemporal Crime Prediction. *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)* (pp. 61-67). Las Vegas,NA: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). Retrieved from https://csce.ucmss.com/cr/books/2017/LFS/CSREA2017/IKE3704.pdf

Eck, J. E., Clarke, R. V., & Guerette, R. T. (2007). Risky facilities: Crime concentration in homogeneous sets of establishments and facilities. *Crime prevention studies, 21*, 225.

Feng, S. Q., Piza, E. L., Kennedy, L. W., & Caplan, J. M. (2019). Aggravating effects of alcohol outlet types on street robbery and aggravated assault in New York City. *Journal of Crime and Justice, 42*(3), 257-273. doi:https://doi.org/10.1080/0735648X.2018.1559076

Fionda, V., & Palopoli, L. (2011). Biological network querying techniques: analysis and comparison. *Journal of Computational Biology, 18*(4), 595-625. doi:https://doi.org/10.1089/cmb.2009.0144

Fox, B., Trolard, A., Simmons, M., Meyers, J. E., & Vogel, M. (2021). Assessing the Differential Impact of Vacancy on Criminal Violence in the City of St. Louis, MO. *Criminal Justice Review, 46*(2), 156-172. doi:https://doi.org/10.1177/0734016821996795

Garnier, S., Caplan, J. M., & Kennedy, L. W. (2018). Predicting dynamical crime distribution from environmental and social influences. *Frontiers in Applied Mathematics and Statistics, 4*, 13. doi: https://doi.org/10.3389/fams.2018.00013

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems, 61*, 115-125. doi:https://doi.org/10.1016/j.dss.2014.02.003

Geyer, F. (2017). Performance evaluation of network topologies using graph-based deep learning. *Proceedings of the 11th EAI International Conference on Performance Evaluation Methodologies and Tools* (pp. 20-27). Venice: ACM.

Graves, A., Jaitly, N., & Mohamed, A. R. (2013). Hybrid speech recognition with deep bidirectional LSTM. *IEEE workshop on automatic speech recognition and understanding* (pp. 273-278). Olomouc: IEEE. doi:10.1109/ASRU.2013.6707742

Groff, E. (2011). Exploring 'near': Characterizing the spatial extent of drinking place influence on crime. *Australian & New Zealand Journal of Criminology, 44*(2), 156-179. doi:https://doi.org/10.1177/0004865811405253

Groff, E. (2013). Measuring a place's exposure to facilities using geoprocessing models: an illustration using drinking places and crime. In M. Leitner, *Crime modeling and mapping using geospatial technologies* (pp. 269–295). Dordrecht: Springer.

Groff, E. R. (2017). Measuring the influence of the built environment on crime at street segments. *Jerusalem Review of Legal Studies, 15*(1), 44-54. doi:https://doi.org/10.1093/jrls/jlx005

Groff, E. R., & La Vigne, N. G. (2002). Forecasting the future of predictive crime mapping. *Crime Prevention Studies, 12*, 29-58.

Groff, E. R., & Lockwood, B. (2014). Criminogenic facilities and crime across street segments in Philadelphia: Uncovering evidence about the spatial extent of facility influence. *Journal of Research in Crime and Delinquency, 51*(3), 277-314. doi:https://doi.org/10.1177/0022427813512494

Grossmann, L., Thompson, M., Kluger, J., Park, A., Walsh, B., Suddath, C., . . . Carbone, N. (2011, November 28). *The 50 Best Inventions*. Retrieved September 2, 2021, from Time: http://content.time.com/time/subscriber/article/0,33009,2099708-13,00.html

Grubesic, T. H., & Mack, E. A. (2008). Spatio-temporal interaction of urban crime. *Journal of Quantitative Criminology, 24*(3), 285-306. doi:https://doi.org/10.1007/s10940-008-9047-5

Haberman, C. P., & Ratcliffe, J. H. (2015). Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts. *Criminology, 53*(3), 457-483. doi:https://doi.org/10.1111/1745-9125.12076

Hagberg, A. A., Schult, D. A., Swart, P. J., Varoquaux, G., Vaught, T., & Millman, J. (2008). Exploring Network Structure, Dynamics, and Function Using NetworkX. *Proceedings of the 7th Python in Science Conference.* SciPy2008.

Hart, T. C., & Miethe, T. D. (2015). Configural Behavior Settings of Crime Event Locations: Toward an alternative conceptualization of criminogenic microenvironments. *Journal of Research in Crime and Delinquency, 52*(3), 373-402. doi:https://doi.org/10.1177/0022427814566639

He, Z., Deng, M., Xie, Z., Wu, L., Chen, Z., & Pei, T. (2020). Discovering the joint influence of urban facilities on crime occurrence using spatial co-location pattern mining. *Cities, 99*, 102612. doi:https://doi.org/10.1016/j.cities.2020.102612

He, Z., Tao, L., Xie, Z., & Xu, C. (2020). Discovering spatial interaction patterns of near repeat crime by spatial association rules mining. *Scientific Reports, 10*(1), 1-11. doi:https://doi.org/10.1038/s41598-020-74248-w

Hipp, J. R., & Kim, Y. A. (2019). Explaining the temporal and spatial dimensions of robbery: Differences across measures of the physical and social environment. *Journal of Criminal Justice, 60*, 1-12. doi:https://doi.org/10.1016/j.jcrimjus.2018.10.005

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780. doi:10.1162/neco.1997.9.8.1735

Huang, C., Zhang, J., Zheng, Y., & Chawla, N. V. (2018). DeepCrime: Attentive hierarchical recurrent networks for crime prediction. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (pp. 1423-1432). Torino: ACM.

Huet, E. (2015, March 1). *Server And Protect: Predictive Policing Firm PredPol Promises To Map Crime Before It Happens*. Retrieved September 27, 2021, from Forbes: https://www.forbes.com/sites/ellenhuet/2015/02/11/predpol-predictive-policing/?sh=610e74fd4f9b

Hunt, J. M. (2016). *Do crime hot spots move? Exploring the effects of the modifiable areal unit problem and modifiable temporal unit problem on crime hot spot stability* . Washington,D.C.: American University.

Irvin-Erickson, Y., & La Vigne, N. (2015). A spatio-temporal analysis of crime at Washington, DC metro rail: Stations' crime-generating and crime-attracting characteristics as transportation nodes and places. *Crime science, 4*(1), 1-13. doi:https://doi.org/10.1186/s40163-015-0026-5

Jacobs, J. (1961). *The death and life of great American cities.* New York City, NY: Vintage .

Jeffery, C. R. (1977). *Crime prevention through environmental design* (Vol. 524). California: Sage Publications.

Jin, G., Sha, H., Feng, Y., Cheng, Q., & Huang, J. (2021). GSEN: An ensemble deep learning benchmark model for urban hotspots spatiotemporal prediction.

Neurocomputing, 455, 353-367. doi:https://doi.org/10.1016/j.neucom.2021.05.008

Johnson, L. T., Taylor, R. B., & Groff, E. R. (2015). Metropolitan local crime clusters: Structural concentration effects and the systemic model. *Journal of Criminal Justice, 43*(3), 186-194. doi:https://doi.org/10.1016/j.jcrimjus.2015.03.002

Johnson, S. D., Bowers, K. J., Birks, D. J., & Pease, K. (2009). Predictive mapping of crime by ProMap: Accuracy, units of analysis, and the environmental backcloth. In D. Weisburd, W. Bernasco, & G. Bruinsma, *Putting crime in its place* (pp. 171-198). New York, NY: Springer.

Johnson, S. D., Bowers, K. J., Birks, D. J., & Pease, K. (2009). Predictive mapping of crime by ProMap: Accuracy, units of analysis, and the environmental backcloth. In D. Weisburd, W. Bernasco, & G. Bruinsma, *Putting crime in its place* (pp. 171-198). New York, NY.: Springer.

Jones, R. W., & Pridemore, W. A. (2019). Toward an integrated multilevel theory of crime at place: Routine activities, social disorganization, and the law of crime concentration. *Journal of Quantitative Criminology, 35*(3), 543-572. doi:https://doi.org/10.1007/s10940-018-9397-6

Kadar, C., Maculan, R., & Feuerriegel, S. (2019). Public decision support for low population density areas: An imbalance-aware hyper-ensemble for spatio-temporal crime prediction. *Decision Support Systems, 119*, 107-117. doi:https://doi.org/10.1016/j.dss.2019.03.001

Kang, H. W., & Kang, H. B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *PloS One, 12*(4), e0176244. doi:https://doi.org/10.1371/journal.pone.0176244

Keating, A. (2008). *Chicago Neighborhoods and Suburbs: A Historical Guide.* Chicago, IL: University of Chicago Press.

Kennedy, L. W., Caplan, J. M., & Piza, E. L. (2015). *A multi-jurisdictional test of risk terrain modeling and place-based evaluation of environmental risk-based patrol deployment strategies.* Newark, NJ: Rutgers Center on Public Security.

Kennedy, L. W., Caplan, J. M., Piza, E. L., & Buccine-Schraeder, H. (2016). Vulnerability and exposure to crime: Applying risk terrain modeling to the study of assault in Chicago. *Applied Spatial Analysis and Policy, 9*(4), 529-548. doi:https://doi.org/10.1007/s12061-015-9165-z

Kim, H. J., Fay, M. P., Feuer, E. J., & Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine, 19*(3), 335-351. doi:https://doi.org/10.1002/(SICI)1097-0258(20000215)19:3<335::AID-SIM336>3.0.CO;2-Z

Kim, Y. A. (2018). Examining the relationship between the structural characteristics of place and crime by imputing census block data in street segments: Is the pain worth the gain? *Journal of Quantitative Criminology, 34*(1), 67-110. doi:https://doi.org/10.1007/s10940-016-9323-8

Kinney, J. B., Brantingham, P. L., Wuschke, K., Kirk, M. G., & Brantingham, P. J. (2008). Crime attractors, generators and detractors: Land use and urban crime opportunities. *Built Environment, 34*(1), 62-74. doi:https://doi.org/10.2148/benv.34.1.62

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *ICLR 2017* (pp. 1-14). Toulon: ICLR. Retrieved from https://arxiv.org/pdf/1609.02907.pdf

Kitchen, P., & Williams, A. (2010). Quality of life and perceptions of crime in Saskatoon, Canada. *Social Indicators Research, 95*(1), 33-61. doi:https://doi.org/10.1007/s11205-009-9449-2

Kleemans, E. R. (2001). Repeat burglary victimisation: results of empirical research in the Netherlands. *Crime Prevention Studies, 12*, 53-68.

Kulldorff, M., Athas, W. F., Feurer, E. J., Miller, B. A., & Key, C. R. (1998). Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health, 88*(9), 1377-1380.

Kulldorff, M., Heffernan, R., Hartman, J., Assunçao, R., & Mostashari. (2005). A space–time permutation scan statistic for disease outbreak detection. *PLoS Medicine, 2*(3), e59. doi:ttps://doi.org/10.1371/journal.pmed.0020059

Lantz, B., & Ruback, R. B. (2017). A networked boost: Burglary co-offending and repeat victimization using a network approach. *Crime & Delinquency, 63*(9), 1066-1090. doi:https://doi.org/10.1177/0011128715597695

Levin, A., Rosenfeld, R., & Deckard, M. (2017). The law of crime concentration: An application and recommendations for future research. *Journal of Quantitative Criminology, 33*(3), 635-647. doi:https://doi.org/10.1007/s10940-016-9332-7

Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv*, 1707.01926. doi:arXiv:1707.01926v3

Lockwood, B. (2012). The presence and nature of a near-repeat pattern of motor vehicle theft. *Security Journal, 25*(1), 38-56. doi:https://doi.org/10.1057/sj.2011.5

Lu, Y., & Chen, X. (2007). On the false alarm of planar K-function when analyzing urban crime distributed along streets. *Social Science Research, 36*(2), 611-632. doi:https://doi.org/10.1016/j.ssresearch.2006.05.003

MacDonald, J. M., Nicosia, N., & Ukert, B. D. (2018). Do schools cause crime in neighborhoods? Evidence from the opening of schools in Philadelphia. *Journal of Quantitative Criminology, 34*(3), 717-740. doi:https://doi.org/10.1007/s10940-017-9352-y

Maki, N., & Okabe, A. (2005). A Spatio-Temporal Analysis of Aged Members of a Fitness Club in a Suburb. *Proceedings of Geographical Information Systems Associations, 14*, 29-34.

Malleson, N., Steenbeek, W., & Andresen, M. A. (2019). Identifying the appropriate spatial resolution for the analysis of crime patterns. *PloS One*, e0218324. doi:https://doi.org/10.1371/journal.pone.0218324

Mazeika, D. M., & Kumar, S. (2017). Do crime hot spots exist in developing countries? Evidence from India. *Journal of Quantitative Criminology, 33*(1), 45-61. doi:https://doi.org/10.1007/s10940-016-9280-2

McCord, E. S., & Ratcliffe, J. H. (2007). A micro-spatial analysis of the demographic and criminogenic environment of drug markets in Philadelphia. *Australian & New Zealand Journal of Criminology, 40*(1), 43-63. doi:https://doi.org/10.1375/acri.40.1.43

McCord, E. S., & Ratcliffe, J. H. (2009). Intensity value analysis and the criminogenic effects of land use features on local crime patterns. *Crime Patterns and Analysis, 2*(1), 17-30.

Mohler, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting, 30*(3), 491-497. doi:https://doi.org/10.1016/j.ijforecast.2014.01.004

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association, 106*(493), 100-108. doi:https://doi.org/10.1198/jasa.2011.ap09546

Morgan, F. (2001). Repeat burglary in a Perth suburb: indicator of short-term or long-term risk? *Crime Prevention Studies, 12*, 83-118.

Nashif, N., & Fatafta, M. (2017, October 4). *The Israeli algorithm criminalizing Palestinians for online dissent*. Retrieved September 27, 2021, from Open Democracy: https://www.opendemocracy.net/en/north-africa-west-asia/israeli-algorithm-criminalizing-palestinians-for-o/

Nelessen, A. C. (1994). *Visions for a new American dream.* Planners Press,.

Newton, A., & Hirschfield, A. (2009). Measuring violence in and around licensed premises: The need for a better evidence base. *Crime Prevention and Community Safety, 11*(3), 171-188. doi:https://doi.org/10.1057/cpcs.2009.12

Ng, W., Minasny, B., Mendes, W. D., & Demattê, J. A. (2020). The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. *Soil, 6*(2), 565-578. doi:https://doi.org/10.5194/soil-6-565-2020, 2020

Nobles, M. R., Ward, J. T., & Tillyer, R. (2016). The impact of neighborhood context on spatiotemporal patterns of burglary. *Journal of Research in Crime and Delinquency, 53*(5), 711-740. doi:https://doi.org/10.1177/0022427816647991

Oberwittler, D., & Wikström, P. O. (2009). Why small is better: Advancing the study of the role of behavioral contexts in crime causation. In D. Weisburd, W. Bernasco, Bruinsma, & G., *Putting crime in its place* (pp. 35-39). New York, NY: Springer.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association, 83*(401), 9-27. doi:10.1080/01621459.1988.10478560

Ogawa, T., Sasaka, Y., Maeda, K., & Haseyama, M. (2018). Favorite video classification based on multimodal bidirectional LSTM. *IEEE Access, 6,* 61401-61409. doi:10.1109/ACCESS.2018.2876710

Ohyama, T., & Amemiya, M. (2018). Applying Crime Prediction Techniques to Japan: A Comparison Between Risk Terrain Modeling and Other Methods. *European Journal on Criminal Policy & Research,, 4*(2), 469-487.

Okabe, A., & Sugihara, K. (2012). *Spatial analysis along networks: statistical and computational methods.* New York,NY: John Wiley & Sons.

Okabe, A., & Yamada, I. (2001). The K-function method on a network and its computational implementation. *Geographical Analysis, 33*(3), 271-290. doi: https://doi.org/10.1111/j.1538-4632.2001.tb00448.x

Olah, C. (2015, Augusr 27). *Understanding LSTM Networks*. Retrieved October 18, 2021, from https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Olligschlaeger, A. M. (1997). Artificial neural networks and crime mapping. In *Crime mapping and crime prevention* (pp. 313-348). Monsey,NY: Criminal Justice Press.

Openshaw, S. (1981). The modifiable areal unit problem. *Quantitative geography: a British view,* 60-69.

Ortega, A., Frossard, P., Kovačević, J., Moura, J. M., & Vandergheynst, P. (2018). Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE,, 106*(5), 808-828. doi:10.1109/JPROC.2018.2820126

Payroll and Timekeeping—Attendance. (1996). Retrieved July 3, 2021, from http://directives.chicagopolice.org/directives/data/a7a57b36-12cf4df7-24112-cf4e-9398046d4f55fbaf.html

Pease, K. (1998). *Repeat victimisation: Taking stock* (Vol. 90). London: Home Office Police Research Group.

Perry, W. L. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations.* Santa Monica,CA: Rand Corporation.

Piquero, A., & Rengert, G. F. (1999). Studying deterrence with active residential burglars. *Justice Quarterly, 16*(2), 451-471. doi:https://doi.org/10.1080/07418829900094211

Polvi, N., Looman, T., Humphries, C., & Pease, K. (1990). Repeat break and enter victimisation: Time course and crime prevention opportunity. *Journal of Police Science and Administration, 17*(1), 8-11.

Polvi, N., Looman, T., Humphries, C., & Pease, K. (1991). The time course of repeat burglary victimization. *The British Journal of Criminology, 31*(4), 411-414. doi:https://doi.org/10.1093/oxfordjournals.bjc.a048138

Pred, A. (1981). Social reproduction and the time-geography of everyday life. *Geografiska Annaler. Series B, Human Geography, 63*(1), 5-22. doi:https://doi.org/10.2307/490994

Ratcliffe, J. H. (2005). Detecting spatial movement of intra-region crime patterns over time. *Journal of Quantitative Criminology, 21*(1), 103-123. doi:https://doi.org/10.1007/s10940-004-1789-0

Ratcliffe, J. H. (2006). A temporal constraint theory to explain opportunity-based spatial offending patterns. *Journal of Research in Crime and Delinquency, 43*(3), 261-291. doi:https://doi.org/10.1177/0022427806286566

Ratcliffe, J. H. (2012). The spatial extent of criminogenic places: a changepoint regression of violence around bar. *Geographical Analysis, 44*(4), 302-320. doi: https://doi.org/10.1111/j.1538-4632.2012.00856.x

Rengert, G. F., & Wasilchick, J. (1985). *Suburban burglary: A time and a place for everything.* Springfield, IL: CC Thomas.

Rengert, G. F., Piquero, A. R., & Jones, P. R. (1999). Distance decay reexamined. *Criminology, 37*(2), 427-446. doi: https://doi.org/10.1111/j.1745-9125.1999.tb00492.x

Roncek, D. W., & Maier, P. A. (1991). Bars, blocks, and crimes revisited: Linking the theory of routine activities to the empiricism of "hot spots". *Criminology, 29*(4), 725-753. doi:https://doi.org/10.1111/j.1745-9125.1991.tb01086.x

Roncek, D. W., & Pravatiner, M. A. (1989). Additional evidence that taverns enhance nearby crime. *Sociology and Social Research, 73*(4), 185-188.

Rosser, G., & Cheng, T. (2019). Improving the robustness and accuracy of crime prediction with the self-exciting point process through isotropic triggering. *Applied Spatial Analysis and Policy, 12*(1), 5-25. doi:https://doi.org/10.1007/s12061-016-9198-y

Rosser, G., Davies, T., Bowers, K. J., Johnson, S. D., & Cheng, T. (2017). Predictive crime mapping: Arbitrary grids or street networks? *Journal of Quantitative Criminology, 33*(3), 569-594. doi:https://doi.org/10.1007/s10940-016-9321-x

Rummens, A., Hardyns, W., & Pauwels, L. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Applied Geography, 86*, 255-261. doi:https://doi.org/10.1016/j.apgeog.2017.06.011

Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing, 323*, 203-213. doi:https://doi.org/10.1016/j.neucom.2018.09.082

Sampson, R. (2012). *Great American City: Chicago and the Enduring Neighborhood Effect.* Chicago, IL: University of Chicago Press.

Schnell, C., Braga, A. A., & Piza, E. L. (2017). The influence of community areas, neighborhood clusters, and street segments on the spatial variability of violent crime in Chicago. *Journal of Quantitative Criminology, 33*(3), 469-496. doi:https://doi.org/10.1007/s10940-016-9313-x

Schnell, C., Grossman, L., & Braga, A. A. (2019). The routine activities of violent crime places: A retrospective case-control study of crime opportunities on street segments. *Journal of Criminal Justice, 60*, 140-153. doi:https://doi.org/10.1016/j.jcrimjus.2018.10.002

Shaw, C. R., & McKay, H. D. (1942). *Juvenile delinquency and urban areas.* Chicago,IL: University of Chicago Press.

Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime “hot spots”: A randomized, controlled trial. *Justice Quarterly, 12*(4), 625-648. doi:https://doi.org/10.1080/07418829500096221

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology, 27*(1), 27-56. doi:https://doi.org/10.1111/j.1745-9125.1989.tb00862.x

Shleifer, S., McCreery, C., & Chitters, V. (2019). Incrementally improving graph wavenet performance on traffic prediction. *arXiv preprint*, 912.07390. doi:arXiv:1912.07390v1

Short, M. B., Brantingham, P. J., Bertozzi, A. L., & Tita, G. E. (2010). Dissipation and displacement of hotspots in reaction-diffusion models of crime. *Proceedings of the National Academy of Sciences, 107*(9), 3961-3965. doi:https://doi.org/10.1073/pnas.0910921107

Short, M. B., D’orsogna, M. R., Brantingham, P. J., & Tita, G. E. (2009). Measuring and modeling repeat and near-repeat burglary effects. *Journal of Quantitative Criminology, 25*(3), 325-339. doi:https://doi.org/10.1007/s10940-009-9068-8

Smith, M. (2018, October 30). *Can we predict when and where a crime will take place?* Retrieved September 27, 2021, from BBC: https://www.bbc.com/news/business-46017239

Song, G., Bernasco, W., Liu, L., Xiao, L., Zhou, S., & Liao, W. (2019). Crime feeds on legal activities: Daily mobility flows help to explain thieves’ target location choices. *Journal of Quantitative Criminology,, 35*(4), 831-854. doi:https://doi.org/10.1007/s10940-019-09406-z

South China Morning Post. (2018, January 28). *Japan trials AI-assisted predictive policing before 2020 Tokyo Olympics*. Retrieved September 27, 2021, from South China Morning Post: https://www.scmp.com/news/asia/east-asia/article/2130980/japan-trials-ai-assisted-predictive-policing-2020-tokyo-olympics

Space and Naval Warfare Systems Center Atlantic. (2013). *Geographic Information Systems and Predictive Policing Application Note.* Washington, DC: U.S. Department of Homeland Security. Retrieved from https://www.dhs.gov/sites/default/files/publications/GIS-Predictive-Policing-AppN_0813-508_0.pdf

Steenbeek, W., & Kreis, C. (2015). Where broken windows should be fixed: Toward identification of areas at the tipping point. *Journal of Research in Crime and Delinquency*, *52*(4), 511-533. doi:https://doi.org/10.1177/0022427815580166

Steenbeek, W., & Weisburd, D. (2016). Where the action is in crime? An examination of variability of crime across different spatial units in The Hague, 2001–2009. *Journal of Quantitative Criminology*, *32*(3), 449-469. doi:https://doi.org/10.1007/s10940-015-9276-3

Stucky, T. D., & Ottensmann, J. R. (2009). Land use and violent crime. *Criminology, 47*(4), 1223-1264. doi: https://doi.org/10.1111/j.1745-9125.2009.00174.x

Timms, D. (1975). *The urban mosaic: towards a theory of residential differentiation.* Cambridge, UK: CUP Archive.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(sup1), 234-240. doi:DOI: 10.2307/143141

Tompson, L., Partridge, H., & Shepherd, N. (2009). Hot routes: Developing a new technique for the spatial analysis of crime. *Crime Mapping: A Journal of Research and Practice, 1*(1), 77-96.

Townsley, M., Homel, R., & Chaseling, J. (2003). Infectious burglaries. A test of the near repeat hypothesis. *British Journal of Criminology*, 615-633. doi:https://doi.org/10.1093/bjc/43.3.615

Tseloni, A., & Pease, K. (1996). Repeat victimization and policing: Whose home gets victimized, where and how often. *Conference of the Swedish Police College.* Stockholm.

Uberti, D. (2021, July 8). *AFTER BACKLASH, PREDICTIVE POLICING ADAPTS TO A CHANGED WORLD*. Retrieved September 27, 2021, from The Wall Street Journal: https://www.wsj.com/articles/after-backlash-predictive-policing-adapts-to-a-changed-world-11625752931

Vandeviver, C., & Bernasco, W. (2017). The geography of crime and crime control. *Applied Geography, 86*, 220-225. doi:10.1016/j.apgeog.2017.08.012

Vandeviver, C., & Steenbeek, W. (2019). The (in) stability of residential burglary patterns on street segments: The case of Antwerp, Belgium 2005–2016. *Journal of Quantitative Criminology, 35*(1), 111-133. doi:https://doi.org/10.1007/s10940-017-9371-8

Wang, B., Luo, X., Zhang, F., Yuan, B., Bertozzi, A. L., & Brantingham, P. J. (2018). Graph-based deep modeling and real time forecasting of sparse spatio-temporal data. *arXiv preprint*.

Wang, X., & Brown, D. E. (2012). The spatio-temporal modeling for criminal incidents. *Security Informatics, 1*(1), 1-17. doi:https://doi.org/10.1186/2190-8532-1-2

Weisburd, D., Bruinsma, G. J., & Bernasco, W. (2009). Units of analysis in geographic criminology: Historical development, critical issues, and open questions. In D. Weisburd, W. Bernasco, & G. J. Bruinsma, *Putting crime in its place* (pp. 3-31). New York, NY: Springer.

Weisburd, D., Groff, E. R., & Yang, S. M. (2012). *The criminology of place: Street segments and our understanding of the crime problem.* Oxford: Oxford University Press.

Weisburd, D., Wyckoff, L. A., Ready, J., Eck, J. E., Hinkle, J. C., & Gajewski, F. (2006). Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. *Criminology, 44*(3), 549-592. doi: https://doi.org/10.1111/j.1745-9125.2006.00057.x

Wheeler, A. P. (2019). Quantifying the local and spatial effects of alcohol outlets on crime. *Crime & Delinquency, 65*(6), 845-871. doi:https://doi.org/10.1177/0011128718806692

Wheeler, A. P., & Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology, 37*(2), 445-480. doi:https://doi.org/10.1007/s10940-020-09457-7

Wilcox, P., & Tillyer, M. S. (2017). Place and Neighborhood Contexts 1. In D. Weisburd, & J. Eck, *Unraveling the Crime-Place Connection* (pp. 121-142). London: Routledge.

Wilcox, P., Land, K. C., & Hunt, S. A. (2003). *Criminal Circumstance: A Dynamic Multicontextual Criminal Opportunity Theory.* New York, NY: Routledge.

Wilson, J. Q., & Kelling, G. L. (1982). Broken windows. *Atlantic monthly, 249*(3), 29-38.

Wong, D. W. (2004). The modifiable areal unit problem (MAUP). In D. Janelle, B. Warf, & K. Hansen, *WorldMinds: Geographical perspectives on 100 problems* (pp. 571-575). Dordrecht: Springer.

Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint*, 1906.00121. doi:arXiv:1906.00121

Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2020). Connecting the dots: Multivariate time series forecasting with graph neural networks. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 753-763). ACM. doi:https://doi.org/10.1145/3394486.3403118

Xu, J., & Griffiths, E. (2017). Shooting on the street: measuring the spatial influence of physical features on gun violence in a bounded street network. *Journal of Quantitative Criminology, 33*(2), 237-253. doi:https://doi.org/10.1007/s10940-016-9292-y

Xu, Y., Fu, C., Kennedy, E., Jiang, S., & Owusu-Agyemang, S. (2018). The impact of street lights on spatial-temporal patterns of crime in Detroit, Michigan. *Cities, 79*, 45-52. doi:https://doi.org/10.1016/j.cities.2018.02.021

Yamada, I., & Thill, J. C. (2004). Comparison of planar and network K-functions in traffic accident analysis. *Journal of Transport Geography, 12*(2), 149-158. doi:https://doi.org/10.1016/j.jtrangeo.2003.10.006

Youstin, T. J., Nobles, M. R., Ward, J. T., & Cook, C. L. (2011). Assessing the generalizability of the near repeat phenomenon. *Criminal Justice and Behavior, 38*(10), 1042-1063. doi:https://doi.org/10.1177/0093854811417551

Youstin, T. J., Nobles, M. R., Ward, J. T., & Cook, C. L. (2011). Assessing the generalizability of the near repeat phenomenon. *Criminal Justice and Behavior, 38*(10), 1042-1063. doi:https://doi.org/10.1177/0093854811417551

Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *IJCAI*, (pp. 3634-3640). Stockholm.

Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv*, 1511.07122. doi: arXiv:1511.07122v3

Yue, H., Zhu, X., Ye, X., & Guo, W. (2017). The local colocation patterns of crime and land-use features in Wuhan, China. *ISPRS International Journal of Geo-Information, 6*(10), 307. doi:https://doi.org/10.3390/ijgi6100307

Zhang, X., Liu, L., Xiao, L., & & Ji, J. (2020). Comparison of machine learning algorithms for predicting crime hotspots. *IEEE Access, 8*, 181302-181310. doi:10.1109/ACCESS.2020.3028420

Zhang, Y., & Cheng, T. (2020). Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. *Computers, Environment and Urban Systems*, 79. doi:https://doi.org/10.1016/j.compenvurbsys.2019.101403

Zhao, X., & Tang, J. (2020). Exploring Spatio-Temporal and Cross-Type Correlations for Crime Prediction. *arXiv preprint*, 2001.06923.

Zhuang, Y., Almeida, M., Morabito, M., & Ding, W. (2017). Crime hot spot forecasting: A recurrent model with spatial and temporal information. *IEEE International Conference on Big Knowledge* (pp. 143-150). Hefei: IEEE. doi:10.1109/ICBK.2017.3