Title
Computational Methods to Understand Mexico's Organized Crime and Drug Violence

Author
Contreras Velasco, Oscar Fernando

Publication Date
2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational Methods to Understand Mexico's Organized Crime and Drug Violence

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Sociology

by

Oscar Fernando Contreras Velasco

2024

ABSTRACT OF THE DISSERTATION

Computational Methods to Understand Mexico's Organized Crime and Drug Violence

by

Oscar Fernando Contreras Velasco

Doctor of Philosophy in Sociology

University of California, Los Angeles, 2024

Professor Ruben Hernandez-Leon, Co-Chair

Professor Gabriel Rossman, Co-Chair

This dissertation is structured as a three-paper dissertation. The first chapter builds on social network analysis and structural balance theory to analyze, with a novel approach, some of the unintended consequences of Mexico's kingpin strategy on the network of criminal organizations. The goal of this chapter is threefold: first, to show that the kingpin strategy is associated with the fragmentation of criminal organizations in Mexico; second, criminal organizations developed a set of structurally balanced arrangements before the government waged a war against them and that the kingpin strategy disrupted such arrangements; third, the fragmentation of criminal organizations also produced a process of clustering of violence. The second chapter examines the hazards faced by undocumented Mexican migrants during their U.S.-Mexico border crossings, particularly in the context of violent confrontations between drug trafficking organizations. By

employing a negative binomial mixed effects model, difference-in-differences analysis, and unsupervised machine learning techniques, the research aims to discern which demographics within the undocumented migrant population are more susceptible to heightened hazards throughout their journey. The third chapter uses a combination of social network analysis (SNA) and machine learning techniques to predict links in Mexico's network of criminal organizations. Specifically, it uses four similarity- based algorithms to estimate the likelihood that a link will be formed between two unconnected organizations in the network. Four algorithms are enhanced with the Node2Vec algorithm and a Deep Neural Network Architecture (DNNs) to improve their predictive capabilities. Of the node-similarity indices implemented, the Preferential Attachment algorithm enhanced with both the Node2Vec and the DNN architectures performed the best. When predicting potential future ties, the best performing algorithm was the Jaccard Coefficient enhanced both by the Node2Vec. I argue that this happens because the Jaccard Coefficient captures the immediate neighbor overlap but does not account for longer paths or the global network structure, while the Node2Vec does capture broader topological and community-based features that might not be apparent from direct connections alone.

The dissertation of Oscar Fernando Contreras Velasco is approved.

David Shirk

Andres Villarreal

Gabriel Rossman, Committee Co-Chair

Ruben Hernandez-Leon, Committee Co-Chair

University of California, Los Angeles

2024

DEDICATION

I dedicate this dissertation to my co-chairs Ruben Hernandez-Leon and Gabriel Rossman, whose unwavering guidance and support have been invaluable throughout my Ph.D.

To my parents, your love and staunch belief in me has been my greatest source of strength. To my little sisters, thank you for pushing me to be a better role model every day.

To Nushy, your patience, understanding, and constant support have been my rock. Thank you for being by my side every step of the way.

To my friends and colleagues, your camaraderie and support have made this journey more enriching and enjoyable. Thank you for being there through the challenges and triumphs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF MAPS

AKNOWLEDGMENTS

I would like to express my deepest gratitude to my co-chairs, Ruben Hernandez-Leon and Gabriel Rossman, for their unwavering guidance, support, and encouragement throughout my Ph.D. journey. Your expertise, patience, and dedication have been invaluable in shaping this dissertation, and I am profoundly grateful for the countless hours you have spent reviewing my work, providing insightful feedback, and motivating me to achieve my best.

I also wish to extend my sincere thanks to my dissertation committee members, Andres Villareal, and David A. Shirk. Your diverse perspectives, constructive criticism, and thoughtful suggestions have significantly enriched my research.

Finally, special thanks to Nathan Jones, who generously shared his expertise on criminal networks and the LANTIA dataset. Without your contribution, chapter 3 of this dissertation would not have been possible. Thank you for your invaluable support.

VITA

Oscar Contreras Velasco earned his Bachelor of Arts in International Relations from Instituto Tecnologico Autonomo de Mexico (ITAM) in 2010. His passion for understanding social dynamics, particularly in the context of organized crime and migration, led him to pursue advanced studies. In 2014, Oscar received his Master of Arts in Social Anthropology from Universidad Iberoamericana, Mexico City. His master's thesis, titled "Police Institution, violence and the culture of terror in Tijuana," explored the ways in which police officers experienced and navigated violence in the context of the drug war in Mexico.

Oscar continued his academic journey at the University of California, Los Angeles (UCLA), where he pursued a second M.A. in Sociology and a Ph.D. in Sociology. Under the mentorship of Professors Ruben Hernandez-Leon and Gabriel Rossman, Oscar's research focused on using computational methods to analyze organized crime and drug violence in Mexico. His dissertation, "Computational Methods to Understand Mexico's Organized Crime and Drug Violence," employs statistics, social network analysis, and machine learning techniques to understand some of the effects of the drug violence and uncover hidden patterns and predict future trends in criminal networks. During his time at UCLA, Oscar received several honors and awards, including the 2023 Kollock Distinguished Teaching Award and multiple Excellence in Teaching Awards. He has presented his work at numerous national and international conferences, contributing valuable insights to the field of sociology.

Beyond academia, Oscar is committed to applying his research to real-world challenges. He has collaborated with various non-governmental organizations and policy think tanks, such as Rice University's Baker Institute, to develop strategies for mitigating the impact of organized crime.

His work has been published in several peer-reviewed journals and featured in media outlets discussing the complexities of crime and migration.

Introduction:

Over the last two decades, incremental developments in computational power have revolutionized the ability to process and analyze complex data, allowing for advanced computational methods to better understand sociological phenomena. Methods such as Social Network Analysis (SNA), Machine Learning, and statistics have become powerful tools to analyze complex datasets, uncover hidden patterns, and forecast future trends. Given that my research interests focused on Mexico's organized crime, violence, and migration, the use of these quantitative methods allowed me to map and analyze the relationships and interactions between individuals, groups, and entities within criminal networks, as well as identify key players and structural vulnerabilities. Machine Learning algorithms also allowed me to understand patterns and anomalies that would be impossible for human analysts to detect due to the scale and complexity of the data involved. In the context of the study of migration, for instance, Machine Learning and statistical techniques allowed me to identify factors influencing migration, and even predict future trends. The overarching questions my dissertation addresses are: How can we understand the complex and violent relationships between Mexico's criminal organizations? What are the main features and patterns that characterize these relationships? What are some of the consequences of this violence on the population, and how should we understand organized crime? How can we deal with the lack of good quality data on criminal organizations in Mexico and elsewhere?

Although each chapter in this dissertation is structured as a single academic paper, all three chapters are related in that they all focus on Mexico, they make use of computational methods, and they all follow a core argument: Mexico's war on drugs, initiated by ex-president Felipe Calderon in 2006, started a process of fragmentation of criminal organizations that led to an increase of violence all over the country and that had pernicious consequences for the Mexican population. I focus on the consequences of this fragmentation and violence on the hazards migrants face when crossing the U.S. – Mexico border. Although I chose to focus on Mexico, the debates raised are broader and consequential to other countries and regions. For instance, the first chapter raises the question about the network of violence amongst criminal organizations evolves and adapts given external shocks in the form of militarization and targeted attacks. This example provides a framework to understand some of the consequences of equivocated state policies in the face of organized crime and violence, which can be applied to many other case studies. Likewise, the second chapter provides a frame to understand the state-like nature of organized crime given its capacity to control a territory, tax its population, and create protection rackets in a context of increased violence. Finally, although the third chapter is mainly a methodological one, it proposes the use of algorithms enhanced with artificial intelligence architectures to better predict the structure of networks and the future formation of links. Below is a brief description of each one of these chapters.

Chapter 1 builds on social network analysis and structural balance theory to analyze some of the unintended consequences of Mexico's kingpin strategy on the network of criminal organizations. I use data on violent conflicts between Mexico's criminal organizations, between 2004 and 2020, from the Uppsala Conflict Data Program (UCDP), and a combination of statistics, social network analysis, GIS, and archival methods to understand the patterns and geography of violent conflicts and alliances before and after the war on drugs. This chapter shows that the kingpin strategy is associated with the fragmentation of criminal organizations in Mexico, and that criminal organizations developed a set of structurally balanced arrangements before the government waged a war against them and that the kingpin strategy disrupted such arrangements, which led to an increase in the number of violent conflicts. Finally, I show that the fragmentation of criminal organizations also produced a process of clustering of violence, where sets of organizations started fighting each other in specific regions of the country, increasing the levels of violence in those geographical spaces.

Chapter 2 examines the hazards faced by undocumented Mexican migrants during their U.S.-Mexico border crossings, in the context of violent confrontations between drug trafficking organizations. I use data from the "Encuesta de Migración y Frontera Norte (EMIF)" Survey, which details migratory flows from Mexico to the United States, census information from Mexico's National Institute of Statistics and Geography (INEGI), and data from the CBP enforcement statistics. By employing a negative binomial mixed effects model, difference-in-differences analysis, and unsupervised machine learning techniques, the chapter aims to discern which demographics within the undocumented migrant population are more susceptible to heightened hazards throughout their journey and the impact of criminal organizations' territorial control in border cities on such hazards. The central questions addressed are: What specific demographic groups among Mexican undocumented migrants are more prone to hazardous border crossings? How does the presence of organized crime groups controlling the border influence the hazards encountered by these migrants? Additionally, how does the choice of crossing region affect the dangers faced during their journey? In this chapter I argue that specific demographics are likely to encounter increased hazards when crossing the U.S.-Mexico border, and that traversing territories disputed by multiple criminal organizations correlates with escalated violence and greater dangers for migrants. This increased hazard is attributed to the incapacity of criminal organizations, which function as quasi-states, to regulate illegal activities and protect their networks in regions where they do not exert undisputed control, which includes overseeing human smuggling. Finally, the research proposes that migrants crossing through the region generally referred to as the "East" are exposed to higher, underlining the significance of geographical factors in their dangerous journey.

Chapter 3 is the product of a collaborative enterprise, where I am first author, that explores social network analysis (SNA) and machine learning techniques to predict links in Mexico's network of criminal organizations. One of the main challenges in the studies of organized crime in countries like Mexico, is the dearth of good quality data on criminal networks that is publicly available. This chapter is an effort to address this lack of data by proposing advanced computational methods that allow researchers to uncover and predict data that might not be available otherwise. We propose the use of four similarity indices enhanced with the Node2Vec algorithm and a Deep Neural Network Architecture (DNNs) to make predictions about the structure and future links in a criminal network. Of the node-similarity indices implemented, we found that the Preferential Attachment algorithm enhanced with both the Node2Vec and the DNN architectures performed the best. We conclude that this is in part due to the power law-like distribution of the criminal network. We found that this criminal network is structured such that nodes with many connections are more likely to gain more and vice versa. This also explains why the Preferential Attachment performed in all its variants performed so well when predicting ties in this network. We argue that such structure can be explained by the existence of hierarchical relationships where a few organizations monopolize activities, control, and subcontract other smaller organizations. When predicting potential future ties, we found that the best performing algorithm was the Jaccard Coefficient enhanced both by the Node2Vec. We argue that this happens because the Jaccard Coefficient captures the immediate neighbor overlap but does not account for longer paths or the global network structure, while the Node2Vec does capture broader topological and community-based features that might not be apparent from direct connections alone. Finally, we argue that the use of this enhanced algorithm can be very powerful to predict future network ties in any network, not only criminal ones.

Chapter 1: Unintended Consequences of State Action: How the Kingpin Strategy Transformed the Structure of Violence in Mexico's Organized Crime

Introduction

Violent conflicts between organized crime groups have become one of the main sources of violent deaths in Mexico for the last two decades. Recent studies have argued that political decentralization coupled with a kingpin strategy pursued by the federal government only worsened violence. In 2007, former president Felipe Calderon announced a war on drugs that sought to uproot these organizations and bring back peace to the population. This meant, among other things, the use of the military to tackle criminal organizations and a kingpin strategy to behead them. But far from ending violence, such a strategy only provoked the fragmentation of criminal organizations, who started fighting each other for territorial control and trafficking routes, spreading the violence across the country (Trejo and Ley, 2019; Rios, 2013; Atuesta and Perez-Davila, 2018). Between 2000 and 2009 there were more than 20,000 murders attributed to organized crime in Mexico (Shirk and Astorga, 2010: 2). The number skyrocketed in the following decade. Only in 2020, 36,773 people suffered violent deaths, which accounts for a 77% increase compared to 2015 and points to a continuing uphill trend (INEGI, 2021).

This chapter aims to build on social network analysis and structural balance theory to analyze some of the unintended consequences of the kingpin strategy pursued by ex-president Calderon, and subsequent administrations, on the structure of conflicts and alliances between criminal organizations. While the detrimental effects of the war on drugs on the population have been widely analyzed, there is less research on how this strategy affected the way these organizations structured their conflicts and alliances[1]. I use data on violent conflicts between Mexico's criminal organizations, between 2004 to 2020, from the Uppsala Conflict Data Program (UCDP). I use a combination of statistics, social network analysis, GIS, and archival methods to understand the patterns and geography of violent conflicts and alliances before and after the war on drugs. What was the effect of the kingpin strategy on the structure of conflicts and alliances among criminal organizations? Were there specific arrangements between criminal organizations that were disrupted after the strategy was implemented? How did this affect the levels and geography of violence in the country?

The goal of this chapter is threefold. First, to show that the kingpin strategy is associated with the fragmentation of criminal organizations in Mexico; second, to show that criminal organizations developed a set of structurally balanced arrangements before the government waged a war against them and that the kingpin strategy disrupted these arrangements, which led to an increase in the number of violent conflicts; third, I will argue that the fragmentation of criminal organizations also produced a process of clustering of violence, where sets of organizations started fighting each other in specific regions of the country, increasing the levels of violence in those geographical spaces.

The rest of the chapter will be structured as follows: the second section will be a brief historical context of the conflicts between the most powerful drug trafficking organizations from the early 2000s to the present (2020). I will describe the main configurations of conflicts and alliances that were produced during this period, interlacing them with the socio-political context

---

[1] Studies that analyze drug related violence in Mexico, have studied some of the social consequences of drug related violence (Trejo and Ley, 2018; Shirk and Astorga, 2010; Guerrero, 2009; Escalante 2009), the role of the state and the impact of political decentralization on the rise of drug related violence (Rios, 2013), and how the fragmentation and cooperation of drug trafficking organizations in Mexico had an effect on violence levels (Atuesta and Perez-Davila, 2017).

that set the stage for such configuration; the third section will address a brief literature review on social networks and structural balance theory, which will serve as a framework for the rest of my analysis; the fourth section describes the datasets used for my analysis and the model specifications; in the fifth section analyzes the results of my statistical and structural balance models. The results of my models will show that the beheading of criminal organizations is associated with their fragmentation, and that their network of conflicts and alliances was structurally balanced before 2010, but that the process of fragmentation ultimately led to an unbalanced structure; in the sixth section, I implement the Louvain method for community detection to show that the strategy of the government produced a regionalization or clustering of violence, with a greater number of violent conflicts occurring throughout the country and with geographically limited scope. Finally, I will include some concluding remarks and policy implications of my findings.

Historical background

Mexico's drug trafficking organizations consolidated their power in the late 1980s and early 1990s, a process facilitated by both suppression of the Colombian cartels and the Caribbean route, and the signing of NAFTA (North American Free Trade Agreement) in 1993. The dominant political party, the PRI, had been in power for over 70 years, in an arrangement that novelist Vargas Llosa called "the perfect dictatorship". This was the time of big and centralized drug trafficking cartels like the Guadalajara, Juarez, Gulf, and Arellano Felix organizations, who controlled vast territories and trafficking routes and were allowed to operate under the protection of various government agencies (Hernandez, 2010). The end of the Caribbean route made Mexico the new natural way up north, and the new trade agreement (NAFTA) made smuggling in big quantities much easier. With more money and power, drug trafficking organizations intensified their activities and increased territorial control, which led to turf wars and increased levels of corrupting power and violence.

In the 2000s, a new ruling party meant the realignment of former agreements between the government and criminal organizations, and the rise of new criminal groups, which intensified turf wars and violence (Trejo and Ley, 2018). Local governments now had more power and autonomy which allowed them to renegotiate their allegiances with old and new criminal organizations. More importantly, in 2007 the new government, under the leadership of ex-president Felipe Calderon Hinojosa, decided to focus its efforts on dismantling organized crime groups. Calderon sent the military to the streets and started a kingpin strategy with the hope that by beheading these criminal organizations, they would remain too weak to sustain their structure and activities. Two famous examples were Operation Baja California, in early 2007, and Joint Operation Michoacan, later that same year. But the kingpin strategy led to further fragmentation of criminal organizations and an increase in violence. The lack of leadership in these organizations created power vacuums that led to splinter groups that now contested the territory and trafficking routes of the original organizations. As a result of these political rearrangements and policies, between 2000 and 2009 there were more than 20,000 murders attributed to organized crime, and more than half were registered between the years 2008 and 2009. Meanwhile, 1,100 police officers and soldiers lost their lives between 2006 and 2009 in the war against these drug cartels (Shirk and Astorga, 2010: 2). Furthermore, the number of organized crime groups skyrocketed from 5, in 2006, to around 60 in 2012 (Trejo and Ley, 2018). Some cartels like the Zetas, and Cartel Jalisco Nueva Generacion (CJNG) became notorious for their capacity to produce violence and rapidly expand their geographical areas of influence. The Zetas splintered from the Gulf Cartel in 2010, following internal disputes and transforming the geography of violence throughout the whole country. Other organizations faced fragmentation right after their leaders were captured or killed by the government or enemy organizations. Such was the case of the Knights Templar, which splintered from La Familia Michoacana in 2011 right after one of their leaders was killed by the military.

When former president Enrique Pena Nieto was elected, in 2012, there were high expectations that he would change Calderon's strategy of militarization, but this did not happen. By the end of Pena Nieto's administration, the levels of homicides reached unprecedented levels: 33,341 only in 2018, which accounts for a 15% increase from the previous year, making it the most violent year in Mexico's history as of that date. That same year, Andres Manuel Lopez Obrador (AMLO) was elected president. He promised to put an end to the militarization of public safety but as of 2022, he has done just the opposite. AMLO increased the budget and role of the military in the war against organized crime.

Social network analysis and structural balance theory

Social network analysis (SNA) has been widely used to study the structure and relationships between criminal organizations. For instance, Natarajan and Belanger (1998) showed that drug trafficking organizations diversify their opportunities through the generation of diverse supplies, substances, routes, and methods. Smith and Papachristos (2016) analyzed the overlap of criminal, personal, and legitimate networks in Chicago's organized crime networks. Other studies have shown that group leaders have both high betweenness and centrality, which suggests that they are very important actors within their networks (Calderoni, 2014; Duijn, Kashirin, and Sloot, 2014;

Hofmann and Gallupe, 2015). More recently, Jones, et al. (2022) analyzed Mexico's organized crime alliance and subgroup network structures and found a bipolar network system. This chapter uses social network analysis, and builds on the notion of structural balance, to understand the effects of the kingpin strategy on the patterns of conflicts and alliances between Mexico's drug trafficking organizations. Structural balance was first formalized by Heider's (1946) balance theory. It was a psychological model of human relations and asserted that one person's relationship ('positive' or 'negative') with another is interdependent upon their evaluation or attitude towards a third person or entity. The smallest social unit of analysis in which one can assess "balance" is a triad. For any given person (i) who has a relationship with another person (j), person i experiences balance (or cognitive consistency) if and only if both i and j view another persona or an entity k unfavorably and i and j have a positive relationship. In any different configuration, person i would experience cognitive incongruence, thus imbalance.

Cartwright and Harary (1956), extended Heider's idea of balance to the theory of social and group structure and defined it as "structural balance". Consider a finite set of actors $N = \{1,2,\ldots,n\}$. Actors cannot have ties with themselves and if i has a relation with j, then j also has a relation with i (ties are undirected). The set of all possible relations between the actors is then the set X^n of all two-element subsets ij of N. A signed graph (hereafter just "graph") $G = (N,X,F)$ combines a set of actors N with a set of relations $X \in X^n$, a subset $F \in X$ of which are positive, relations in its complement $E = X \backslash F$ being negative. Cartwright and Harary (1956) argued that within triads (individuals or groups), balance can be achieved if two of the three relations are negative or if all relations among the three actors are positive. In terms of signs, triads + - - and +++ are balanced, while triads - - - and ++ - are imbalanced (See Figure 1). This logic is reflected in aphorisms about how parties may swear allegiance or declare war based on feuds or loyalties with third parties: “The friend of a friend is a friend, the friend of an enemy is a friend, the enemy of a friend is an enemy, and the enemy of an enemy is a friend” (Van De Rijt, 2011).

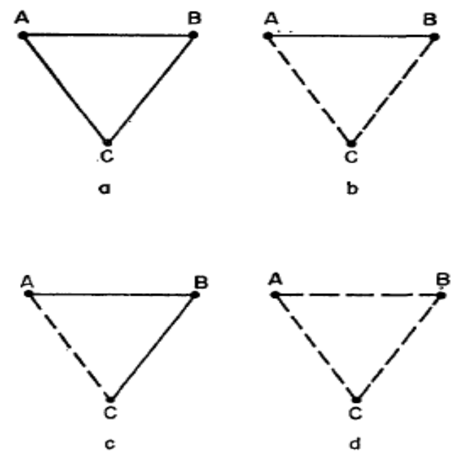


Figure. 1. Structural Balance. Structures *a* and *b* are balanced, while c and c are not balanced. Cartwright and Harary (1956).

Cycles (triangular relationships in this case) containing an odd number of negative edges are defined as unbalanced, and cycles containing 0 or an even number of negative edges will be balanced. This would imply that the signed network is totally balanced, or structurally balanced. Many signed networks, however, can have the majority of balanced cycles without reaching total balance. Instead of simply stating whether a network is balanced or unbalanced, I will show that some networks can reach ‘partial balance’, which will be defined as a state where the amount of balanced cycles is greater than unbalanced ones. Cartwright and Harary (1956) measure this with the degree of balance, which is simply the fraction of balanced cycles:

$$D(G) = \frac{\sum_{k=3}^{n} O_k^+}{\sum_{k=3}^{n} O_k}.$$

Where $\sum_{k=3}^{n} O_k^+$ is the sum of balanced cycles (closed walks), and $\sum_{k=3}^{n} O_k$ represents the total number of cycles. I assume that a walk of length k in G denotes a sequence of nodes *V0, V1,…, Vk-1, Vk* such that for each *i*=1, 2,…, k there is an edge from *Vi-1 to Vi*. If *V0 = Vk*, the sequence is a closed walk of length k. A cycle of length k will happen when all the nodes in a closed walk are distinct except the endpoints (Aref and Wilson, 2017). Having established the basic assumptions of structural balance theory, I will introduce the dataset and the model specifications I used to test the effects of the government strategy on the network of conflicts and alliances between criminal organizations.

Datasets and model specifications

Sample gathering for this analysis consisted of two stages: first, I use the dataset provided by the Uppsala Conflict Data Program (UCDP)[2], which is one of the most comprehensive datasets on violent conflicts that are publicly available and tracks violent conflicts worldwide since 1989. It consistently tracks violent conflicts in Mexico between non-state actors from 2004 to 2020. Choosing this period allowed me to observe the structure of conflicts and alliances between criminal organizations before and after the federal government decided to wage the war on drugs. The initial sample included 7442 violent events between 43 criminal organizations, accounting for

[2] For more information on this dataset visit https://ucdp.uu.se/

an estimated 59,139 violent deaths. I filtered out violent conflicts that included organizations that had no proven links to drug trafficking or other organized crime activities. Such is the case of some civilian organizations or vigilante groups, even though reports have linked some of their members with groups such as La Familia Michoacana and Los Caballeros Templarios (Herrera, 2019). In this part of the chapter I only included organizations of which I had enough information about their conflicts and alliances. My final sample to test for structural balance includes a total of 5,302 violent events between 11 drug trafficking organizations, from 2004 to 2020, which represents over 70% of the violent events in the original dataset and accounts for 40,055 violent deaths. These organizations are: Sinaloa Cartel, Cartel Jalisco Nueva Generacion (CJNG), Beltran Leyva Organization (BLO), Juarez Cartel, Los Caballeros Templarios, La Familia Michoacana, Tijuana Cartel, Los Zetas, Gulf Cartel, Cartel Independiente de Acapulco (CIDA), and Guerreros Unidos. Each event identifies a dyad of two rival organizations, and each year represents one subset (17 in total).

The second stage of my data collection was focused on using archival methods to reconstruct any positive relationship between these organizations. Because the UCDP only allows me to reconstruct violent conflicts, I had to make use of historical archives, news media outlets, and government reports, to validate every violent conflict presented in my dataset, as well as alliances between these organizations for each year[3]. In particular, I used three credible sources of investigative journalism: Insight Crime, Animal Politico, and Borderland Beat, as well as two of the most renowned newspapers in Mexico: El Universal, and El Reforma. I also used secondary sources from journalists and academics whose systematic and in-depth work helped me understand

[3] I used Python to run all my analysis and to graph the relationships between criminal groups. In particular, the Networkx package allowed me to create, manipulate, and study the structure of these relationships.

the complex relationships between some of these organizations during the last two decades. By using social network analysis and building on structural balance theory, I was able to model the structure of conflicts between these Mexican drug trafficking organizations from 2004 to 2020. Negative relationships are represented by red lines connecting nodes when such conflicts are violent, yellow lines connecting nodes when such conflicts are non-violent, and green lines connecting nodes when there is a positive relationship (alliances). Nodes represent drug trafficking organizations. Only green edges represent positive relationships, while yellow and red represent negative relationships.

A key argument in this chapter is that the kingpin strategy pursued by the Mexican government had an adverse effect on the levels of violence and the fragmentation of drug trafficking organizations. A similar argument, focusing on the militarization of public safety and the decentralization of the political system, has been elaborated by several scholars (Atuesta and Ponce, 2017; Herrera, 2019; Rios and Shirk, 2012; Trejo and Ley, 2020; Atuesta and Perez-Davila, 2018). To test my argument, I created a second dataset based on the same sample of relationships between 11 criminal organizations from 2004 to 2020. The dataset consists of 187 observations and I use organization/year as my unit of analysis. "Fragmentation" is the dependent variable, and it represents whether a given organization suffers fragmentation in any given year from 2004 to 2020. The outcome is binary:1 for "Fragmentation" and 0 for "No fragmentation". The model includes 4 explanatory variables: "Military operation", represents whether there is evidence that the federal government was at war with a criminal organization during a particular year. Specifically, I consider military operations against a specific city or state where one or more criminal organizations are present. I also consider specific government actions by the federal police against a specific criminal organization. An example of these military operations is "Operativo

Baja California" (Operation Baja California), sanctioned in 2007, to send the military to the state of Baja California, which was considered the bastion of the Tijuana Cartel. "Military operation" is also a binary variable: 1 for "Military operation" and 0 for "No military operation the organization"; "Border control" represents whether an organization controlled a border city during a given year. It is a binary variable: 1 for "controls a border city" and 0 for "does not control border city"; "War" represents whether an organization was at war with another organization during a given year. It is also a binary variable: 1 for "war" and 0 for "no war"; "Beheading" represents whether the leader of an organization was killed or extradited in a given year. It is a binary variable: 1 for "leader extradited or killed" and 0 for "leader not extradited nor killed"; finally. I fitted a nested model with four regressions: the first is a logistic regression with "Fragmentation" as the dependent variable, and "Beheading" as the independent variable; the second, third, and fourth models are multivariate logistic regressions with "Fragmentation" as the dependent variable.

The role of the government is crucial when analyzing violent conflicts between drug trafficking organizations, but rather than understanding the role of the government only in opposition to that of organized crime groups, I side with Trejo and Ley (2018), and Herrera (2019), who demonstrate that criminals and state agents collude and create a 'gray zone of criminality' that allows for organized crime to coexist with the state. Criminal organizations in Mexico have historically operated with the protection of government agents, but these relationships became more complex with the decentralization of political power in the 2000s. Slack and Campbell (2016), for instance, argued that the Mexican clientelist political system created the conditions for violence to erupt when the one-party system ended. This has led criminal organizations to decentralize their power structures into regional organizations and to seek the protection of local and state governments. Such findings show a complex scenario where government institutions can work with criminal groups in a non-cohesive fashion while pursuing conflicting agendas between the local, state, and federal levels, and sometimes even within the same institution. The assumption of this paper is that, for every conflict between two drug trafficking organizations that is analyzed, local governments might be embedded to some capacity, directly or indirectly in those conflicts, and that the kingpin approach is a strategy exclusively used by the federal government with the help of the military.

The “Kingpin strategy”: fragmentation and structural imbalance

In this section, I test the hypothesis that there is an association between the kingpin strategy and the fragmentation of criminal organizations. Figure 1. shows the Phi Correlation Coefficients between the dependent variable and all the proposed predictors[4]. The correlation matrix shows that there is a strong positive correlation between the beheading of an organization and its fragmentation (.96). Next, I fit a simple logistic regression and three multivariate logistic regressions. Table 2. shows the results of my nested model[5]. The simple logistic regression (Model 1) shows that extraditing or killing the leader of an organization (‘Beheading’) is associated with an increase in the log odds of its fragmentation by .75 ($p<.01$), compared to when there is no ‘Beheading’. Similarly, the multivariate logistic regressions (Models 2 and 3) show that, holding all other predictors constant, the ‘Beheading’ of an organization is associated with an increase in the log odds of its fragmentation by .75($p<.01$), compared to when there is no ‘Beheading’ present. Finally, the results of Model 4 should be considered with caution because one of its predictors

[4] The Phi coefficient is a statistical measure used to analyze the association between two binary variables. The interpretation is similar to that of the Pearson correlation (Osborn, 2006).
[5] I used several R packages to fit both the simple logistic and the multivariate logistic regressions.

('War') is highly correlated with 'Military Operation' (.92) and 'Border Control' (.73), and multicollinearity might be present. The results of the simple logistic regression and the multivariate logistic regression are consistent with Atuesta and Ponce (2107), who found a causal relationship between interventions by security forces in Mexico and an increase in the number of criminal organizations.

Table 1. Contingency Table

	Fragmentation		Total
	0 (N=169)	1 (N=18)	
Military_Operation			
0	56 (100%)	0 (0%)	56 (30%)
1	113 (86%)	18 (14%)	131 (70%)
Total	169 (90%)	18 (10%)	187 (100%)
Beheading			
0	163 (100%)	0 (0%)	163 (87%)
1	6 (25%)	18 (75%)	24 (13%)
Total	169 (90%)	18 (10%)	187 (100%)
Border_Control			
0	90 (93%)	7 (7%)	97 (52%)
1	79 (88%)	11 (12%)	90 (48%)
Total	169 (90%)	18 (10%)	187 (100%)
War			
0	44 (100%)	0 (0%)	44 (24%)
1	125 (87%)	18 (13%)	143 (76%)
Total	169 (90%)	18 (10%)	187 (100%)

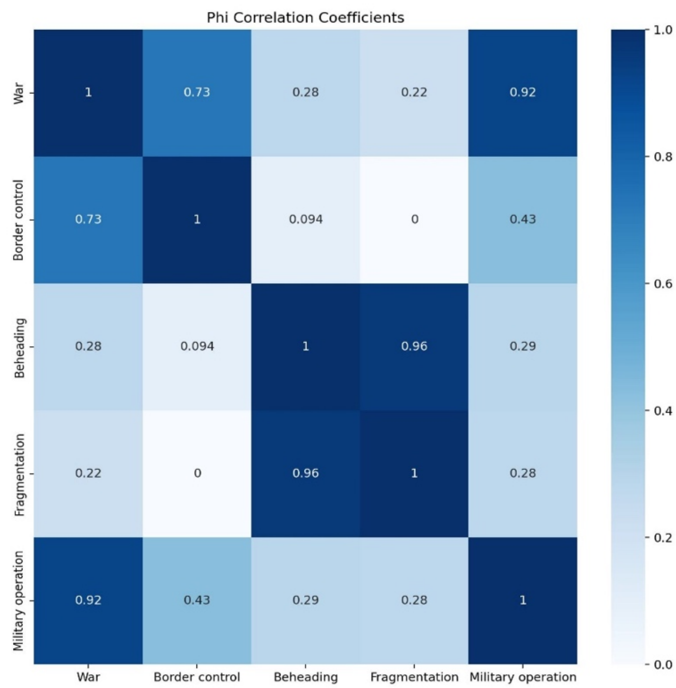


Figure 2. Correlation Matrix

Table 2. Effects of four predictors on the fragmentation of criminal organizations

Logistic Regression Results

	Fragmentation			
	1	2	3	4
Beheading	.75***	.74***	.74***	.75***
	(.03)	(.03)	(.04)	(.04)
Military operation		.02	.02	.05
		(.03)	(.03)	(.04)
Border Control			−.01	−0.00
			(.02)	(.03)
War				−.04
				(.05)
Constant	0.00	−.01	−.01	−0.00
	(.01)	(.02)	(.02)	(.02)
Observations	187	187	187	187
Log Likelihood	82.14	82.45	82.57	82.87
Akaike Inf. Crit.	−160.27	−158.90	−157.13	−155.73
Notes:	*P < .05 **P < .01 ***P < .001			

If beheading increases the likelihood of an organization splitting, we should observe larger networks of criminal organizations after 2007 in Mexico, which is when the federal government started its war on drugs and focused its efforts on a kingpin strategy. Focusing on the same sample of 11 organizations, I test my second hypothesis: that the network of conflicts and alliances between criminal organizations was structurally balanced before the government waged war against them, but the kingpin strategy led to the fragmentation of many of these organizations, which ultimately led to an unbalanced structure. Table 3 summarizes the results of the structural balance analysis. Two distinctive periods can be observed: the first, from 2004 to 2010, shows structural balance because all triangular relationships are balanced; the second from 2011 to 2020, does not show structural balance, but the ratio of balanced triads is always higher than that of unbalanced triads, which can be understood as "partial balance" (Aref and Wilson, 2018). In general, the number of violent edges is lower than that of non-violent edges for almost every year (except for the first four years), and as the number of edges increases, the relative number of violent edges seems to decrease. This means that the number of violent conflicts is relatively small compared to what it could potentially be, even when the number of organizations in the network increase.

Table 3. Balanced and unbalanced triangular relations

between drug trafficking organizations.

Year	Nodes	Edges	Violent edges	Conflicting edges	Triangular relations	Balanced	Unbalanced	Degree of Balance	Balance
2004	4	6	3	3	4	4	0	1	Structural
2005	4	6	3	3	4	4	0	1	Structural
2006	4	6	3	3	4	4	0	1	Structural
2007	4	6	3	3	4	4	0	1	Structural
2008	5	10	3	4	10	10	0	1	Structural
2009	5	10	3	4	10	10	0	1	Structural
2010	6	15	6	8	20	20	0	1	Structural
2011	9	36	8	21	84	56	28	0.666666667	Partial
2012	10	45	9	24	120	80	40	0.666666667	Partial
2013	10	45	7	24	120	80	40	0.666666667	Partial
2014	10	45	7	24	120	80	40	0.666666667	Partial
2015	10	45	8	23	120	80	40	0.666666667	Partial
2106	10	45	8	23	120	80	40	0.666666667	Partial
2017	10	45	5	23	120	80	40	0.666666667	Partial
2018	9	36	6	17	84	58	26	0.69047619	Partial
2019	9	36	6	17	84	58	26	0.69047619	Partial
2020	9	36	6	17	84	58	26	0.69047619	Partial

The period from 2004 to 2007, before the war on drug trafficking, follows a configuration of violent conflicts that is structurally balanced, with a small number of conflicting organizations that controlled large territories: The Sinaloa Cartel, Juarez Cartel, Gulf Cartel, and the Tijuana Cartel. This is consistent with what historians and experts assert on drug-related violence during this period (Jones, 2013; Hernandez, 2010; Escalante, 2009; Guerrero, 2009; Serrano, 2010; Blancornelas, 2010). When the Guadalajara Cartel disintegrated, in the late 1980s, three organizations emerged: The Sinaloa Cartel, The Tijuana Cartel, and the Juarez Cartel (the Gulf Cartel already existed), and violent conflicts between them quickly ignited. By the early 2000s, the Sinaloa Cartel started to consolidate as the most powerful drug trafficking organization in Mexico, and probably the world. As a result, the rest of the organizations forged alliances to fight off the Sinaloa Cartel. Figure 3. shows the conflicting relationships and alliances between these four organizations between 2004 and 2007. There is a total of 4 nodes (organizations) and 6 edges (relationships), 3 of which are negative (violent). All four of the 4 triangular relationships are structurally balanced.

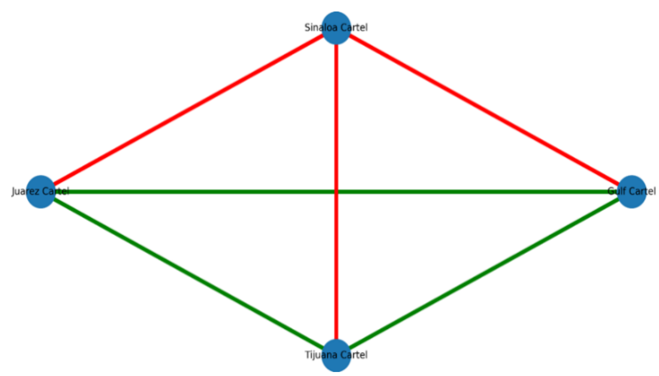


Figure. 3. Structural balance in conflicts and alliances between the most powerful drug trafficking organizations between 2004 and 2007. Red edges represent violent conflicts and green edges represent alliances. Only red edges represent negative relationships.

In 2007, the federal government declared a war on drug trafficking organizations and started a military campaign to behead and disarticulate the most important cartels. As a result, in just a few years, the leaders of some of the most important cartels started to be apprehended and extradited or killed in shootouts with the government. Their absence quickly ignited internal disputes over who would replace them and fragmentation followed. The 'balanced arrangement', observed between organizations from 2004 and 2007, started to experience important changes: in 2008, when the Beltran Leyva Organization (BLO) appears for the first time as another important player in the network of violent conflicts. The BLO was originally an organization working under the Sinaloa Cartel. They were in charge of providing security and building a team to fight off one of Sinaloa's main rivals in the early 2000s, the Gulf Cartel. But in 2008 internal disputes led the two organizations to split and started a bloody war that spilled over several states and that lasted over a decade. During this same period, the Tijuana cartel experienced several blows to its leadership: first, in 2002, when Benjamin Arellano Felix was captured by the federal government. His brother, Ramon, was killed a few months earlier in a shootout with the police. Then, in 2006, another brother and successor to the leadership of the organization was captured with the help of the DEA (Blancornelas, 2010). Finally, in 2008, the Tijuana cartel also split and two factions emerged. The second faction was backed by the Sinaloa Cartel who wanted a foothold in Tijuana.

The networks of conflicts and alliances between drug trafficking organizations in 2008 and 2009 are exactly the same (Figure 4). They both have a total of 5 nodes (organizations) and 10 edges, 3 of which are violent and 4 are conflicting edges (violent and non-violent). There are 10 triangular relationships in both networks. 10 of these relationships are balanced and 0 are unbalanced, which means that both networks are structurally balanced.

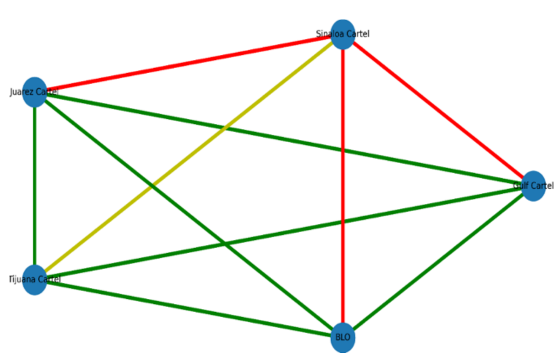


Figure. 4. Conflicts and alliances, in 2008 and 2009, are structurally balanced. Red edges represent violent conflicts, yellow edges represent non-violent conflicts, and green edges represent alliances. Only red edges represent negative relationships.

But this balanced arrangement between criminal organizations did not last for long. The year 2010 marked a shift in the configuration of conflicts and alliances: The Zetas split from the Gulf Cartel and started a violent war that spilled over the whole country. The Zetas were formed in 1999, when Osiel Cardenas Guillen, a notorious leader of the Gulf Cartel, lured more than thirty soldiers from the Mexican special forces to work for him (Castellanos, 2013). Cardenas Guillen was captured by the government and extradited to the U.S. in 2007. His brothers, Antonio and Eduardo Costilla, took charge. But in 2010 Antonio was gunned down by the military and Eduardo took charge of the Gulf Cartel. This created an internal fraction and the Zetas decided to start their own operations, which fueled a war with the Gulf cartel. Dissatisfied, the Gulf cartel forged an alliance with the Sinaloa cartel and the Familia Michoacana to fight off the Zetas (Correa-Cabrera, 2017).

In the state of Michoacan, La Familia Michoacana had been fighting off the Zetas for several years, but in 2010 their leader, Nazario Moreno Gonzales was allegedly gunned down in a shootout with the federal police. This led to a split among cartel leaders. One faction stayed with La Familia Michoacana, and the other created a new organization: Los Caballeros Templarios. They called themselves a “self-defense” movement focused on expelling other organizations from Michoacán and took the name from the medieval military-religious order the Knights Templar (Avalos, 2022). Another vicious organization, Cartel Jalisco Nueva Generacion (CJNG), also emerged in 2010, after former Sinaloa Cartel capo Ignacio Coronel was killed by the Mexican military. The CJNG has been associated with the use of extreme forms of violence against enemies, civilians, and state agents alike. By 2011 they were actively engaged in violent conflicts with other organizations. For instance, they claimed authorship of a massacre of 35 people in Veracruz that same year. They also killed around 15 police officers during an ambush, in 2015 (Avalos, 2022).

The network of conflicts and alliances shows significant changes in 2011. Figure 5. shows a total of 9 nodes and 36 edges, 8 of which are violent, and 21 are conflicting edges. The total number of triangular relationships is 84, of which 55 are balanced, and 29 are unbalanced. This is the first year in which the network does not show structural balance. The degree of balance is .654, which accounts for 'partial balance'.

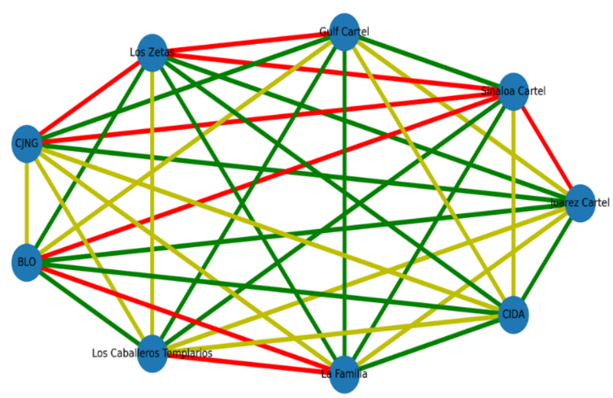


Figure 5. Network of violent conflicts in 2011. The degree of balance is .654, which accounts for 'partial balance'. Red edges represent violent conflicts, yellow edges represent non-violent conflicts, and green edges represent alliances. Only red edges represent negative relationships.

The period from 2012 to 2017 shows almost the exact same network of conflicts and alliances between these organizations with a total of 10 nodes and 45 edges (See Figure 6). For instance, the year 2012 shows 9 violent edges and a total of 24 conflicting edges (violent and non-violent). Although 2013 and 2014 show only 7 violent edges, the total of conflicting edges continues to be 24. The years 2015 and 2016 both show 8 violent edges and a total of 23 conflicting edges (violent and non-violent). Finally, 2017 shows 5 violent edges and 23 conflicting edges. Despite these small variations, all the networks in this period show a degree of balance of .683. This is because 82 triadic relationships are balanced, while only 38 are imbalanced.

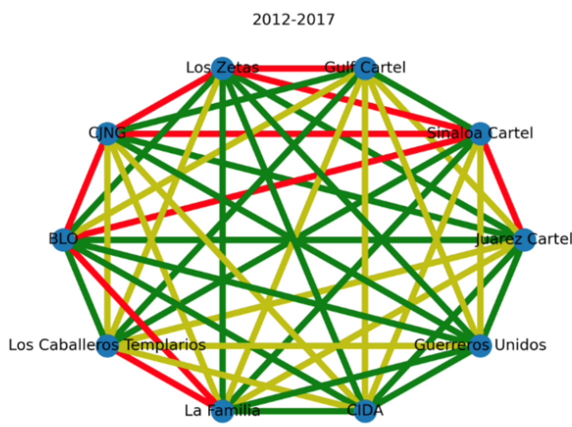


Figure 6. Network of conflicts and alliances between 2012 and 2017. Red edges represent violent conflicts, yellow edges represent non-violent conflicts, and green edges represent alliances. Only red edges represent negative relationships.

The period from 2018 to 2020 follows a similar network to the previous period. The most notable difference is the absence of Los Caballeros Templarios, who dissolved into small cells in 2017 (Dittmar, 2020). These three networks all have 9 nodes and 36 edges, 6 of them are violent, and 17 are conflicting nodes (violent and non-violent). For all these years the degree of balance is .69, with 58 balanced triads and 26 unbalanced triads (See Figure 7).

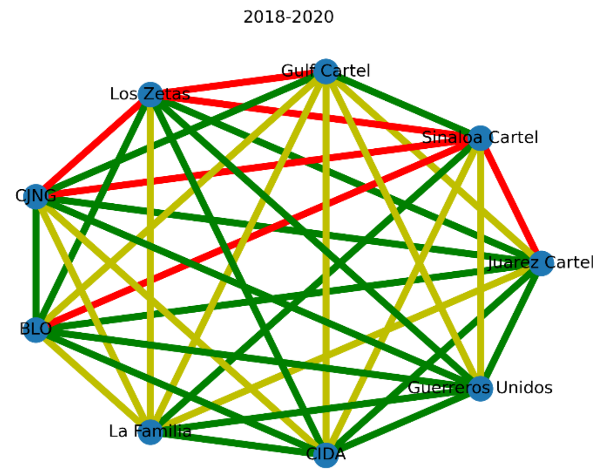


Figure 7. The network of conflicts and alliances between 2018 and 2020 shows a partial balance of .69. Red edges represent violent conflicts, yellow edges represent non-violent conflicts, and green edges represent alliances. Only red edges represent negative relationships.

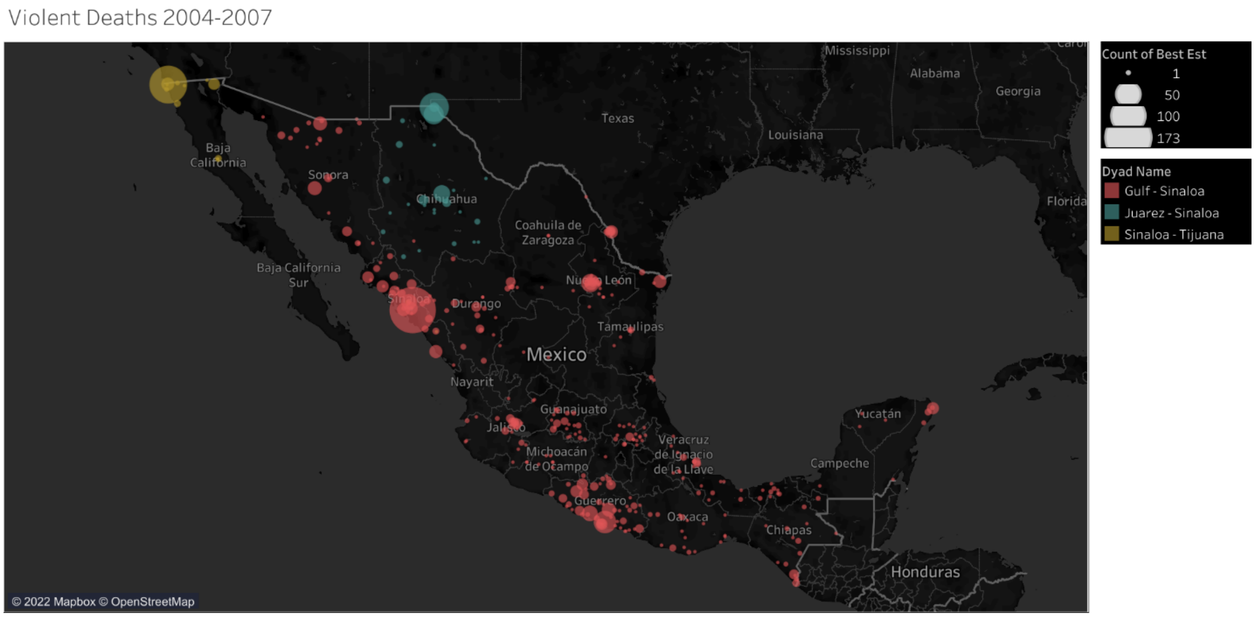
This analysis suggests that, after 2010, this network of criminal organizations stopped being structurally balanced. As the federal government continued beheading drug trafficking organizations across the country, the balanced arrangement that existed between them became increasingly more difficult to handle, until it reached a tipping point in 2010. After that year, every year in the sample shows networks that are structurally imbalanced.

Fragmentation and clustering of violence

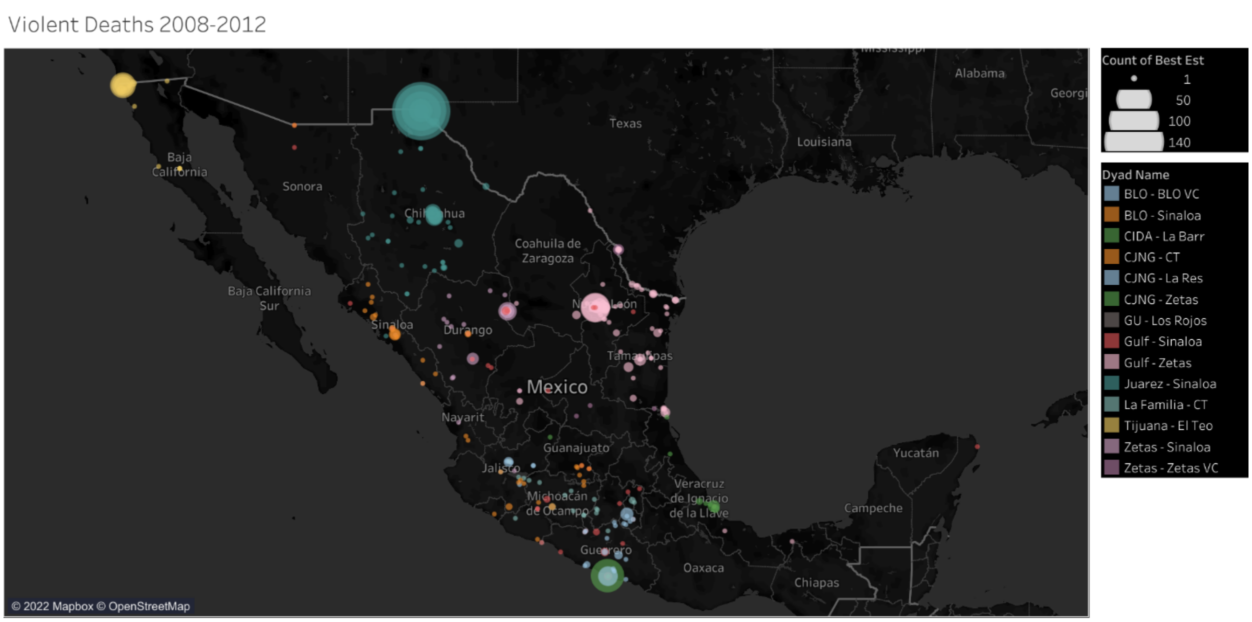
In the last section, I showed that the kingpin strategy was associated with the fragmentation of criminal organizations and that it disrupted the balanced arrangements in the network of conflicts and alliances. In this section, I test my third hypothesis: that the fragmentation of criminal organizations was associated with a 'regionalization' or clustering of violence. I use the original dataset of 7442 violent events between 43 criminal organizations, accounting for an estimated 59,139 violent deaths. The results show that, from 2004 to 2007, there is a total of four organizations partaking in 3 violent conflicts. Much of the violence happened in border cities and along drug trafficking routes. Cities like Tijuana and Ciudad Juarez were of particular importance because they were important entryways into the United States, and drug trafficking organizations have historically fought to gain or maintain their control. Likewise, ports like Acapulco and Mazatlán are vital because they allow organizations to control big shipments of drugs and precursors into Mexico (See Map 1). The number and the spread of violent events between the Sinaloa and the Gulf cartels are remarkable, as they were both striving to take control of most of the country and trafficking routes. By 2012, when Calderon's presidency ended, the number of violent conflicts had increased to 14, and the number of organizations partaking in the violence reached 19[6]. There is also a gradual clustering of conflicts, with a notable increase of violence in central and southern Mexico, specifically The Bajio region and Tierra Caliente. Veracruz continues to be a violent state, with conflicts between CJNG and the Zetas (See Map 2).

Map 1. Geography of deaths related to violent events between criminal organizations (2004-2007)

[6] Some authors argue that the number of groups reached almost 60 (Trejo and Ley, 2021).

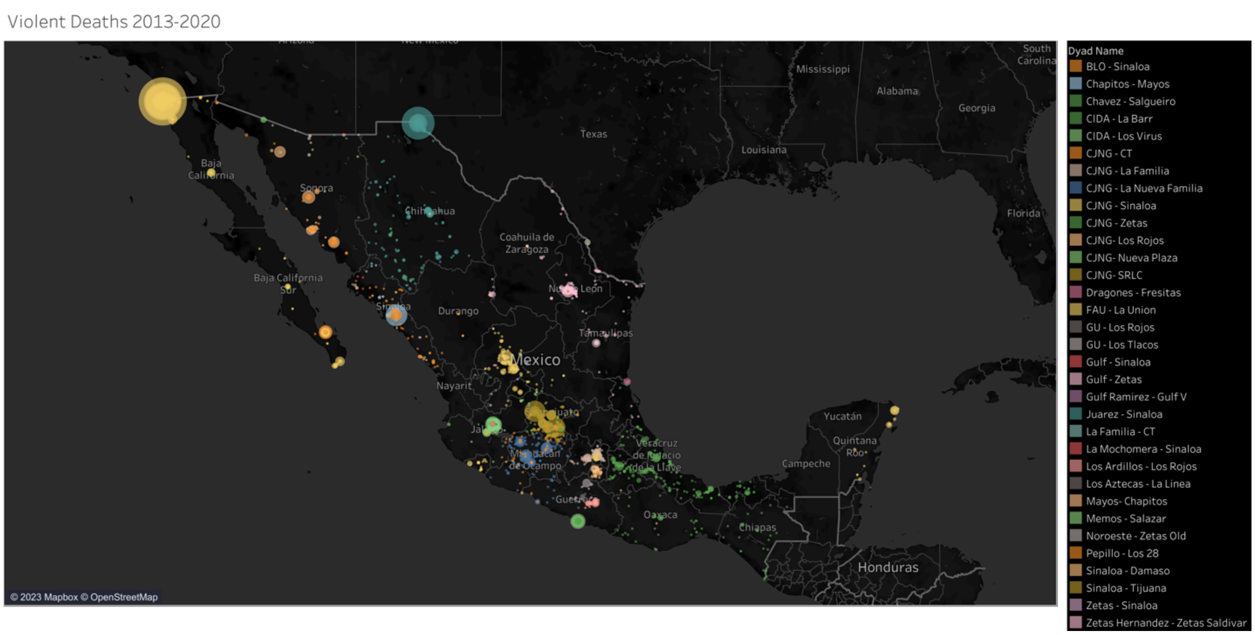


Map 2. Geography of deaths related to violent events between criminal organizations (2008-2012)



When former president Enrique Pena Nieto took office, in 2012, he promised to end the violence and the militarization of the country, but the use of the military continued and so did the kingpin strategy. In 2018, President Andres Manuel Lopez Obrador also promised to end the militarization of the country, but he has so far shown nothing but the same old strategy. The consequences of how both administrations have handled organized crime in Mexico can be observed in the new geography of violence. The period from 2013 to 2020 shows a further spillover of violent conflicts across the country, with Oaxaca and Chiapas becoming part of the contested states. There is a total of 34 violent conflicts between over 40 organizations. Despite the multiplicity of organizations fighting each other in different parts of the country, it is notable the amount of geographical clustering of these conflicts. Only two conflicts seem to reach more than one geographical region: The Gulf – Zetas, and the Sinaloa – CJNG conflicts. The vicious fighting between CJNG and Sinaloa for Tijuana's turf is also notable (See Map 3).

Map 3. Geography of deaths related to violent events between criminal organizations (2013-2020)



Map 3. shows that by 2020 there are far more organizations involved in violent conflicts than in previous years (a total of 33), but it also shows an interesting pattern of clustering of conflicts. To test for the clustering of violence, I run a modularity test to observe whether there is

an increase in the number of clusters of conflicts in the network between 2004 and 2020. In social network analysis, clusters or communities are sub-networks where a node is directly connected to any other node of the sub-network. The number of communities can be obtained through the modularity of a network. The modularity of a partition in an undirected, unweighted network is defined as:

$$Q = \frac{1}{L}\sum_{C}(Lc - \frac{k_C^2}{4L})$$

Where the sum runs over all clusters of the partition, Lc is the number of internal links in cluster C, kc is the total degree of the nodes in C, and L is the number of links in the network. I use the Louvain community detection algorithm, which allows for modularity optimization. It randomly orders all nodes in the network, and then removes and inserts each node in a different community. Networks with higher modularity will normally have dense connections within communities but sparse connections between nodes in different communities (Menczer, Fortunato, and Davis; 2020). Figure 8. gives an example of what the Louvain algorithm produces. It shows the network of violent conflicts between organizations in 2012, where each color represents a cluster such that all nodes with the same color belong to the same cluster of conflicts. The total amount of clusters for that year is 5. Finally, Figure 9. shows how the number of clusters of violent conflicts consistently increased from 1 in 2004, to 7 in 2020. This confirms the hypothesis that not only do we observe a greater number of conflicting organizations from 2004 to 2020, but that these organizations cluster together in smaller and more localized conflicts over time.

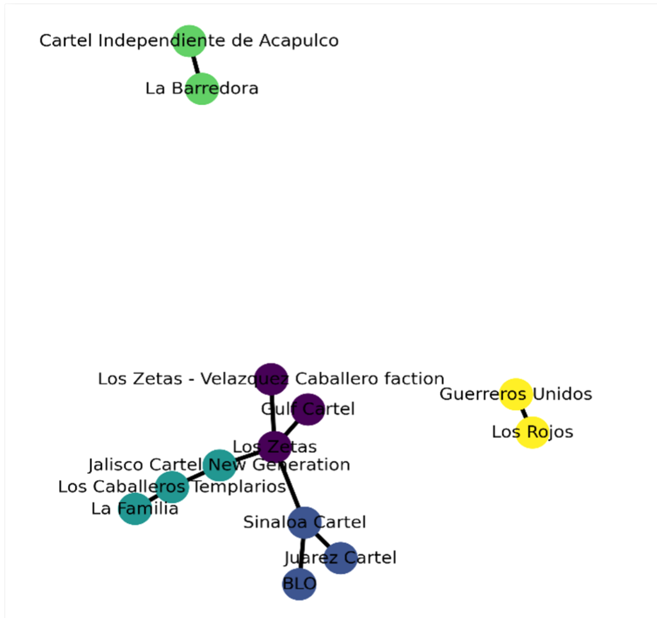


Figure 8. The Louvain Algorithm for Modularity Optimization shows a total of 5 clusters of violence in 2012

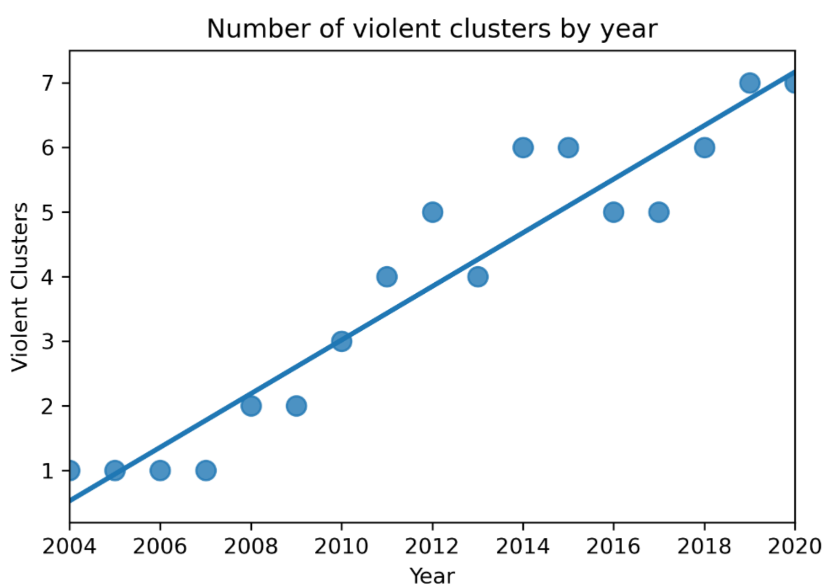


Figure 9. Clusters of Violence from 2004 to 2020

Chapter Conclusions

Several conclusions can be drawn from the present chapter. First, the results of the simple logistic regression and multivariate logistic regression models show that the beheading of an organization due to extradition or killing of the leader has a significantly positive effect on its immediate fragmentation. This is consistent with the historical evidence of how Mexico's criminal organizations suffered increased levels of fragmentation once the government started its kingpin strategy and how it led to increased levels of violence. It is also consistent with previous studies that show a positive relationship between increased levels of fragmentation of criminal organizations and violence (Atuesta and Perez-Davila, 2017).

Second, when testing for structural balance, there seem to be two distinct periods: the first from 2004 to 2010, and the second from 2011 to 2020. The first period consistently shows structurally balanced networks, while the second does not. More specifically, the period from 2004 to 2007, follows a configuration of violent conflicts that is structurally balanced, with a small number of conflicting organizations that controlled large territories. In 2007, the federal government declared a war on drug trafficking organizations and started a military campaign to behead and disarticulate the most important cartels. As a result, in just a few years, the leaders of some of the most important cartels started to be apprehended and extradited or killed in shootouts with the government. Their absence quickly ignited internal disputes over who would replace them and fragmentation followed. After 2010 the network stops being structurally balanced and more organizations become a part of it as a consequence of fragmentation. As the federal government continued beheading drug trafficking organizations across the country, the balanced arrangements that existed between them became increasingly more difficult to sustain. The year 2011 seems to be the tipping point where the network stops being structurally balanced.

An unavoidable question is whether the absence of militarization and a kingpin strategy would have sustained such a structural balance or if other variables would lead to the same effect of fragmentation and increase in violence. Regardless, both the theory and the present findings show that more conflicting organizations inevitably increase the possibility for the network not to be balanced, and unbalanced networks will tend to increase tensions within the network compared to balanced networks (Cartwright and Harary, 1956), which translates into more violence and more deaths of civilians. These results show that the kingpin strategy produced unintended consequences by disrupting the network of conflicts and alliances between criminal organizations, creating a situation prone to chaos and disorder in the network, and increasing the levels of lethal violence throughout the country.

Third, the geographical analysis of violent conflicts and the analysis of the modularity of conflicts show that the total amount of clusters of violent conflicts consistently increased from 1 in 2004, to 7 in 2020. Likewise, there is a consistent spread of violence to new regions of the country, particularly in the south. This confirms that not only do we observe a greater number of conflicting organizations, but that these organizations tend to cluster together in smaller and more localized conflicts. The trend is unambiguous and shows that the kingpin strategy not only was unsuccessful in bringing down the violence, but it also had the unintended consequence of fostering a much more complex network of conflicts with an increasing number of clusters of violence.

Literature on organized crime-related violence has shown that the Mexican clientelist political system created the conditions for violence to erupt when the one-party system ended (Rios, 2013; Herrera, 2019; Trejo and Ley, 2019; Shirk and Astorga, 2010; Slack and Campbell, 2016). Consequently, criminal organizations decentralize their power structures into regional organizations and seek the protection of local and state governments. This chapter builds on social network analysis and structural balance theory in a novel fashion to expand the argument: the kingpin strategy and the political decentralization unintentionally fostered the fragmentation of criminal organizations which increased the number of violent conflicts throughout the country. This also shifted the geography of violence and a process of clustering of conflicts. The transformation was structural because it altered the network of criminal organizations, geographical because it shifted the geography of violence, and organizational because it was associated with the fragmentation of criminal organizations.

These findings also have important public policy implications: it is clear that the kingpin strategy needs to be reimagined in light of the growing evidence of its failure, but the "how" might not be easy to answer. Going back to a structure where a few organizations controlled most of the territory and had balanced arrangements between them and the federal government is unlikely. Many new smaller criminal organizations have established arrangements with local governments and might be too embedded in the local dynamics of power, which makes uprooting them increasingly difficult. Instead of following a uniform strategy, the federal government should acknowledge the complexities and the local and specific conditions under which violence and organized crime happen. Criminal groups are multiple and heterogeneous, and it is unrealistic to think that they can all be tackled the same way. For instance, I show that some organizations are more violent than others, which can be observed by the geographical expansion of their conflicts and their central position in the network. That is the case of the Sinaloa cartel, the Zetas, and the Cartel Jalisco Nueva Generacion (CJNG). There are also some organizations that have a greater capacity to link different clusters of violence, even though they might not seem to be that central in the network. Such types of organizations function as 'brokers' in the network and can be thought of as ‘super-spreaders’ of violence (Smith and Papachristos, 2020). This line of analysis is compelling and should be revisited in any future research agenda on criminal organizations and violence.

A practical, yet controversial, public policy approach, could be to focus all military action only against those organizations that are so violent that represent an imminent threat to the safety and well-being of the population and the stability of the state. Of course, this would have to be accompanied by strict laws and actions that tackle the financial structure of criminal organizations, as well as all forms of corruption. Finally, the judicial system would have to be reformed to regain the trust of the population. A country where 93.2 % of crimes committed go unreported, and only 1.1% are resolved (INEGI, 2022) is a country where organized crime is set to thrive.

References

Aref, S., & Wilson, M. C. (2019) Balance and frustration in signed networks. Journal of Complex Networks 7, 2, 163-189. doi:10.1093/comnet/cny015

Atuesta, L.H., & Pérez-Dávila, Y.S. (2018). Fragmentation and cooperation: the evolution of organized crime in Mexico. *Trends in Organized Crime*, 21, 235-261.

Atuesta, L.H., & Ponce (2017) Meet the *Narco*: increased competition among criminal organisations and the explosion of violence in Mexico, Global Crime, 18:4, 375-402, DOI: 10.1080/17440572.2017.1354520

Ávalos, H. S. (2022, July 19). *Familia Michoacana*. InSight Crime. Retrieved November 17, 2022, from https://insightcrime.org/mexico-organized-crime-news/familia-michoacana-mexico-profile/

Blancornelas, Jesus (2010). *El Cartel*. Random House Mondadori.

Bowman, P., Previde, S., & Dittmar, P. (2021, August 12). *Las nuevas facciones criminales detrás de la violencia en México.* InSight Crime. Retrieved November 10, 2022, from https://es.insightcrime.org/noticias/algunos-principales-grupos-surgieron-fragmentacion-criminal-mexico/

Calderoni, Francesco (2014). Social Network Analysis of Organized Criminal Groups. *Encyclopedia of Criminology and Criminal Justice*. ISBN : 978-1-4614-5689-6.

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological Review, 63*(5), 277–293. https://doi.org/10.1037/h0046049

Castellanos, Guillermo (2013). *Historia del narcotráfico en Mexico*. Mexico City: Aguilar.

Correa-Cabrera, Guadalupe (2017). *Los Zetas Inc. Criminal Corporations, Energy, and Civil War in Mexico*. Austin, Texas: University of Texas Press.

Duijn, P. A. C., Kashirin, V., & Sloot Peter, ma. (2014). The relative ineffectiveness of criminal network disruption. *Scientific Reports, 4*(4238), 1–15.

Escalante, F. (2009). 'Homicidios 1990-2007', en *Nexos* 381. Septiembre, México.

Guerrero, E. (2009). 'Las tres guerras. Violencia y narcotráfico en México', en *Nexos* 381, Septiembre. México.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology: Interdisciplinary and Applied, 21,* 107–112. https://doi.org/10.1080/00223980.1946.9917275

Hernandez, Anabel (2010). Los Senores del Narco. Mexico, D.F., Grijalbo.

Herrera, J. (2019). Cultivating Violence: Trade Liberalization, Illicit Labor, and the Mexican Drug Trade. *Latin American Politics and Society, 61*(3), 129-153. doi:10.1017/lap.2019.8

Hofmann, D. C., & Gallupe, O. (2015). Leadership protection in drug trafficking networks. *Global Crime, 16*(2), 123–138.

Jones, Nathan, Irina Chindea, Daniel Weisz Argomedo, and John Sullivan. "A Social Network Analysis of Mexico's Dark Network Alliance Structure." Journal of Strategic Security 15, no. 4

(2022). https://doi.org/10.5038/1944-0472.15.4.2046.

Jones, Nathan P. "The Unintended Consequences of Kingpin Strategies: Kidnap Rates and the Arellano-Félix Organization." Trends in Organized Crime 16, no. 2 (2013): 156–76. https://doi.org/10.1007/s12117-012-9185-x.

INEGI Instituto Nacional de Estadística, Geografía e Informática (2021). [Web.] Retrieved from the Library of Congress, https://lccn.loc.gov/2005544414.

INEGI Instituto Nacional de Estadística, Geografía e Informática (2022). Retrieved April 22, 2023, from https://www.inegi.org.mx/programas/envipe/2022/.

Menczer, F., Fortunato, S., & Davis, C. (2020). *A First Course in Network Science*. Cambridge: Cambridge University Press. doi:10.1017/9781108653947

Natarajan, M., & Belanger, M. (1998). Varieties of drug trafficking organizations: A typology of cases prosecuted in New York City. *Journal of Drug Issues*, *28*, 1005-1026.

Osborn, C. E. (2006). Statistical Applications for Health Information Management. United Kingdom: Jones and Bartlett Publishers.

Rios Contreras, Viridiana. 2013. How Government Structure Encourages Criminal Violence: The causes of Mexico's Drug War. Doctoral dissertation, Harvard University

Ríos V., Shirk D. A.. 2012. *Drug Violence in Mexico: Data and Analysis through 2011*. San Diego, CA: Justice in Mexico Project.

Serrano, M. (2010). 'El problema del narcotráfico en México: una perspectiva latinoamericana', en Gustavo Vega y Blanca Torres (coord.) Los *grandes problemas de México XII: Relaciones Internacionales*, México, El Colegio de México.

Shirk, David A., and Astorga, Luis (2010). Dru Trafficking Organizations and Counter-Drug Strategies in the U.S.-Mexican Context. Evolving Democracy. Center for U.S.-Mexican Studies, UC San Diego.

Slack, J., and Campbell, H. (2016) On Narco-coyotaje: Illicit Regimes and Their Impacts on the US–Mexico Border. *Antipode*, 48: 1380– 1399. doi: 10.1111/anti.12242.

Smith, Chris & Papachristos, A.V. (2022) Violence brokers and super-spreaders: how organised crime transformed the structure of Chicago violence during Prohibition, Global Crime, 23:1, 23-43, DOI: 10.1080/17440572.2021.1998772

Smith, Chris M. & Andrew V. Papachristos, 2016. "Trust thy Crooked Neighbor: Multiplex in Chicago Organized Crime Networks. American Sociological Review 2016, Vol. 81(4) 644–667.

Trejo, Guillermo and Ley, Sandra (2018). Why Did Drug Cartels Go to War in Mexico? Subnational Party Alternation, the Breakdown of Criminal Protection, and the Onset of Large-Scale Violence. Comparative Political Studies, Vol. 51(7), 900-937.

Van De Rijt, Arnout, (2011). "The Micro-Macro Link for the Theory of Structural Balance." *Journal of Mathematical Sociology* 35(1): 94-113.

Chapter 2: How organized crime's territorial control affects undocumented migration in the U.S.-Mexico border.

Introduction

This research draws on the EMIF (Encuestas de Migracion y Frontera Norte) Survey on migratory flows from Mexico to the United States, census data from Mexico's National Institute of Statistics and Geography (INEGI), and data from the U.S. Customs and Border Protection (CBP) to understand the effects of violent conflicts on vulnerable populations, and more specifically, on undocumented migrants along the U.S.-Mexico border. I use data from 2015 to 2019 to analyze the types of hazards Mexican undocumented migrants face when crossing the U.S.-Mexico border in the context of the territorial control of drug trafficking organizations. I fit a negative binomial mixed effects model and difference in differences, as well as unsupervised machine learning, to understand how specific demographics within the undocumented migration are more vulnerable and suffer from heightened hazards in their journey and how some types of territorial control that criminal organizations exert on the border cities they control increase the hazards these migrants experience. The questions this chapter answers are "Which demographics are more vulnerable to hazardous crossings among Mexican undocumented migrants? How does the presence of organized crime groups that control the border affect the hazards these migrants encounter when crossing the border? How does the region of crossing affect the hazards migrants experience on their journey?

I argue that migrants who are unemployed at the time of crossing to the United States, are older, female, unemployed, and less educated, will have a higher probability of encountering hazards when crossing the US-Mexico border (H1). Likewise, I show that crossing the border through territories contested by two or more criminal organizations will be associated with higher levels of violence and higher hazards for migrants (H2). This happens because, compared to a contested control of a territory, criminal organizations that have full control of a territory have more capacity to control criminal activities while better protecting their smugglers and undocumented migrants. Uncontested control also allows cartels to have more control over human smugglers who might want to take advantage of migrants during their journey. In that sense, large criminal organizations like the ones controlling or fighting over the control of the US-Mexico border function as quasi-states that seek territorial control and creating protection rackets. Finally, I argue that migrants who cross the border through the region broadly defined as the "East" will encounter higher levels of hazards associated with their crossing (H3). While there is much research done on the effects of this violence on the general population (Escalante, Ley, Jones, Rios), and on the violence as a cause of internal migration and displacement (Rios, 2014; Hernandez), there is less research on the indirect effects of these violent conflicts between criminal organizations on the hazards associated with the border crossings of undocumented migrants.

The rest of the chapter is structured as follows: the first section will discuss the literature review regarding human smuggling and organized crime in the context of Mexican undocumented migration to the United States. The second section will introduce a brief historical context that describes the importance and dimension of Mexican migration to the United States during the last few decades and how drug trafficking organizations took over the human smuggling business. In the third section, I describe both datasets and the model specifications for my analysis. In the fourth section, I present the results of my models: I fit a negative binomial fixed effects model and implement two unsupervised machine-learning algorithms to underline geographical differences and the hazards migrants might encounter along the border. Finally, I developed a section with conclusions and insights for future research.

Literature Review

According to the Smuggling of Migrants Protocol of the United Nations High Commissioner for Refugees (UNHCR) people smuggling "involves the facilitation of a person's illegal entry into a State, for financial or other material benefit" (Article 3(a)). As Gabriela E. Sanchez (2015) argues, human smuggling facilitators effectively navigate the constraints of their marginalization by fulfilling an essential need within an also marginal community: the need for unregulated mobility". (p. 6). According to the author, illicit markets like smuggling provide opportunities for the exercise of agency "through immigrants' adaptation and resistance in a highly constrained context" (Sanchez, 2016, p. 60).

Some authors have argued that drug cartels and human smugglers are distinct businesses although drug trafficking has influenced migration and human smuggling (Izcara Palacios, 2012; Sanchez, 2015). For instance, Martinez (2016) shows that the familiarity of coyotes as part of the local networks of migrants has been compromised and gradually changed by unknown guides who are controlled by the more powerful drug cartels. The incentives have also changed: traditionally there was a more vertical and straightforward relationship where the migrant paid the smuggler to get them safely to the other side of the border and the smuggler had the economic incentive to get paid for a proper service. There was also the factor of trust, as many smugglers belonged to the local network of migrants, and trust and reputation played an important role in their business. Now migrants do not necessarily know or trust their smugglers, and the relationship between drug trafficking organizations and smugglers is rather complex although hierarchical. Slack and Campbel (2016) for instance show how criminal organizations overlap, interact, and self-govern, alluding to the theoretical discussion about the parallel governance of criminal organizations. They argue that now “there is a de facto hierarchy based on power and profitability of specific clandestine activities. This leads to an illicit regime of narco-governmentality, one that functions in opposition to or in collusion with the law and whose basis is survival, profitability, and evading arrest”. (p. 3). These parallel mechanisms of governance established by illicit regimes represent a form of governmentality that is enacted much like the State (Slack and Campbell, 2019).

Charles Tilly (1086) famously compared the historical role of war-making and state-making as organized crime in part due to its capacity to tax those who were ruled and protect them against external threats, and when those threats, while putting in place mechanisms to enforce such taxation. According to Von Lampe (2016) Quasi-governmental structures regulate behavior, provide protection against predatory criminals and against law enforcement, offer dispute resolution services, and in return, they tax illegal income (186). Similarly, large criminal organizations in Mexico not only can tax and control a territory, but they can also accumulate political capital and broad public support, which makes them capable of acting like quasi-states.

Violence by criminal organizations is part of an enforcing mechanism for them to control and punish those trying to evade control or taxation (Tilly, 1985). Thomas Shelling (1971) argued that the business of organized crime is the business of protection and that violence is only used when that protection fails. A similar argument is posed by Shortland and Varesse (2018) when analyzing piracy along the coastlines of Somalia. These “protectors” of piracy run profit-maximizing protection businesses that enable pirates to operate in their domains without restraints. Following this line of argument, Mexican territories that are fully controlled by a single drug trafficking organization will be expected to be less prone to the use of violence and will have more control over human smugglers and other criminal activities, which in turn translates into a better outcome for migrants. On the other hand, territories, where the control is contested by two or more criminal organizations, will show more use of violence and less control over human smugglers and other criminal activities, which will lead to higher hazards experienced by migrants in the form of assaults, abandonment or different forms of oversight and negligence.

My chapter builds on this theoretical discussion and proposes an explanatory framework to understand why some border crossings are more dangerous than others for migrants who cross with the help of smugglers along the U.S.-Mexico border, namely the type of territorial control by criminal organizations. Two types of territorial control will be discussed. Those territories controlled by a single cartel, and those where such control is contested by two or more drug trafficking organizations. The key difference lies in their capacity to produce mechanisms of control and governance.

Context

The phenomenon of Mexican undocumented migration to the United States has been widely studied and there are many estimates that gives us a picture of the size and importance of this phenomenon. For instance, according to the Migration Policy Institute (MPI), the number of Mexican unauthorized immigrants in the United States declined over the last decade from 7.8 million in 2007, to 5.3 million in 2017, and 5.2 million in 2021. One of the major factors driving down the overall population was that more left the U.S. than arrived. Despite this decline, Mexicans still represent a much larger percentage of all undocumented immigrants than any other nationality (46% in 2021). Despite the relative decrease in the number of undocumented migrants entering the United States, there still is a large demand for smugglers by Mexican immigrants, especially given the increased border securitization of the border by U.S. authorities, which forces migrants to cross through more remote and hazardous terrains.

In this context, human smuggling has been a consequence of the long process of continuous escalation of border control efforts by both countries to stem drug trafficking and undocumented immigration. This pushed migrants to attempt border crossing along more remote and difficult terrain and to rely more on professional smugglers (Andreas, 2022). I addition, the United States further militarized the border in response to the terrorist attacks on the World Trade Center and the Pentagon, in 2001. As border security increased, human smuggling required more sophistication and more knowledge about routes deep in the desert or other inhospitable areas and so forced migrants to rely even more on human smugglers (also known as “coyotes”). Smugglers started to lead migrants through more hazardous terrain with less surveillance (Spener, 2004; Guerette and Clarke, 2005). According to Peter Andreas (2022), the sharp escalation of border policing during the last two decades is related to a “loss-of-control narrative” which has focused partly on the growing power and wealth of Mexican drug-trafficking groups and “the development of more daring and better-organized migrant-smuggling efforts” (p. 5).

During the early 2000s, drug trafficking organizations realized this was a highly profitable business with a captive market of hundreds of thousands of migrants who cross the U.S.-Mexico border without documents every year. Dudley, Asmann, and Dittmar (2023) argue that “For years, community-based coyote networks provided a valued service, often across generations of migrants and their families. But as hazards rose, so did the need for more specialized networks that operated in less-traveled areas.” (p. 6). Higher hazards require more specialization and more infrastructure, which translates into higher fees because larger criminal organizations began to collect fees for passing through their territory. The U.S-Mexico border became particularly dangerous for migrants. "As the coyotes sought new routes, they had to expand their networks. Their new allies ranged from trusted partners to unknown criminal groups." (p. 7).

Overall, the changing nature of human smuggling has brought about negative consequences for migrants who depend on smugglers to cross the border and experience heightened hazards and throughout their journey, and this part of my research intends to address the nuanced relationship between the rise and fragmentation of organized crime groups, their struggles for territorial control, and its consequences for undocumented migration.

Dataset and Model Specifications

The goal of this chapter is to understand the different levels of hazard undocumented migrants experience when crossing the U.S.-Mexico border in the context of the territorial control that criminal organizations exert in this region. I build on the EMIF Norte survey which is done on a quarterly basis and released every year. It uses multistage sampling, which is a type of probabilistic sampling that allows for statistical inference[7]. The survey includes a different number of questions depending on the year of the survey, but the period I chose (2015-2019) has around 260 questions that cover topics ranging from their demographics, their journey to the United States, their labor status before, during, and after their journey, and their future plans once back in Mexico. In the section related to the migrants' journey, the EMIF survey also includes questions that address

[7] For more information about the EMIF Norte survey: https://www.colef.mx/emif/diseno.html

different types of hazards they faced on their last journey into the United States before being deported back into Mexico.

I focused on those respondents who said they used a coyote or smuggler to cross the border without documents up to 12 months prior to their interview. My dataset includes a total of 39 variables and 4945 observations that describe the migrants' demographics, different aspects of their journey to the border, and the hazards they encountered when crossing the border. In the section related to the migrants' journey, the EMIF survey also includes questions that address different types of hazards they faced on their last crossing into the United States before being deported back into Mexico. Those hazards are: being abandoned by a smuggler, attacked by an animal, assaulted, suffering from extreme cold or heat, drowning, falling off a cliff or hill, getting lost, car accident or vehicle asphyxia, and suffering from lack of water or food.

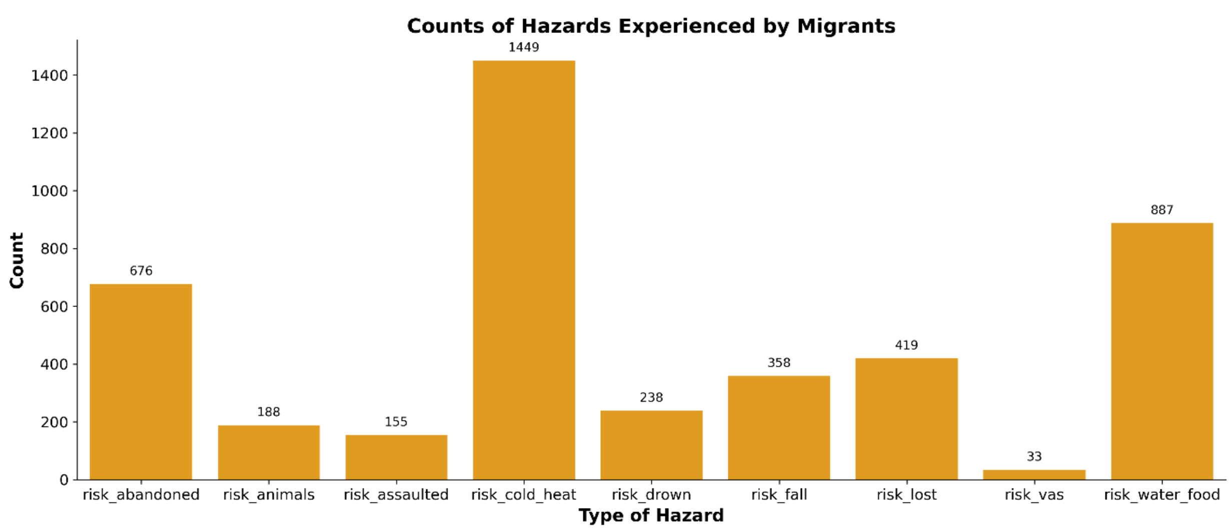


Figure 10. Distribution of Types of Hazards

H1: Contested Territories will experience higher levels of violence.

For my first hypothesis, I wanted to test the causal relationship between a territory being contested and the levels of violence in that territory, but I encountered a problem with the EMIF data: all contested territories remained contested from 2015 to 2019 and all uncontested territories remained uncontested. This lack of variability forced me to take another approach. Instead, I decided to find a different variable that could be used as a proxy of how contested a territory is. I built another dataset from the Mexican Census (National Institute of Statistics and Geography (INEGI)). This dataset includes 198 observations and seven variables that describe the homicide rates of the six Mexican states that border the United States from 1990 to 2020.

Table 4. INEGI Dataset Descriptive Statistics

	vars	n	mean	sd	median	trimmed	mad	min	max
Year	1	198	2006.00	9.55	2006.00	2006.00	11.86	1990.00	2022.00
Homicides	2	198	753.77	897.45	438.00	557.66	362.50	73.00	6407.00
State*	3	198	3.50	1.71	3.50	3.50	2.22	1.00	6.00
Contested	4	198	0.42	0.50	0.00	0.41	0.00	0.00	1.00
StateN	5	198	3.50	1.71	3.50	3.50	2.22	1.00	6.00
Population*	6	198	18.83	10.36	18.00	18.88	13.34	1.00	36.00
hom_rt	7	198	24.15	25.54	16.51	18.92	11.23	2.36	188.08

The variable “Contested” was created with the help of archival methods to reconstruct the violent conflicts between organizations along the U.S.-Mexico border in this period. I looked for any reference to violent confrontations between two or more organized crime groups. I used five credible sources of investigative journalism: Insight Crime, Animal Politico, Proceso, Sin Embargo, and Borderland Beat, as well as three of the most renowned newspapers in Mexico: El Universal, El Reforma, and Milenio. I also reviewed the work of specialists in the topic like historian Luis Astorga (2010) and journalist Anabel Hernandez (2012). If I could not find any evidence or reference to major and ongoing confrontations between two or more organizations in a state each year, I could safely assume that there was uncontested control by a single organization.

For instance, the Tijuana Cartel, led by the Arellano Felix Organization, had virtually uncontested control of the State of Baja California since its creation in 1987. Then, in 2008 after all its leaders were captured or killed, the organization splintered into two factions that fought for the control of Tijuana. Since then, no single organization has been able to fully control the state. Multiple newspapers and specialists have analyzed the different criminal groups that have tried to take over the main plazas in this state, making it a contested territory. So up until 2007, I coded that state as 0 for Uncontested, and 1 thereafter.

Next, I used a Difference in Differences (DiD) analysis to understand the effect of a territory being contested by several criminal organizations and the levels of violence in the form of homicide rate. This method estimates the effect of a specific intervention or treatment, and the basic idea is to compare the average change over time in the outcome variable for the treatment group to the average change over time for the control group.

The DiD analysis revealed a positive relationship between a territory being contested and the rate of homicides, suggesting that there is a causal relationship between these two variables. These results are consistent with literature that shows that regions where rival organizations fought for the control of territory and trafficking routes are prone to observe more violence in the form of homicides (Atuesta, 2017; Herrera, 2019; Rios, 2012). Once I established this causal relationship, I used the homicide rate as a predictor with the other dataset. Specifically, I used it as a proxy of how contested a territory is in each period.

```
Panel variable: staten (strongly balanced)
 Time variable: year, 1990 to 2022
         Delta: 1 unit

. xtreg log_hom_rt contested post treatXpost, fe

Fixed-effects (within) regression               Number of obs     =        198
Group variable: staten                          Number of groups  =          6

R-squared:                                      Obs per group:
     Within  = 0.5947                                         min =         33
     Between = 0.1524                                         avg =       33.0
     Overall = 0.3709                                         max =         33

                                                F(3, 189)         =      92.45
corr(u_i, Xb) = 0.0389                          Prob > F          =     0.0000

------------------------------------------------------------------------------
  log_hom_rt | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   contested |  -.3318126   .2339245    -1.42   0.158    -.7932508    .1296256
        post |  -.1464503   .0800925    -1.83   0.069    -.3044404    .0115398
  treatXpost |   1.424941   .2425345     5.88   0.000     .9465186    1.903363
       _cons |   2.459196   .0572191    42.98   0.000     2.346326    2.572066
-------------+----------------------------------------------------------------
     sigma_u |  .60537008
     sigma_e |  .42643759
         rho |  .66835356   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(5, 189) = 63.70                     Prob > F = 0.0000
```

Fig 11. Difference in Differences Results

H2: Contested territories are associated with higher hazards for migrants.

For my second hypothesis, I turned to the EMIF dataset[8], which includes a total of 39 variables and 4945 observations that describe the migrants' demographics, different aspects of their journey to the border, and the hazards they encountered when crossing the border. I also created two variables from the CBP data on enforcement statistics[9]. These two variables ("Aprehensions" and "Sector") describe the number of apprehensions CBP officers had with undocumented migrants along specific CBP sectors of the border. Finally, I matched geographically the different cities of

[8] For more information on the EMIF Norte dataset visit https://www.colef.mx/emif/.
[9] For more information on the CBP Enforcement Statistics visit https://www.cbp.gov/newsroom/stats/cbp-enforcement-statistics.

crossing on the Mexican side with the sectors on the US side. This specific variable allowed me to control for the effect of border enforcement by the U.S. government on the hazards faced by migrants. There is a wide literature on the negative effects of border militarization on undocumented migration and I wanted to make sure that my model was not mixing this effect with the effect of violence on the Mexican side of the border (Dunn, 2021; Slack et al., 2016; Meierotto, 2014; Michalowski, 2007).

I also decided to construct a hazard index that could be used as a measure of how hazardous these crossings were. To build my Hazard Index, I implemented a Cronbach alpha reliability test, which is a measure of internal consistency and reliability. I implemented factor analysis to test whether different configurations of hazards could yield a more internally consistent index. Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. I tested different models going from 1 to 9 factors. The Factor Analysis scree plot (Figure 5) shows the eigenvalues scored by each factor when added into the model. It seems like the model with two factors is preferable given the "elbow point". However, when analyzing the alpha scores of reliability, it turned out that none of the models was able to yield a substantially higher alpha score than the index that included all 9 factors. This index with all nine items scored .701 and .699 the raw and standardized alphas respectively, which is the threshold for an internally reliable index. I decided to use it as a predictor with all nine hazards.

Table 5. EMIF Dataset Descriptive Statistics

	count	mean	std	min	max
Year	4945.0	2016.76	1.43	2015.0	2019.0
Age	4945.0	29.38	8.58	15.0	64.0
Educ	4945.0	2.29	1.14	0.0	10.0
Experience	4945.0	1.12	2.36	0.0	43.0
Companions	4945.0	-1.32	11.05	-39.0	40.0
Male	4945.0	0.88	0.32	0.0	1.0
indig	4945.0	0.15	0.36	0.0	1.0
eng	4945.0	0.27	0.45	0.0	1.0
Smugg_decision	4945.0	0.29	0.46	0.0	1.0
Married	4945.0	0.32	0.47	0.0	1.0
alone	4945.0	0.59	0.49	0.0	1.0
Risk_Index	4945.0	0.89	1.40	0.0	9.0
Contested	4945.0	0.36	0.48	0.0	1.0
b_mod2	4945.0	0.49	0.50	0.0	1.0
City	4945.0	8.65	6.40	1.0	23.0
Region2	4945.0	0.79	0.41	0.0	1.0
Reason_easy	4945.0	0.42	0.49	0.0	1.0
Reason_origin	4945.0	0.05	0.21	0.0	1.0
Reason_destination	4945.0	0.16	0.37	0.0	1.0
Reason_family	4945.0	0.06	0.23	0.0	1.0
rel_border	4945.0	0.17	0.38	0.0	1.0
emp	4945.0	0.71	0.45	0.0	1.0
risk_cold_heat	4945.0	0.29	0.46	0.0	1.0
risk_water_food	4945.0	0.18	0.38	0.0	1.0
risk_drown	4945.0	0.05	0.21	0.0	1.0
risk_fall	4945.0	0.07	0.26	0.0	1.0
risk_animals	4945.0	0.04	0.19	0.0	1.0
risk_lost	4945.0	0.08	0.28	0.0	1.0
risk_vas	4945.0	0.01	0.08	0.0	1.0
risk_abandoned	4945.0	0.14	0.34	0.0	1.0
risk_assaulted	4945.0	0.03	0.17	0.0	1.0
Homicides	4945.0	1198.64	786.40	244.0	2978.0
hom_rt	4945.0	37.05	21.39	8.0	83.0
Ln_hom	4945.0	6.36	0.55	5.0	7.0
Region3	4945.0	0.67	0.47	0.0	1.0

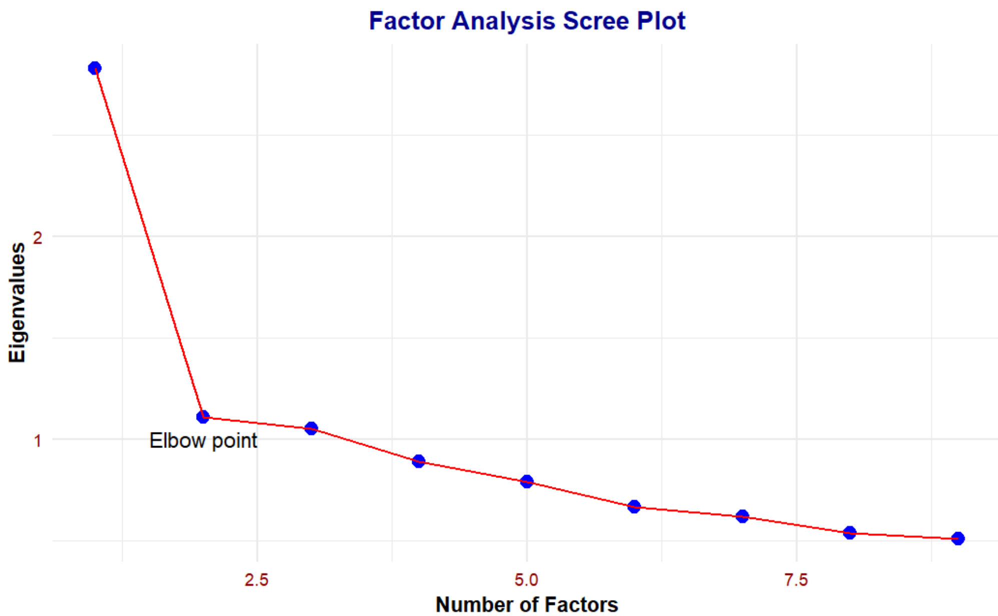


Figure 12. Factor Analysis

Because the distribution of the Hazard Index is right-skewed and shows overdispersion (meaning that the variance is larger than the mean), I decided that a negative binomial should be the best model to fit the data. However, I also wanted to control for unobserved heterogeneity associated with the crossing sites, so I decided to include fixed effects for the variable City. First, I tested Hazard Index as a function of homicide rate, controlling for demographic variables such as age, experience, number of companions, educational attainment, marital status, sex, employment, and whether they identified themselves as indigenous. This model gave me valuable information about how different demographic characteristics influenced the hazards migrants experienced when crossing the border.

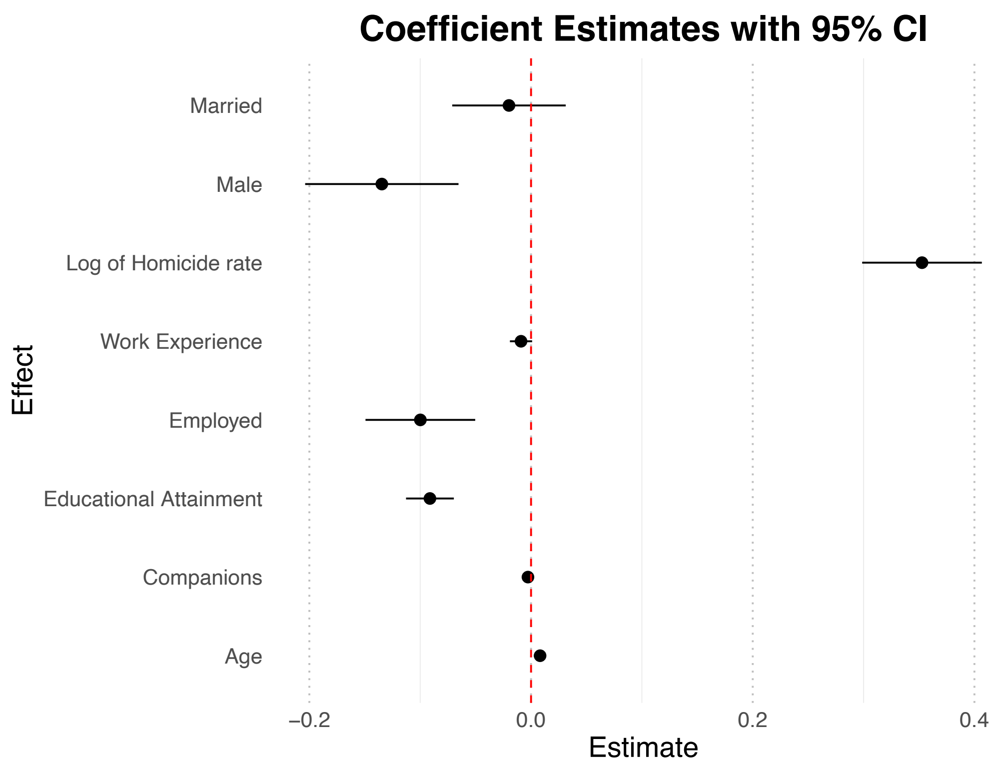


Figure 13. Results of Negative Binomial Model of Number of Hazards Experienced with Fixed Effects for "City"

The result of the model shows that migrants that are female, with lower levels of education, who are unemployed, and are older will face more hazards when crossing the border. All these characteristics can be associated with higher degrees of vulnerability. These results and hypotheses assume that lower levels of education are associated with lower levels of socioeconomic status (SES). All these variables are related to the concept of vulnerability and are used in this research as a proxy to understand differentiated vulnerabilities within the migrant population, which is already a vulnerable one. This is an important finding because undocumented migration is not homogeneous and within this population, it is possible to find even more vulnerable sub-populations that are exposed to higher levels of hazard when crossing the U.S.-Mexico border.

In terms of the homicide rate, let's remember that it is a proxy of how contested a territory is, so the results show that more contested territories are associated with a higher hazard index, and this is consistent with the idea that contested territories will negatively affect the control of drug trafficking organizations over many illegal activities, including the business of human smuggling, which in turn will increase the hazards migrants will face when crossing the border. Finally, when going through the coefficients of my fixed effects model, I observed that cities in the east of the country tended to be associated with a higher hazard index than those in the west. I decided to test a last hypothesis in relation to the regions of crossing, trying to observe if there are any regions or states along the US-Mexico border that are more dangerous than others for migrants.

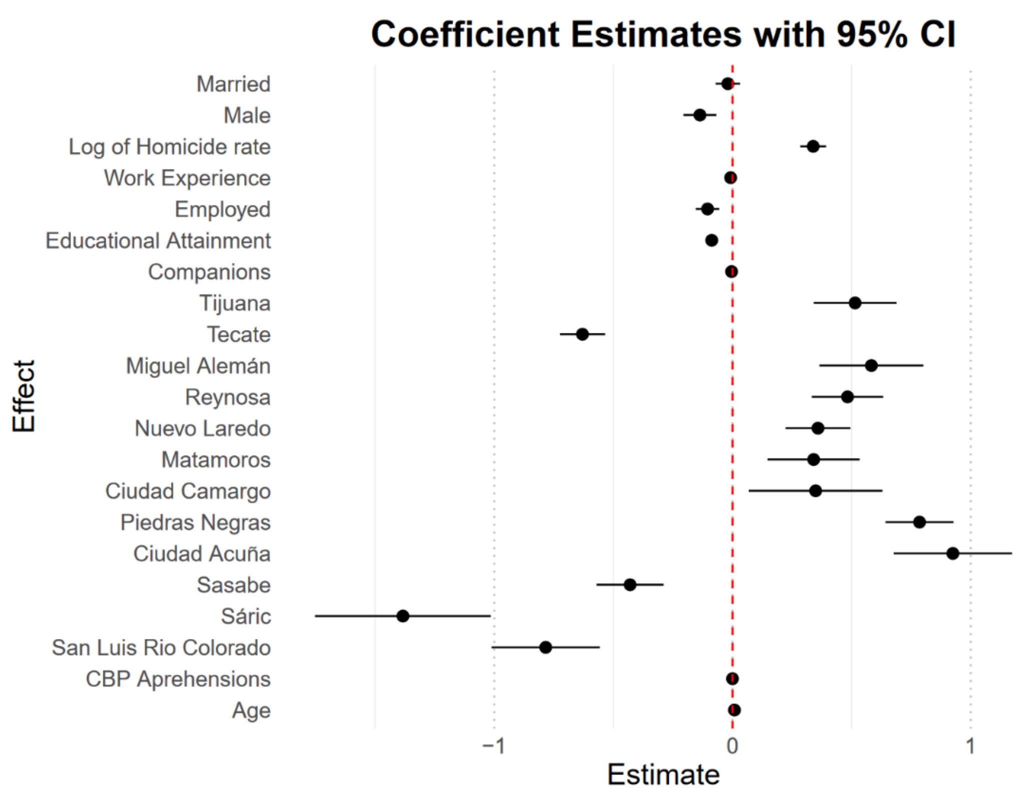


Figure 14. Territories in the "East" of the Country are Associated with a Higher Hazard Index.

H3: Certain regions of the border will be more dangerous than others.

To test my third hypothesis. I implemented two unsupervised machine learning algorithms: the K-means and Hierarchical Clustering. I gave a unique identifier to each city and implemented both algorithms considering the hazard index and the city of crossing. I implemented the Elbow Method to determine the optimal number of clusters. The basic idea is to run the clustering algorithm for a range of cluster numbers (K) and for each value, calculate the sum of squared distances from each point to its assigned center (Within Cluster sum of squares, WCSS). As more clusters are added, the WCSS tends to decrease because the points are closer to the centers they're assigned to. The rate of decrease sharply changes at a certain point, creating an "elbow" in the graph. I also implemented the silhouette analysis to confirm the optimal number of clusters. The silhouette analysis provides a graphical representation of how well each observation lies within its cluster. In this case, the silhouette analysis shows that additional clusters do not increase significantly the silhouette score, which is an indicator that 2 clusters is the best choice. I was able to devise two big regions that differentiate themselves in terms of the hazards posed to migrants. They were broadly defined as East and West.

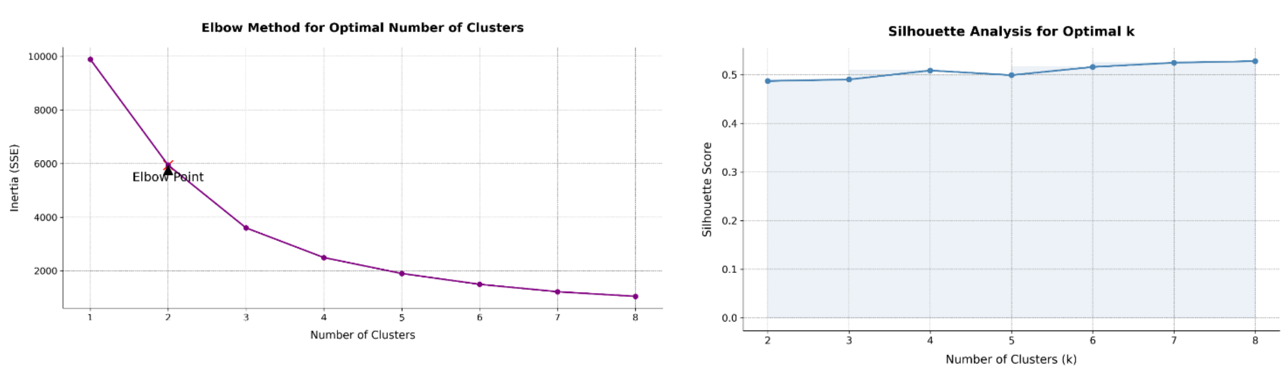


Figure 15. Elbow Metod and Silhouette Analysis

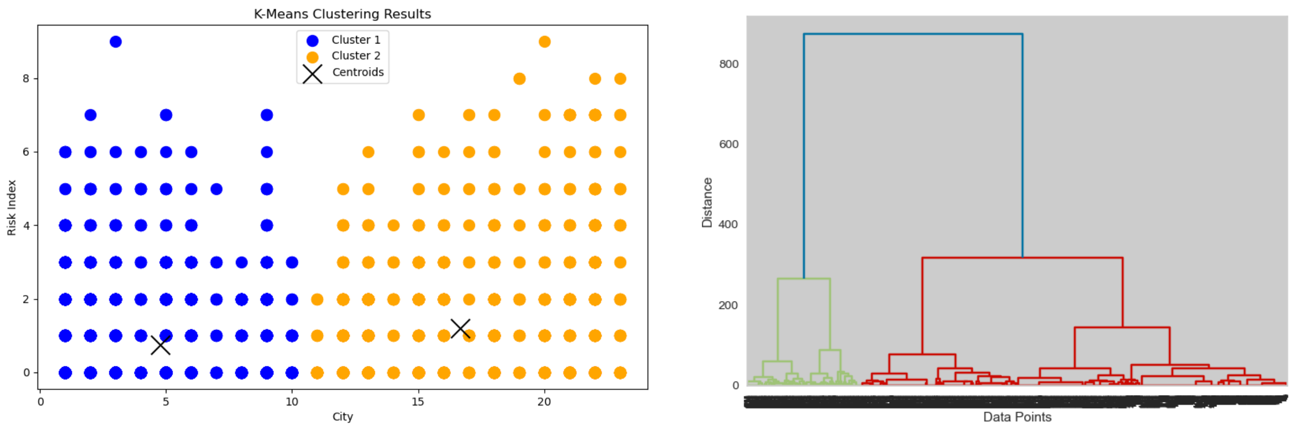


Figure 16. K-Means Clustering and Hierarchical Clustering

Finally, I implemented hierarchical clustering which is another unsupervised machine learning method. The advantage of this method over K means is that it does not require specifying the number of clusters in advance. The results are usually presented in a dendrogram, which is a tree-like diagram that records the sequences of merges or splits and helps to interpret the data by visually showing the clusters at each level of the hierarchy (Figure 11).

When analyzing geographically these two clusters, the regional division is clear, and it fits almost perfectly the division between what the Sinaloa Cartel controlled in this period and the states where no single organization had the monopoly of the use of violence. Is possible to see the region that was controlled by a single criminal organization from 2015 to 2019 (blue), which was the Sinaloa Cartel, and the region that was contested (orange) by multiple criminal organizations like Cartel del Noreste, Cartel de los Zetas, Cartel del Golfo, etc. And when looking at the distribution of the hazard index by region, there is also a clear difference between the East and the West. The East has a good portion of observations at higher values of the Hazard Index compared to the West. But

I also wanted to test if this observed difference was statistically significant. I fitted a negative binomial model of the Hazard Index as a function of the Region of crossing controlling for the same variables as in the previous model. The model summary shows that The West is negatively associated with the Hazard Index compared to the East, which is what I expected (Figure 13).

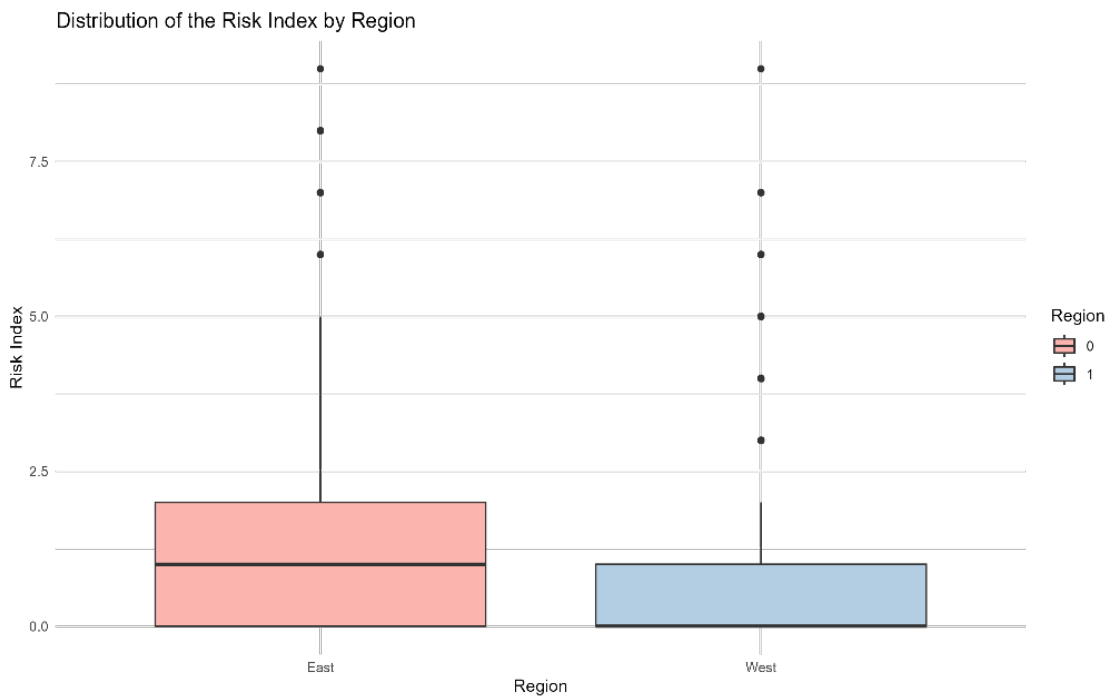


Figure 17. Box Plot of Risk Index by Region

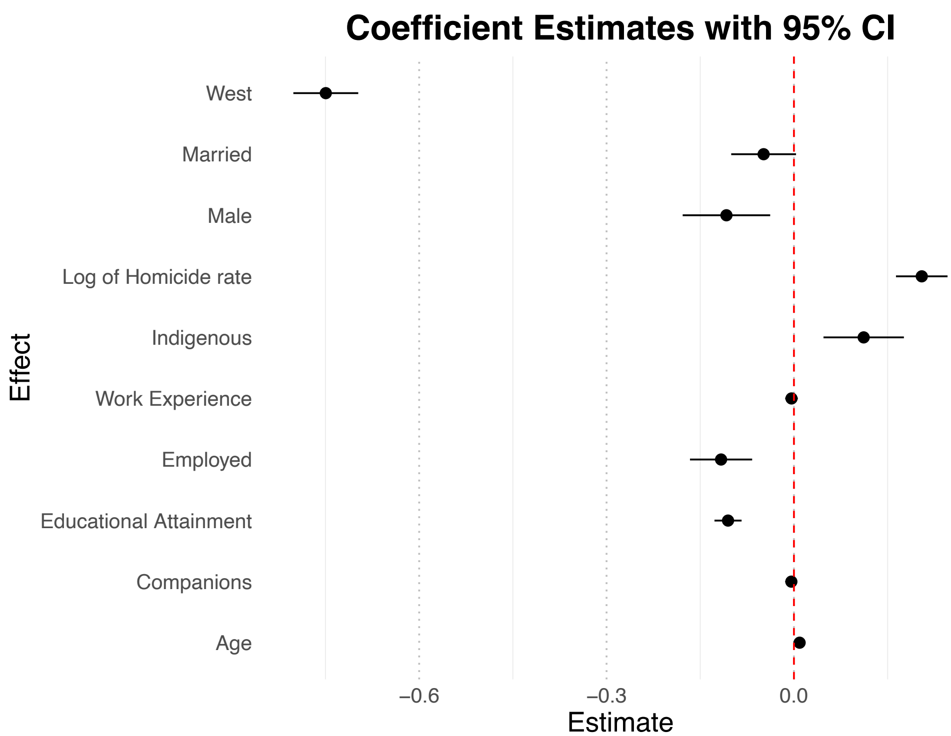


Figure 18. Hazard Index as a Function of Region of Crossing

Chapter Conclusions

This research draws on the EMIF (Encuestas de Migracion y Frontera Norte) Survey on migratory flows from Mexico to the United States, census data from Mexico's National Institute of Statistics and Geography (INEGI), and data from the U.S. Customs and Border Protection (CBP) to understand the effects of violent conflicts on vulnerable populations, and more specifically, on undocumented migrants along the U.S.-Mexico border. I used computational methods, including statistics and unsupervised machine learning, to show that territories of crossing that are contested by two or more drug trafficking organizations are associated with increased hazards for undocumented migrants. Specifically, migrants who are unemployed at the time of crossing to the

United States, are older, female, unemployed, and less educated, will have a higher probability of encountering hazards when crossing the US-Mexico border (H1). Likewise, I show that crossing the border through territories contested by two or more criminal organizations will be associated with higher levels of violence and higher hazards for migrants (H2). I argue that this happens because, compared to a contested control of a territory, criminal organizations that have full control of a territory have more capacity to control criminal activities while better protecting their smugglers and undocumented migrants. Uncontested control also allows cartels to have more control over human smugglers who might want to take advantage of migrants during their journey. In that sense, large criminal organizations like the ones controlling or fighting over the control of the US-Mexico border function as quasi-states that seek territorial control and creating protection rackets. Finally, I argue that migrants who cross the border through the region broadly defined as the "East" will encounter higher levels of hazards associated with their crossing (H3), and this and this happens because, during the period analyzed, the East was contested by competing criminal organizations, which increased the levels of violence in comparison to the West, which was mostly controlled by a single organization (the Sinaloa Cartel).

Finally, the negative binomial fixed effects model (H2) also showed that border enforcement was positively associated with higher hazards for migrants, and this is consistent with the literature that points to the relationship between the militarization of the border and the increased risks posed to migrants (Heyman and Campbell, 2012; Slack, et al., 2016). However, the purpose of the present chapter was to portray an additional side of the complex story: how the territorial control of drug trafficking organizations negatively impacts vulnerable populations, such as migrants, when such control is contested, which raises the discussion about the state-like nature of organized crime.

References

Aref, S., & Wilson, M. C. (2019) Balance and frustration in signed networks. Journal of Complex Networks 7, 2, 163-189. doi:10.1093/comnet/cny015

Atuesta, L.H., & Pérez-Dávila, Y.S. (2018). Fragmentation and cooperation: the evolution of organized crime in Mexico. *Trends in Organized Crime*, 21, 235-261.

Atuesta, L.H., & Ponce (2017) Meet the *Narco*: increased competition among criminal organisations and the explosion of violence in Mexico, Global Crime, 18:4, 375-402, DOI: 10.1080/17440572.2017.1354520

Ávalos, H. S. (2022, July 19). *Familia Michoacana*. InSight Crime. Retrieved November 17, 2022, from https://insightcrime.org/mexico-organized-crime-news/familia-michoacana-mexico-profile/

Andreas P (2000) Border Games: Policing the US–Mexico Divide. Ithaca: Cornell University Press

Astorga L (2005) El siglo de las drogas: el narcotráfico, del Porfiriato al nuevo milenio. México City: Plaza y Janés.

Blancornelas, Jesus (2010). *El Cartel*. Random House Mondadori.

Bowman, P., Previde, S., & Dittmar, P. (2021, August 12). *Las nuevas facciones criminales detrás de la violencia en México.* InSight Crime. Retrieved November 10, 2022, from https://es.insightcrime.org/noticias/algunos-principales-grupos-surgieron-fragmentacion-criminal-mexico/

Castellanos, Guillermo (2013). *Historia del narcotráfico en Mexico*. Mexico City: Aguilar.

Correa-Cabrera, Guadalupe (2017). *Los Zetas Inc. Criminal Corporations, Energy, and Civil War in Mexico*. Austin, Texas: University of Texas Press.

Campbell H (2009) Drug War Zone: Frontline Dispatches from the Streets of El Paso and Juárez. Austin: University of Texas Press.

De León J (2015) The Land of Open Graves: Living and Dying on the Migrant Trail. Berkeley: University of California Press.

Dunn, T. J. (2021). The militarization of the US-Mexico border in the twenty-first century and implications for human rights. In *Handbook on human security, borders and migration* (pp. 35-53). Edward Elgar Publishing.

Duijn, P. A. C., Kashirin, V., & Sloot Peter, ma. (2014). The relative ineffectiveness of criminal network disruption. *Scientific Reports, 4*(4238), 1–15.

Escalante, F. (2009). 'Homicidios 1990-2007', en *Nexos* 381. Septiembre, México.

Guerrero, E. (2009). 'Las tres guerras. Violencia y narcotráfico en México', en *Nexos* 381, Septiembre. México.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology: Interdisciplinary and Applied, 21,* 107–112. https://doi.org/10.1080/00223980.1946.9917275

Hernandez, Anabel (2010). Los Senores del Narco. Mexico, D.F., Grijalbo.

Herrera, J. (2019). Cultivating Violence: Trade Liberalization, Illicit Labor, and the Mexican Drug Trade. *Latin American Politics and Society, 61*(3), 129-153. doi:10.1017/lap.2019.8

Hofmann, D. C., & Gallupe, O. (2015). Leadership protection in drug trafficking networks. *Global Crime, 16*(2), 123–138.

Jones, Nathan P. "The Unintended Consequences of Kingpin Strategies: Kidnap Rates and the Arellano-Félix Organization." Trends in Organized Crime 16, no. 2 (2013): 156–76. https://doi.org/10.1007/s12117-012-9185-x.

INEGI Instituto Nacional de Estadística, Geografía e Informática (2021). [Web.] Retrieved from the Library of Congress, https://lccn.loc.gov/2005544414.

INEGI Instituto Nacional de Estadística, Geografía e Informática (2022). Retrieved April 22, 2023, from https://www.inegi.org.mx/programas/envipe/2022/.

Mendoza, Natalia (2012). Microhistoria de la violencia en Altar, Sonora. Las bases sociales del crimen organizado y la violencia en México.

Meierotto, L. (2014). A disciplined space: The co-evolution of conservation and militarization on the US-Mexico border. *Anthropological Quarterly*, 637-664.

Michalowski, R. (2007). Border militarization and migrant suffering: A case of transnational social injury. *Social Justice*, *34*(2 (108), 62-76.

Natarajan, M., & Belanger, M. (1998). Varieties of drug trafficking organizations: A typology of cases prosecuted in New York City. *Journal of Drug Issues*, *28*, 1005-1026.

Rios Contreras, Viridiana. 2013. How Government Structure Encourages Criminal Violence: The causes of Mexico's Drug War. Doctoral dissertation, Harvard University

Ríos V., Shirk D. A.. 2012. *Drug Violence in Mexico: Data and Analysis through 2011*. San Diego, CA: Justice in Mexico Project.

Serrano, M. (2010). 'El problema del narcotráfico en México: una perspectiva latinoamericana', en Gustavo Vega y Blanca Torres (coord.) Los *grandes problemas de México XII: Relaciones Internacionales*, México, El Colegio de México.

Shirk, David A., and Astorga, Luis (2010). Dru Trafficking Organizations and Counter-Drug Strategies in the U.S.-Mexican Context. Evolving Democracy. Center for U.S.-Mexican Studies, UC San Diego.

Slack, J., and Campbell, H. (2016) On Narco-coyotaje: Illicit Regimes and Their Impacts on the US–Mexico Border. *Antipode*, 48: 1380– 1399. doi: 10.1111/anti.12242.

Slack, J., Martínez, D. E., Lee, A. E., & Whiteford, S. (2016). The Geography of Border Militarization: Violence, Death and Health in Mexico and the United States. *Journal of Latin American Geography*, *15*(1), 7–32. http://www.jstor.org/stable/43964648

Tilly C. War Making and State Making as Organized Crime. In: Evans PB, Rueschemeyer D,

Skocpol T, eds. *Bringing the State Back In*. Cambridge University Press; 1985:169-191.

Trejo, Guillermo and Ley, Sandra (2018). Why Did Drug Cartels Go to War in Mexico? Subnational Party Alternation, the Breakdown of Criminal Protection, and the Onset of Large-Scale Violence. Comparative Political Studies, Vol. 51(7), 900-937.

Chapter 3: Enhancing Similarity-based Algorithms with Node2Vec and Deep Neural Network architectures to Predict Links in Mexico's Organized Crime Network[10]

Introduction

During the past two decades, the number of criminal organizations in Mexico has spiked from fewer than 10 large cartels in 2004 to more than 150 in 2020 and 387 criminal groups in 2021.[11] Several studies have argued that the militarization and kingpin strategy of the Mexican federal government led to this dramatic increase (Atuesta and Perez-Davila, 2017), while others have pointed to the process of political decentralization as one of the main factors (Herrera, 2022). Given the context of heightened violence in the country, the configuration of alliances between these organizations provides important information about the state of Mexico's organized crime (Jones et al., 2022). Therefore, some important questions arise, given the current network of alliances between these organizations: How can we better understand the structure of this network?

---

[10] This chapter is the result of a collective working paper, where I am first author, that was released in 2023 by Rice University's Baker Institute. The original working paper was written by Oscar Contreras Velasco, Nathan P. Jones, Daniel Weisz Argomedo, John P. Sullivan, and Chris Callaghan. To access the original working paper visit: https://www.bakerinstitute.org/research/use-similarity-based-algorithms-predict-links-mexican-criminal-networks

[11] These figures are sourced from Lantia Consultores, a prestigious Mexican analytical firm. See Nathan P. Jones, Irina Chindea, Daniel Weisz Argomedo, and John P. Sullivan, "Mexico's 2021 Dark Network Alliance Structure: An Exploratory Social Network Analysis of Lantia Consultores' Illicit Network Alliance and Subgroup Data," Research Paper (Houston: Rice University's Baker Institute, April 11, 2022), https://doi.org/10.25613/KMGB-NC83; Nathan P. Jones, Irina Chindea, Daniel Weisz Argomedo, and John P. Sullivan, "A Social Network Analysis of Mexico's Dark Network Alliance Structure," *Journal of Strategic Security* 15(4) (2022): 76–105, https://doi.org/10.5038/1944-0472.15.4.2046; Nathan P. Jones, W. Layne Dittmann, Jun Wu, and Tyler Reese, "A Mixed Methods Social Network Analysis of a Cross-Border Drug Network: The Fernando Sanchez Organization (FSO)," *Trends in Organized Crime* 23(2) (2020): 154–82, https://doi.org/10.1007/s12117-018-9352-9; June S. Beittel, "Mexico: Organized Crime and Drug Trafficking Organizations" (Washington, D.C.: Congressional Research Service, June 7, 2022), https://sgp.fas.org/crs/row/R41576.pdf.

Which new alliances are more likely to happen next? What is the probability that these new alliances will happen? Which metrics are better suited to make such predictions and which methods can we use to better understand the structure of these networks given the dearth of good quality information about them?

This chapter uses a combination of social network analysis (SNA) and machine learning techniques to predict links in Mexico's network of criminal organizations. Specifically, we use four similarity- based algorithms to estimate the likelihood that a link will be formed between two unconnected organizations in the network. We enhanced these four algorithms with the implementation of the Node2Vec algorithm and a Deep Neural Network Architecture (DNNs) to improve their predictive capabilities. We also included a standalone Node2Vec algorithm as well as its enhanced version with the help of a Deep Neural Network architecture. Of the node-similarity indices implemented, we found that the Preferential Attachment algorithm enhanced with both the Node2Vec and the DNN architectures performed the best when predicting existing ties. We conclude that this is in part due to the power law-like[12] distribution of the criminal network. We found that this criminal network is structured such that nodes with many connections are more likely to gain more and vice versa. This also explains why the Preferential Attachment performed in all its variants performed so well when predicting ties in this network. Such a structure can be explained by the existence of hierarchical relationships where a few organizations monopolize activities, control, and subcontract other smaller organizations. When predicting potential future ties, the best performing algorithm was the Jaccard Coefficient enhanced both by the Node2Vec.

[12] A power law-like distribution is a type of distribution where a small number of events or values occur very frequently, and a large number of events or values occur infrequently.

We argue that this happens because the Jaccard Coefficient captures the immediate neighbor overlap but does not account for longer paths or the global network structure, while the Node2Vec does capture broader topological and community-based features that might not be apparent from direct connections alone. The use of this enhanced algorithm can be very powerful to predict future network ties in any network, not only criminal ones.

The rest of the chapter is organized by sections: first, we review the literature on social network analysis and machine-learning techniques to understand criminal networks and violence; second, we provide the context of Mexico's organized crime dynamic; third, we detail the five algorithms that we use in our analysis. These comprise four similarity-based indices, four enhanced versions of these same indices using a Deep Neural Network Architecture, and an algorithm known as Node2Vec which is powered by the skip-gram model used in natural language processing; in a fourth section, we summarize the data and the methods we use to analyze the network and evaluate the performance of all four algorithms; A fifth section shows the results of our analysis and discuss the strengths of each algorithm. We determine that the Node2Vec algorithm shows the best predictive capacity given the "area under the curve" (AUC) score. The enhanced Preferential Attachment index shows the second best "area under the curve" (AUC) score, given the structure of the criminal network; finally, we lay out the conclusions and policy recommendations stemming from our research.

Social Network Analysis and Machine Learning

Social network analysis (SNA) methodologies have become crucial for developing and adapting techniques in criminal network analysis (CNA) (Lim, Abdullah, and Jhanjhi, 2021). SNA methodologies combine graph theory, which "provides the conceptual constructs, methods, and techniques for the analysis of graphs," with the "application of analytical techniques and visualization tools developed specifically for the analysis of social and other networks." (Lim et al., 2012). As Renée van der Hulst explains, "In addition to visualizations of network graphs, SNA is an arithmetical technique that analyzes relational patterns of nodes (actors) and connections (ties) based on mathematical computations." (van der Hulst, 2009).

The use of SNA to better understand criminal networks has become increasingly popular in law enforcement worldwide (Sparrow, 1991). SNA methodologies are applied to a wide range of cases, from illegal cannabis operations in the Netherlands to uncovering the unintended consequences of the kingpin strategy on the Fernando Sanchez organization (better known as the Tijuana cartel) (Dujin and Kerks, 2014). Most recently, Nathan P. Jones, Irina Chindea, Daniel Weisz Argomedo, and John P. Sullivan used SNA to demonstrate differential alliance structures within Mexico's bipolar illicit network system (Jones et al., 2022). The differences threaded out by SNA techniques helped develop specific recommendations to weaken and disrupt each unique alliance network. Overall, there is a growing field of opportunity to use SNA methodologies alongside machine-learning techniques to better understand criminal networks and develop strategies to weaken them.

Machine learning has been applied in many different fields to produce predictions. Today machine-learning algorithms are part of our daily lives (for example, in internet search results and the facial recognition software many smartphones use) (Surden, 2014). Machine-learning techniques are "a set of mathematical models to solve high non-linearity problems of different topics: prediction, classification, data association, [and] data conceptualization" (Torres, Comesana, and Garcia-Nieto, 2019). They manage to uncover generalizable patterns due to their ability to reveal complex structures that were not specified in advance (Mullainathan and Spiess, 2017). Machine learning is closely tied to "predictive analytics," as researchers utilize existing data to predict the likelihood of different outcomes. Most importantly, machine-learning algorithms are designed to improve their performance on a given task over time as they build up their library of relevant data with more examples (Surden, 2017).

Machine-learning algorithms can be either unsupervised or supervised. Most real-life applicable machine-learning algorithms use supervised variants. A supervised variant is "a prediction model developed by learning a dataset where the label is known, and accordingly, the outcome of unlabeled examples can be predicted." (Uddin et al., 2019). In an unsupervised variant, on the other hand, the data is unlabeled, and the goal is to find a hidden structure from within the data (Mohammed et al., 2016).

SNA and supervised machine-learning techniques are rapidly becoming critical tools for a number of fields, including: law enforcement and criminal justice, cybersecurity, and national security. When it comes to law enforcement, machine-learning techniques are used to predict where crimes may occur based on past criminal activity (Berman, 2018). A specific example of this occurs in Chicago, where the police department uses "predictive analytics to identify not only places that are particularly vulnerable to crime, but also people more likely to be involved in gun violence." (Berman, 2018). Researchers can also bridge SNA methodologies and machine-learning techniques to understand criminal networks better and predict their possible evolution. In addition, courts use machine-learning techniques "to establish individual risk profiles and to predict the likelihood that a particular individual will re-offend." (Deeks, Lubell, and Murray, 2019: p. 26). In cybersecurity, machine-learning techniques have been successfully applied to detect malware, identify the authorship of phishing attacks, and even detect phishing emails. (Torres et al., 2019).

Machine learning can also be applied to text data, which has been growing at a faster rate with the surge in internet and social media use. Finally, in areas related to national security, machine-learning techniques can be applied to abstract or encoded text data used to identify potential threats. (Mohammed et al., 2016). With the rise of disinformation campaigns and foreign election interference, machine-learning techniques can be used to “automatically detect, analyze, and disrupt disinformation, weed out nefarious content and block bots.” (Horowitz et al., 2018: p. 7).

Mexico’s Organized Crime Network Structure

In 2007, ex-president Felipe Calderon Hinojosa decided to focus its efforts on dismantling organized crime groups and sent the military to the streets. Along with the militarization of public safety, Calderon started a “kingpin strategy” with the hope that by beheading these criminal organizations, they would remain too weak to sustain their structure and activities. But the kingpin strategy led to further fragmentation of criminal organizations and an increase in violence. The lack of leadership in these organizations created power vacuums that led to splinter groups that now contested the territory and trafficking routes of the original organizations (Contreras, 2023).

Currently, Mexican organized crime is dominated by drug cartels and allied gangs that cause insecurity, challenge state solvency, and capitalize on illicit global economic flows (Sullivan, 2023). These cartels are poly-crime organizations that intersect with networked criminal enterprises comprising corrupt politicians and transnational criminal organizations. Jones et al. examined the relationships and alliance structures of many of these cartel-gang networks, with an emphasis on the contrasting natures of the Sinaloa cartel and the Jalisco New Generation cartel (CJNG) networks (Jones et al., 2022).

While some argue that drug cartels do not exist, the violence and insecurity resulting from criminal competition is profoundly tangible (Zavala, 2022). The violence resulting from the illicit economy (especially the drug trade) and corruption has been described in terms of crime wars and criminal insurgencies — on the one hand, criminal networks battle each other and the state and, on the other hand, alter the nature of the state and state sovereignty (Tilly, 1985; Sullivan, 2022; Sullivan and Elkus, 2008). The immediate result is a diminished state solvency, with insecurity compounded by weakened perceptions of state legitimacy, and a lack of state capacity. These dynamics play out in a battle for control of lucrative drug trafficking "plazas"[13] which often become the bases for criminal enclaves, where cartels enjoy territorial control and exercise criminal governance — often in league with corrupt government officials (Sullivan, 2019)

Similarity-based Algorithms, Deep Neural Networks, and Node2Vec

In this chapter, we compare the performance of four of the most famous and widely used similarity-based indices. We focus on these algorithms because of their capacity to capture different forms of social capital, such as trust and expectations (Coleman, 1988), and access to information (Burt, 2000). All these forms of social capital are fundamental for alliances to form between criminal

[13] "Plaza" refers to a specific city or geographic location along the U.S.-Mexico border that is used to smuggle illicit drugs from Mexico into the United States: https://www.justice.gov/archive/ndic/pubs32/32781/dtos.htm (note 9).

organizations and can be good predictors of new relationships. Below are brief descriptions of each of the algorithms used.

Adamic-Adar Index

The Adamic-Adar algorithm is an improvement of the simpler algorithm that counts common neighbors. The Common Neighbors algorithm assumes that nodes i and j are more likely to have an edge (or link) if they have many common neighbors. The Adamic-Adar index refines simple counting by assigning more weight to less-connected neighbors (Zhou, Lu, and Zhang, 2009) and is defined as:

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log k(z)}.$$

where Sxy is the similarity score between the nodes x and y, and Γ(x) denotes the set of neighbors of x. Note that k(z) is the degree k of node z.

Resource Allocation Index

The resource allocation algorithm assumes that nodes can send resources to each other. Their common neighbors play the role of transmitters, and each transmitter has a unit of resource that will distribute equally among all its neighbors. The similarity between nodes x and y can be defined as the amount of resource y receives from x:

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z)}.$$

Both the Adamic-Adar index and resource allocation index penalize the contributions of common neighbors with high degrees. The difference between them will be small if the degree, k(z), is small, and large if k(z) is large. This means that the prediction results of both of these indices will be similar when the average degree is small, while for networks with a large average degree, the resource allocation index will normally perform better (Zhou, Lu, and Zhang, 2009). The Resource Allocation index considers that each node can send a fraction of a resource (any form of social capital, for instance) through its neighbors. The difference between this algorithm and the Adamic-Adar index is that the latter penalizes neighbors that are more densely connected. In the context of criminal networks, forms of social capital-like information, innovative ideas, and trust are key to understanding the network structure (Campana, 2016). For this reason, we argue that both algorithms can be particularly useful in detecting potential new connections in this network.

Preferential Attachment Index

Following the Barabasi-Albert model (Barabasi, 2009), the preferential attachment algorithm is commonly used to generate evolving or static scale-free networks (with power law-degree distribution or without growth), where the probability that a new link is connecting x and y is proportional to k(x) and k(y). The similarity index for this algorithm is defined as:

$$s_{xy} = k(x) \times k(y)$$

Because this index requires less information than all the others, it also has a minimal computational complexity. This index will be especially useful for networks that reflect the rich-club phenomenon, where large-degree nodes will be densely connected to each other, and small-degree nodes will be sparsely connected to each other (Bass et al., 2013).

Jaccard Index

The Jaccard index is the proportion of shared nodes between A and B relative to the total number of nodes connected to both A and B. This index is defined as:

$$s_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}.$$

In the context of criminal networks, the Jaccard Coefficient assumes that criminal organizations are inclined to create new alliances among themselves given the shared allies they already have. It is a metric that can be related to the notion of trust and shared expectations: The more shared friends you have, the more likely you are to trust someone. An advantage of the Jaccard index is that it may be more intuitively understood than the other algorithms because it yields probabilities.

Deep Neural Networks

Deep Neural Networks (DNNs) are sophisticated artificial intelligence models that consist of multiple layers of interconnected nodes, that allow to approximate complex functions and capture intricate patterns in data. They have become extremely popular in multiple applications that go from image recognition to natural language processing. DNNs normally consist of an input layer, several hidden layers (hence the "deep" term), and an output layer. Each layer contains neurons or nodes, where each neuron in one layer is connected to all neurons in the subsequent layer, forming a dense network. The strength of such connections is represented by weights, which can be adjusted during training (Liu and Gamal, 2017). The computation of each neuron involves two steps: in the linear transformation each neuron receives inputs form the previous layer, multiplies these inputs by its weights, and adds a bias term:

$$z_i^{(l)} = \sum_j w_{ij}^{(l)} x_j^{(l-1)} + b_i^{(l)}$$

Where "x" are the inputs form the previous layer, "w" are the weights, and "b" is the bias for neuron i in layer l. The second step involves the activation function that is applied to each neuron's output. Common choices for this function are the sigmoid, ReLU, and tanh functions. The activation function is given by:

$$a_i^{(l)} = \sigma(z_i^{(l)})$$

Finally, through adjusting the weights and biases it is possible to minimize the loss function (the difference between the network's prediction and the actual target values). The loss function uses algorithms like gradient descent, given by the function:

$$\theta_{\text{new}} = \theta_{\text{old}} - \eta \nabla J(\theta)$$

Where θ represents the parameters of the model (such as weights). η (eta) is the learning rate, or the size of the steps towards the minimum of the loss function. Finally, $\nabla J(\theta)$ is the gradient of the loss function J with respect to the parameter θ. Enhancing the four algorithms with DNNs involved using the regular predictions as inputs into the DNNs architecture and adjusting the weights and biases to minimize a loss function. These values were then passed through the hidden layers to get an output with the predicted values. This method essentially leverages the learning capabilities of DNNs to capture more complex patterns and interactions in the data than the similarity indices alone.

The Node2Vec algorithm

The Node2Vec algorithm was introduced by Grover and Leskoec (2016) and extends the foundational principles of the Word2Vec model to graph data. This provides a very powerful method for learning continuous feature representations for nodes in networks. The algorithms operate in two main stages: random walks and stochastic gradient descent optimization of a skip-gram model. The random walk allows the algorithm to explore diverse neighborhoods and capture the equivalence and structural roles of nodes within the graph. Two parameters, p and q, govern the random walk and control the likelihood of revisiting the immediate previous node and exploring outward. After the random walk is initialized (starts in node v), the algorithm selects the next node x based on transition probabilities π_{xv} between nodes and continues until the walk length is reached:

$$\pi_{vx} = \begin{cases} \frac{1}{p} & \text{if } x \text{ is the previous node} \\ 1 & \text{if } x \text{ is a neighbor of } v \text{ but not the previous node} \\ \frac{1}{q} & \text{if } x \text{ is not a neighbor of } v \end{cases}$$

The algorithm treats the sequences of nodes generated by random walks as sentences, with nodes treated as words. Then, the Node2Vec utilizes the skip-gram architecture[14] to maximize the log-probability of observing a network neighborhood (denoted by Ns(u)) for a node u conditioned on its feature representation (Meng, 2020).

$$\max \sum_{u \in V} \log \Pr(N_S(u) | f(u))$$

Where f(u) denotes the feature representation of a node u. The probability of observing a neighborhood node v ∈ Ns(u) given the feature representation of u is modeled with a SoftMax function:

$$\Pr(v|f(u)) = \frac{\exp(f(v) \cdot f(u))}{\sum_{w \in V} \exp(f(w) \cdot f(u))}$$

Finally, the embeddings generated by the Node2Vec algorithm can be used for diverse purposes like measuring node similarities, performing node classification and link prediction tasks. The

---

[14] I simple words, the skip-gram takes a target word and tries to predict the surrounding context words. The window size determines the span of words on either side of the target word to be analyzed (Zhang, Liu, and Bis, 2019)

advantage of this algorithm over the rest analyzed in this chapter is that it captures both local and global graph structures which allows for learning more complex graph topologies (Meng, 2020).

Methodology

The method we used to implement the different algorithms followed a three-layered logic: first we implemented the four similarity-based algorithms (Jaccard Index, Preferential Attachment, Adamic Adar Index, and Resource Allocation). These traditional similarity indices compute scores based on direct and indirect relationships between nodes, primarily using immediate topology, such as shared neighbors; in the second layer, we implemented the Node2Vec algorithm on top of each one of these indices. The vector embeddings produced by Node2Vec are concatenated with scalar scores from traditional indices to form a feature vector for each node pair. This new feature vector contains context-aware embeddings along with direct topological similarity measures (see Figure 1 as an example of Node2Vec). This combination gives the model a richer set of features to learn from, which provides a deep insight into each node's positional roles in the larger network and captures hidden patterns that simple indices might miss. The integration of the Node2Vec and traditional similarity indices allows predictions that are based on observable structural properties and on complex patterns learned through embeddings. The third layer of algorithms we tested incorporates Deep Neural Networks (DNNs) in addition to the Node2Vec implementation. We used Node2Vec to first extract meaningful features from the graph structure and then we fed those features into a DNN to perform link predictions. Additional to these 12 implementations, we implemented a Node2Vec algorithm on its own to assess its predictive capacity, as well as a Node2Vec with a DNN architecture.

In total, we implemented 14 algorithms that were tested for accuracy using the Area Under the Curve (AUC), also known as the AUC of the Receiver Operating Characteristic (ROC) curve, alongside a 5-fold cross validation for robustness. We tested all 14 algorithms both with their capacity to correctly predict the formation of existing ties as well as non-existing ones. The AUC is a widely used metric for evaluating the performance of classification models. It includes two parameters: the True Positive Rate, which measures the proportion of actual positives correctly identified by the model (TP / (TP + FN)); the second parameter is the False Positive Rate, which measures the proportion of actual negatives incorrectly classified as positives by the model (FP/ (FP + TN)). The score goes from 0 to 1, where .5 means that the model has no ability to discriminate between positive and negative cases so it's just as good as random guessing. A score of 1 would represent a prefect model because it correctly classifies all positive and negative cases.

Figure 19 – Node2Vec Implementation
This example depicts the color-coded communities exhibiting homophily discovered by node2vec in Les Misérables Network. (Leskovek, 2016).

To improve the robustness of our analysis, instead of just splitting the data into test and training sets, we used the k-fold cross validation method with 5 “folds”. This method divides the data into five equal or nearly equal parts and then each part is used as a test se once, and as part of the training set four times. This process is repeated five times, with each fold being used as the test set once. We evaluated and compared the accuracy scores of each predictive algorithm by computing their AUC score.

Finally, we wanted to predict both existing ties but also potential ones. Using these indices to predict existing ties is important because it gave us insight into the underlying hierarchical and organizational structure of the network. At the same time, predicting potential ties would give us insight into the possible directions in which a criminal network might expand, and potential future characteristics the network might display. In terms of public policy, predicting potential ties could help anticipate and counteract such evolutions and adaptations. This could provide a strategic advantage for policy makers and enforcement agencies by anticipating future changes in the network. For these reasons we included two independent sets of predictions, one for existing ties and one for potential future ties.

Dataset

The data set we analyzed in this chapter was provided by the consulting firm Lantia Consultores, which specializes in public safety, organized crime, and violence.[15] It includes data on connections

[15] The authors received the data from Lantia Consultores in 2021 via a subscription purchased by Rice University’s Baker Institute Center for the U.S. and Mexico.

between criminal organizations in Mexico in 2021. As described in a previous research paper published by the Baker Institute, the initial relational data set came in two edge lists, one with alliance (undirected) data and another with subgroup (directed) data.[16] As before, we combined the two sets for an overall sense of illicit network relations in Mexico.

While the data included 395 organized crime groups, including 387 from the alliance data and additional from the subgroup data, many of those organized crime groups were isolates with no alliances or subgroup relationships. Previous research using similarity indices to make link predictions eliminated isolates to run the analyses (Zhou, Lu, and Zhang, 2009). We decided to follow the same approach in our paper, as similarity indices can only make predictions about connected nodes.

Finally, we binarized the data. This resulted in a network that consisted of 176 nodes and 227 edges. Nodes represent criminal organizations in the network, and edges represent the positive relationships between them. These positive relationships can be understood either as alliances between criminal organizations, or hierarchical relationships in which some criminal groups follow the orders of other, more powerful organizations (see Figure 1).

In SNA, embedding measures help researchers understand a whole population and how the network's structure affects actors in the network (Morselli, Giguere, and Petit, 2007). Statistics

---

[16] A directed edge means that the connection flows in one or both directions such as one person initiating a phone call to another. An undirected tie is used when a relationship exists but there is no data on the direction of the relationship. For example, an investigator may know two people talked on the phone but not who started the call. Nathan P. Jones et al., "Mexico's 2021 Dark Network Alliance Structure."

that measure how individuals are embedded in larger social structures include density, efficiency, clustering, and transitivity (see Table 1). The present criminal network shows a low-density score (0.015), which implies that there are relatively few actual connections compared to all the potential connections. The low average degree of the network (1.29) tells us that each criminal organization has on average, just over one alliance. Likewise, the network's low average clustering coefficient (0.140) suggests a low likelihood that two organizations allied with the same organization are themselves allied with each other.

These network statistics can be more meaningful when comparing different types of criminal networks. For instance, multiple studies show that individuals in drug trafficking networks tend to have lower path length[17] and clustering coefficients than other criminal organizations (Morselli, Giguere, and Petit, 2008). In terms of density and centrality, studies also show that drug trafficking networks, whose primary purpose is to make money, tend to favor efficiency (higher density), while networks with more ideological goals or a longer time to act, favor sparseness with fewer central actors (Bright and Delaney, 2013). Similarly, Dorn, Levi, and King (2005) concluded that drug traffickers are driven by profit and are more likely to have a durable core in their network with several connections to diverse groups and individuals. This differs from ideologically motivated criminals, who will not show such diversity of connections (Bichler, Malm, and Cooper, 2017).

[17] Path length is the distance between two nodes, measured as the number of edges between them: https://www.futurelearn.com/info/courses/social-media/0/steps/16047.

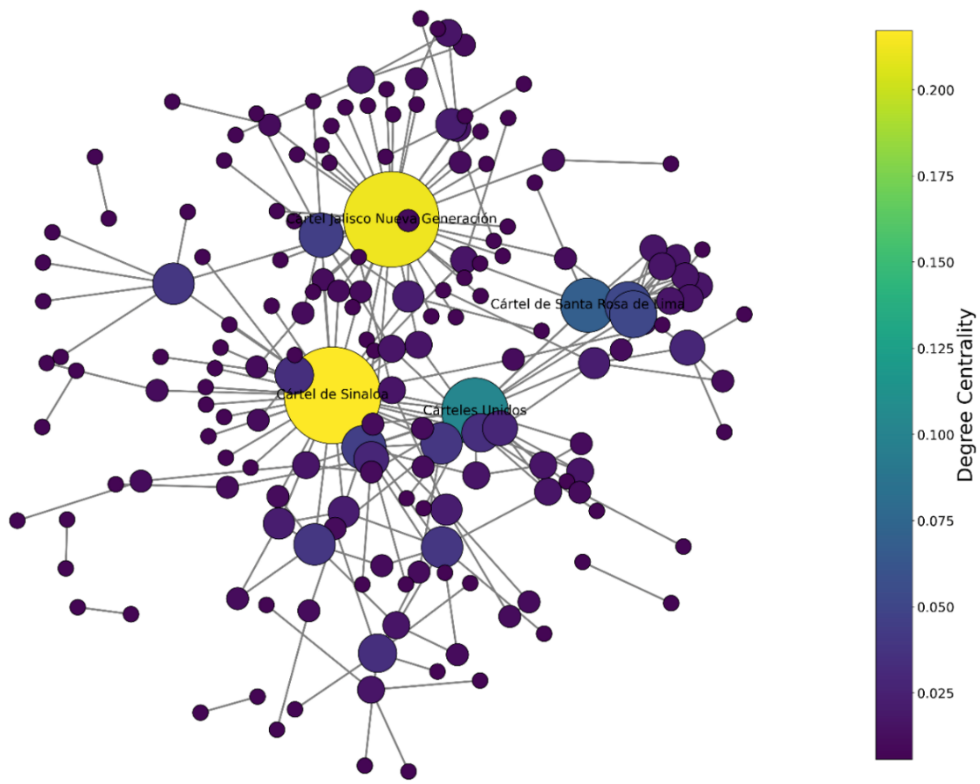


Figure 20 — Mexico's Organized Crime Network (2021).
This network includes 176 nodes and 226 edges. Only the most central nodes (degree centrality) were labeled. Elaboration based on Lantia Consultores 2021 data.

Table 1 — Network Statistics

Table 1 — Network Statistics

Parameter	Nodes	Edges	Density	Avg. Degree	Diameter	Components	Avg. Clustering
Network	176	227	0.015	1.290	10	9	0.140

In addition to describing the overall structure of the network, one of the primary uses of graph theory in SNA is the identification of central actors in the network (Wasserman and Faust, 1994) Here we focus on the structural and locational properties of nodes in the network using three of the most popular centrality measures: degree, betweenness, and closeness (see Table 2). Degree centrality measures the number of connections a node has and shows that prominent actors have the most ties to other actors in the graph. Betweenness centrality measures the number of times a node lies on the shortest path between all pairs of actors. This measure is typically used to identify brokerage positions in a network (Currie et al., 2014). Finally, closeness centrality measures the geodesic distance from each node to all other actors in a network (Wasserman and Faust, 1994). This measure is helpful when trying to identify nodes that spread things (like information) more efficiently through the network (Golbeck, 2013).

In this specific network, the five organizations with the highest degree centrality scores are the Sinaloa cartel (0.217), the Jalisco New Generation cartel (0.211), Cárteles Unidos (0.103), the Santa Rosa de Lima cartel (0.068), and the Nueva Plaza cartel (0.051). These represent the most active organizations in terms of positive ties. The Sinaloa cartel has a history of embedding itself in dense networks and may also be doing this to balance the threat of the CJNG (Jones et al., 2022). The CJNG has a hierarchical alliance structure that is less dense, which may explain its high degree centrality but low closeness centrality.[18] Additionally, there is an important subgroup of

[18] Degree centrality refers to the number of connections for a node. Closeness centrality indicates how close a node is to all other nodes. Betweenness centrality measure how many times a node lies on the shortest path between other nodes. See https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/. Closeness centrality is traditionally a measure that works on fully connected networks. However, the Python library Networkx adapts its closeness centrality algorithm so it can be implemented on disconnected

organizations in the Tierra Caliente region aligned with Cárteles Unidos and the Sinaloa cartel. This may explain the low position of the CJNG in terms of closeness centrality. Interestingly, three of the cartels with high closeness centrality are positioned between the CJNG and Sinaloa, despite their low degree and betweenness centralities.

Table 2 — Centrality Measures

Degree Centrality		***Betweenness Centrality***		***Closeness Centrality***	
Organization	**Score**	**Organization**	**Score**	**Organization**	**Score**
Sinaloa	0.217	Sinaloa	0.568	Sinaloa	0.374
CJNG	0.211	CJNG	0.380	Los Negros	0.330
Cárteles Unidos	0.103	Cárteles Unidos	0.187	Cártel del Poniente	0.324
Cártel de Santa Rosa de Lima	0.068	Los Negros	0.102	Esquema Gan	0.321
Nueva Plaza	0.051	La Nueva Familia Michoacana	0.089	Los Erres	0.321

Finally, the degree rank plot and the degree histogram (see Figure 2) allow us to observe a degree distribution like that of power law distributions. This means that the network is highly centralized, with a few nodes monopolizing most of the connections. This characteristic is consistent with research showing that drug trafficking networks have higher centralization and density than other criminal networks, specifically terrorist networks (Xu and Chen, 2012). Other studies also show that this centralization increases with the threat of law enforcement targeting (Morselli, Giguere, and Petit, 2008).

---

networks. For more information about the implementation of this algorithm you can review the Neworkx documentation: https://bit.ly/44tx1c3.

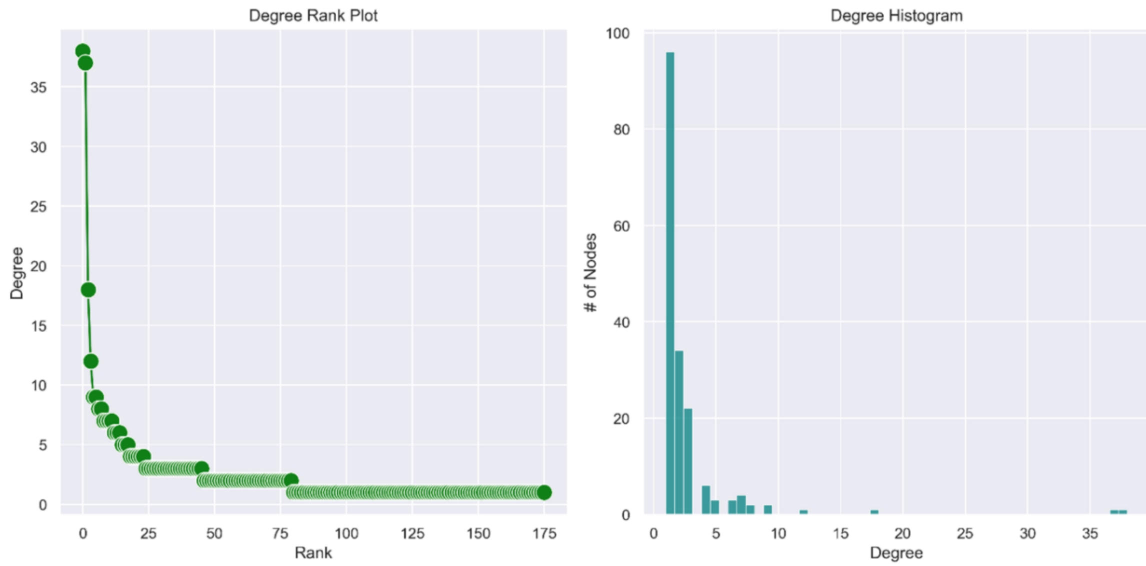


Figure 21 — Highly Centralized Criminal Network

Results

Figure 4 shows the results of the comparisons of AUC scores between the four similarity indices and their enhancements. The focus here is on the predictive capabilities of algorithms on the formation of existing ties. The best performing algorithm is Preferential Attachment enhanced by the Node2Vec and the DNN architectures (AUC = .908), the second-best performing algorithm is Preferential Attachment enhanced only by Node2Vec (AUC = .841), and the third best performing algorithm is the Node2Vec enhanced by the DNN (AUC =.813). Interestingly, even the Preferential Attachment algorithm without any enhancements showed a relatively high accuracy (AUC = .79) The fact that the Preferential Attachment similarity index tends to perform better than the rest is consistent with the power-law like distribution of the network, where a few nodes exhibit many connections, and many nodes show very few connections. In other words, the fact that the

Preferential Attachment is better at understanding and predicting links in the network is related to the very structure of the network.

Another interesting result is that both enhancements (Node2Vec + DNN) were better than one (just Node2Vec), and that one enhancement was better than just implementing the traditional similarity index. Also, the Node2Vec algorithm did a better job predicting existing ties than the rest of the similarity indices, except for Preferential Attachment.

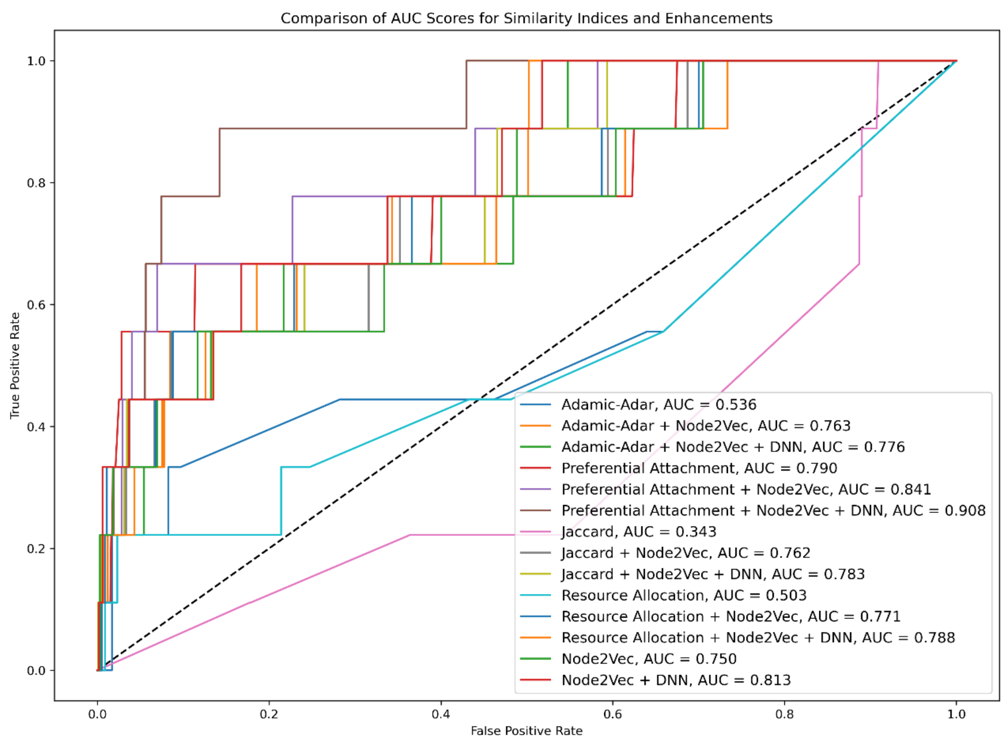


Figure 22. AUC Scores for Similarity Indices and Enhancements for the Lantia Dataset

Figure 5 describes the predictive capabilities of algorithms on the formation of potential ties. The focus here is on the predictive capabilities of algorithms on the formation of ties that are not present in the network but that might happen. In this case, the best performing algorithm was the Jaccard Coefficient enhanced both by the Node2Vec and the DNN (AUC = .946), the second best was the Node2Vec algorithm (AUC = .934), and the third best was the Node2Vec enhanced with the DNN (AUC = .933). Interestingly, in all cases, the algorithms enhanced both by the Node2Vec and the DNN performed worse than the ones enhanced just by the Node2Vec. This could be happening because the use of the DNN might increase the likelihood of overfitting the data, which reduces the AUC scores. An unexpected finding is related to the best performing algorithm. We already established that the Preferential Attachment similarity index best captures the structure of the network. However, when predicting new potential links, the Jaccard Coefficient enhanced by the Node2Vec performed better. The procedure of this enhanced algorithm is as follows: first the algorithm computes the Jaccard Coefficient for every pair of nodes that aren't already connected by an edge. The coefficient helps identify pairs of nodes that have a high likelihood of connection based on their shared networks. Next, the algorithm focuses on those filtered nodes and computes the cosine similarity between their Node2Vec embeddings. Pairs with high cosine similarity scores will be likely to form a connection. An explanation of why this enhanced algorithm can yield such a high prediction accuracy score is that the Jaccard Coefficient captures the immediate neighbor overlap but does not account for longer paths or the global network structure, while the Node2Vec does capture broader topological and community-based features that might not be apparent from direct connections alone. In other words, the combination of the Jaccard Coefficient and the Node2Vec seems to be very powerful because it seems to consider both local and broader topological features, which leads to a more accurate prediction model. This layered approach can be more effective than using each method separately because it first narrows the potential node pairs that share neighbors and then applies the Node2Vec embeddings to further refine the prediction by considering the broader network structure.

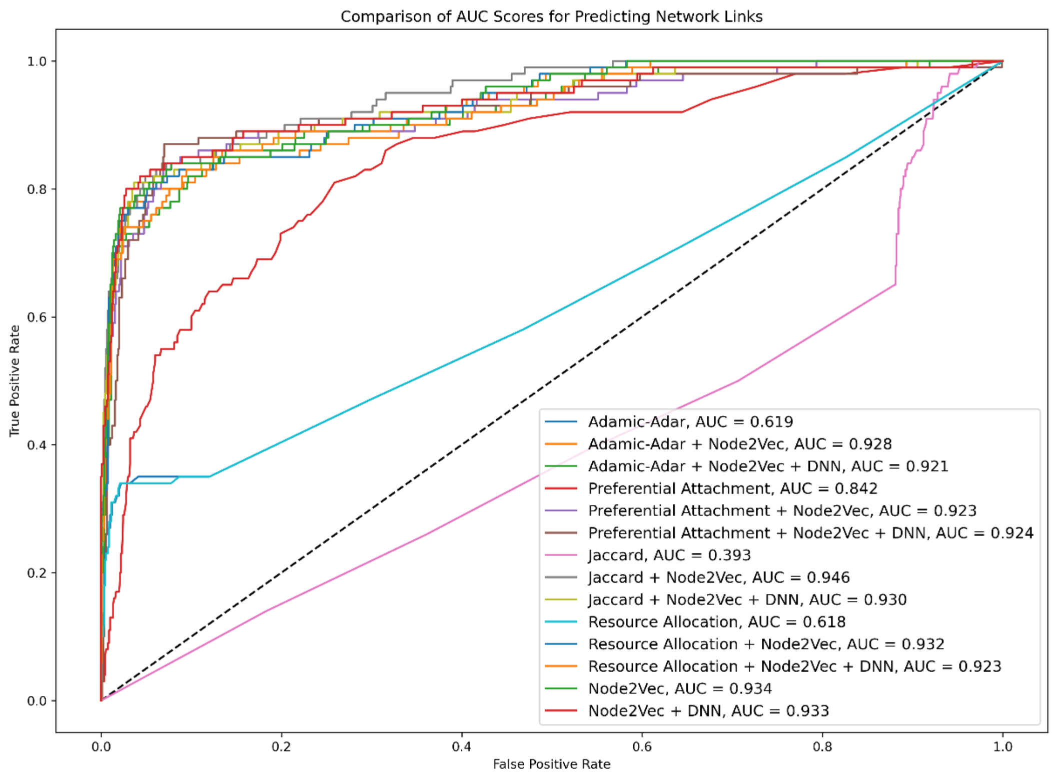


Figure 23 — Comparison of AUC Scores for Predicting Links for the Lantia Dataset

Next, we tested the same algorithms on the BACRIM dataset, which is an open-source dataset developed by the Center for Research and Teaching in Economics (CIDE). This dataset describes Mexico's network of violent relationships between criminal organizations in 2021, and it has a total of 67 nodes and 87 edges[19]. We followed the same methodology as with the previous dataset and tested the capacity of these algorithms to predict both existing and future links. Figure 6 describes the predictive capabilities of these algorithms on the formation of existing ties. The best

[19] For more information on this dataset visit https://ppdata.politicadedrogas.org/.

performing traditional similarity index was Preferential Attachment (AUC = .819). However, when enhancing the index with the Node2Vec and the DNN architectures, the Preferential Attachment increased its accuracy score (AUC = .857). The worst performance algorithm was The Jaccard Coefficient (AUC = .414), but when enhanced by both the Node2Vec and DNN architectures, the algorithm increased its accuracy score dramatically (AUC = .878). This result is consistent with the previous dataset where the Preferential Attachment algorithm was better at understanding and predicting existing links in the network given the inherent structure of the network.

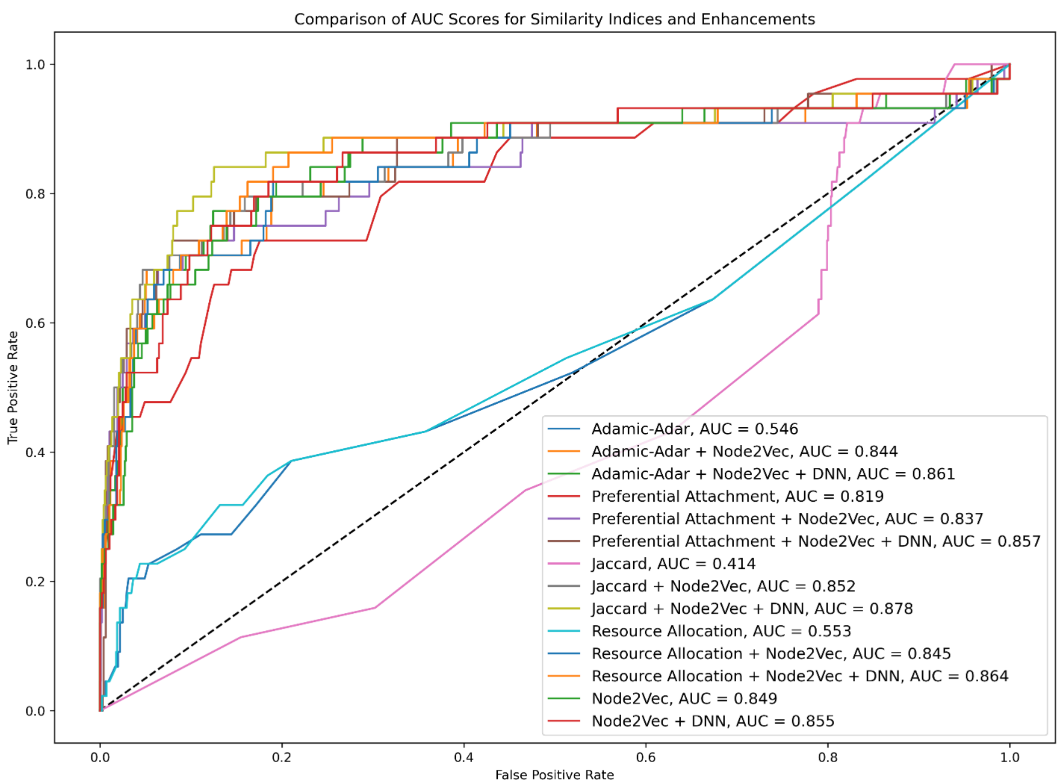


Figure 24. Scores for Similarity Indices and Enhancements for the Bacrim Dataset

Finally, Figure 7 shows that the best performing algorithm to predict future links in the BACRIM dataset was the Jaccard Coefficient enhanced by the Node2Vec and the DNN (AUC = .849). This is also consistent with the results of the previous dataset. The Jaccard Coefficient enhanced by the Node2Vec and DNN is more powerful than other algorithms because it considers both local and broader topological features of the network, which leads to a more accurate prediction model. The results of the BACRIM dataset confirm the power of combining an algorithm that considers local and structural topological features.

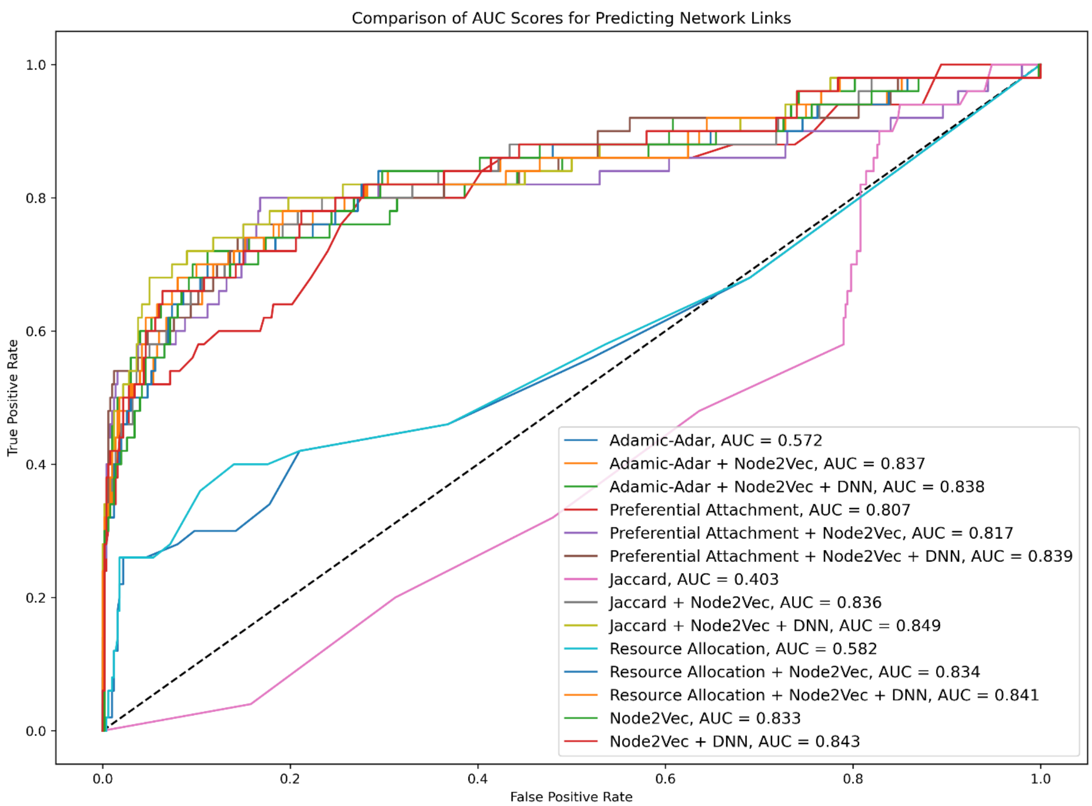


Figure 25. Comparison of AUC Scores for Predicting Links for the Bacrim Dataset

Chapter Discussion and Conclusions

This chapter uses a combination of social network analysis (SNA) and machine learning techniques to predict links in Mexico's network of criminal organizations. Specifically, we use four similarity- based algorithms, as well as their enhanced versions, to estimate the likelihood that a link will be formed between two unconnected organizations in the network. Of the traditional node-similarity indices implemented, we found that in both datasets tested (Lantia and BACRIM) Preferential Attachment performed better when predicting existing ties. Overall, the enhanced algorithms with both the Node2Vec and the DNN architectures performed better than their traditional versions. Also, the Node2Vec algorithm proved very powerful and outperformed all traditional similarity indices. We also concluded that the high accuracy score of the Preferential Attachment algorithms is caused by the power law-like distribution described in both criminal networks. We hypothesize that such a structure could be explained by the existence of hierarchical relationships where a few organizations monopolize activities, control, and subcontract other smaller organizations, but other tests would be required to confirm such hypothesis.

When predicting potential future ties, the best performing algorithm in both networks was the Jaccard Coefficient enhanced both by the Node2Vec and the DNN. We argue that this happens because the Jaccard Coefficient captures the immediate neighbor overlap, while the Node2Vec captures broader topological and community-based features that might not be apparent from direct connections alone, so the use of this enhanced algorithm can be very powerful to predict future network ties in any network, not only criminal ones. This layered approach can be more effective than using each method separately because it first narrows the potential node pairs that share neighbors and then applies the Node2Vec embeddings to further refine the prediction by considering the broader network structure.

In the case of Mexico's network of criminal organizations, Mexico finds itself in a difficult situation wherein its security forces have proven capable of kingpin strikes that can disaggregate large criminal hierarchies, but its local and state institutions lack the ability to address the many smaller predatory networks that have proven highly resilient in the overall criminal structure (Ernst, 2019). In this context, network analysis of the macro-organized crime structure is increasingly important. Because network analysis and machine-learning algorithms can forecast illicit network alliances, they are a valuable tool for public and private sector security services.

Our analysis demonstrates that, when applied to dark networks, predictive algorithms such as the Adar-Adamic, Preferential Attachment, Jaccard Coefficient, and Resource Allocation indices, enhanced by Node2Vec and DNN architectures, may have impressive properties from an intelligence perspective. These algorithms may not just predict future alliances or edges within illicit networks: They may also point intelligence analysts toward missing data that exists but is not available to them due to the clandestine nature of organized crime networks. They could also improve the quality of data analytics by intelligence analysts tasked with fighting these illicit groups. As organized crime networks become increasingly enmeshed as a mechanism to expand their resilience, this type of analysis will play a critical role in combating them.

We thus recommend that policymakers and intelligence agencies incorporate these machine-learning techniques into their analysis of illicit networks. A key contribution from this analysis is the ability of predicted alliances to point us toward what is likely missing data on dark networks. This information could next be used in non-kinetic intelligence gathering strategies to confirm likely relationships, or as part of other kinetic options.[20] This important policy recommendation is based on a simple and intuitive finding and means that these algorithms can help us to design better strategies to address complex illicit network structures. This also serves as a reminder to analysts to be aware of this acute missing data problem and to remember that dark networks are almost inevitably denser than our data suggests.

[20] "The kinetic approach involves aggressive and offensive measures to eliminate or capture network members and their supporters, while the non-kinetic approach involves the use of subtle, non-coercive means for combating dark networks:" (Roberts, and Everton, 2014).

References

Atuesta, L. H., & Pérez-Dávila, Y. S. (2018). Fragmentation and Cooperation: The Evolution of Organized Crime in Mexico. *Trends in Organized Crime, 21*(3), 235–261. https://doi.org/10.1007/s12117-017-9301-z

Barabási, A.-L. (2009). Scale-Free Networks: A Decade and Beyond. *Science, 325*. https://barabasi.com/f/303.pdf

Berman, E. (2018). A Government of Laws and Not of Machines. *Boston University Law Review, 98*(5), 1277–1356. https://dx.doi.org/10.2139/ssrn.3098995

Bichler, G., Malm, A., & Cooper, T. (2017). Drug Supply Networks: A Systematic Review of the Organizational Structure of Illicit Drug Trade. *Crime Science, 6*(2). https://doi.org/10.1186/s40163-017-0063-3

Bright, D. A., & Delaney, J. J. (2013). Evolution of a Drug Trafficking Network: Mapping Changes in Network Structure and Function across Time. *Global Crime, 14*(2–3), 238–260. https://doi.org/10.1080/17440572.2013.787927

Burt, R. S. (2000). The Network Structure of Social Capital. *Research in Organizational Behavior, 22*, 345–423. https://doi.org/10.1016/S0191-3085(00)22009-1

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *The American Journal of Sociology, 94*, S95–S120. http://www.jstor.org/stable/2780243

Currie, G., Burgess, N., White, L., Lockett, A., Gladman, J., & Waring, J. (2014). A Qualitative Study of the Knowledge-Brokering Role of Middle-Level Managers in Service Innovation: Managing the Translation Gap in Patient Safety for Older Persons' Care. *Health Services and Delivery Research, 2*(32). https://doi.org/10.3310/hsdr02320

Deeks, A., Lubell, N., & Murray, D. (2019). Machine Learning, Artificial Intelligence, and the Use of Force by States. *Journal of National Security Law and Policy, 10*(1), 1–26. https://ssrn.com/abstract=3285879

Dorn, N., Levi, M., & King, L. (2005). Literature Review on Upper Level Drug Trafficking. London: Home Office. https://www.ojp.gov/ncjrs/virtual-library/abstracts/literature-review-upper-level-drug-trafficking

Duijn, P. A. C., & Klerks, P. P. H. M. (2014). Social Network Analysis Applied to Criminal Networks: Recent Developments in Dutch Law Enforcement. *Networks and Network Analysis for Defence and Security*, 121–159. https://doi.org/10.1007/978-3-319-04147-6_6

Ernst, F. (2019). Mexico's Hydra-headed Crime War. *International Crisis Group*. https://www.crisisgroup.org/latin-america-caribbean/mexico/mexicos-hydra-headed-crime-war

Everton, S. F. (2012). Disrupting Dark Networks. New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139136877

Faust, K., & Wasserman, S. (1994). Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815478

Fuxman Bass, J. I., Diallo, A., Nelson, J., Soto, J. M., Myers, C. L., & Walhout, A. J. M. (2013). Using Networks to Measure Similarity between Genes: Association Index Selection. *Nature Methods, 10*, 1169–1176. https://doi.org/10.1038/nmeth.2728

Gamal, A. E., Liu, X., & Yang, D. (2017). Deep Neural Network Architectures for Modulation Classification. *51st Asilomar Conference on Signals, Systems, and Computers*, 915–919. https://doi.org/10.1109/ACSSC.2017.8335483

Grover, J. Leskovec. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.

Golbeck, J. (2013). Analyzing the Social Web. Waltham, MA: Elsevier.

Herrera, J. S. (2022). The Limits of Resistance to Criminal Governance: Cyclical Violence and the Aftermath of the Autodefensa Movement in Michoacán, Mexico. *Global Crime*, 1–25. https://doi.org/10.1080/17440572.2021.2024805

Horowitz, M. C., Allen, G. C., Saravalle, E., Cho, A., Frederick, K., & Scharre, P. (2018). Artificial Intelligence and International Security. Center for a New American Security. https://csdsafrica.org/wp-content/uploads/2020/06/CNAS_AI-and-International-Security.pdf

Jones, N. P., Dittmann, W. L., Wu, J., & Reese, T. (2018). A Mixed-methods Social Network Analysis of a Cross-border Drug Network: The Fernando Sanchez Organization. *Trends in Organized Crime, 23*, 154–182. https://doi.org/10.1007/s12117-018-9352-9

Lim, M., Abdullah, A., & Jhanjhi, N. Z. (2021). Performance Optimization of Criminal Network Hidden Link Prediction Model with Deep Reinforced Learning. *Journal of King Saud University, 33*(10), 1202–1210. https://doi.org/10.1016/j.jksuci.2019.07.010

Martínez Torres, J., Iglesias Comesaña, C., & García-Nieto, P. J. (2019). Review: Machine Learning Techniques Applied to Cybersecurity. *International Journal of Machine Learning and Cybernetics, 10*, 2823–2836. https://doi.org/10.1007/s13042-018-00906-1

Mohammed, M., Khan, M. B., & Bashier, E. B. M. B. (2016). Machine Learning. Boca Raton: CRC Press. https://doi.org/10.1201/9781315371658

Morselli, C. (2009). Inside Criminal Networks. New York: Springer.

Morselli, C. (2014). Crime and Networks. New York: Routledge.

Morselli, C., Giguère, C., & Petit, K. (2007). The Efficiency/Security Trade-off in Criminal Networks. *Social Networks, 29*(1), 143–153. https://doi.org/10.1016/j.socnet.2006.05.001

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives, 31*(2), 87–106. https://pubs.aeaweb.org/doi/pdf/10.1257/jep.31.2.87

Rios Contreras, V. (2012). How Government Structure Encourages Criminal Violence: The Causes of Mexico's Drug War. (Ph.D. thesis, Harvard University). http://search.proquest.com/docview/1417075396?accountid=7064

Roberts, N., & Everton, S. F. (2014). Strategies for Combating Dark Networks. *Journal of Social Structure, 12*. http://www.cmu.edu/joss/content/articles/volume12/RobertsEverton.pdf

Sparrow, M. K. (1991). The Application of Network Analysis to Criminal Intelligence: An Assessment of the Prospects. *Social Networks, 13*(3), 251–274.

Sullivan, J. P., & Elkus, A. (2008). State of Siege: Mexico's Criminal Insurgency. *Small Wars Journal*. https://smallwarsjournal.com/jrnl/art/state-siege-mexicos-criminal-insurgency

Sullivan, J. P. (2012). From Drug Wars to Criminal Insurgency: Mexican Cartels, Criminal Enclaves and Criminal Insurgency in Mexico and Central America. Implications for Global Security. Working Paper No. 9. Paris: Fondation Maison des sciences de l'homme. https://shs.hal.science/halshs-00694083/document

Sullivan, J. P. (2019). Criminal Enclaves: When Gangs, Cartels or Kingpins Try to Take Control. Stratfor Threat Lens. https://bit.ly/3YLHSgo

Sullivan, J. P. (2019). The Challenges of Territorial Gangs: Civil Strife, Criminal Insurgencies and Crime Wars. *Revista do Ministério Público Militar, 31*(44). https://bit.ly/3P9lQRl

Sullivan, J. P. (2023). The Information Age: Transnational Organized Crime, Networks, and Illicit Markets. *Journal of Strategic Security, 16*(1), 51–71. https://digitalcommons.usf.edu/jss/vol16/iss1/4

Tilly, C. (1985). War Making and State Making as Organized Crime. In Evans, P., Rueschemeywer, D., & Skocpol, T. (Eds.), Bringing the State Back In. Cambridge: Cambridge University Press, 169–191. https://doi.org/10.1017/CBO9780511628283

Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing Different Supervised Machine Learning Algorithms for Disease Prediction. *BMC Medical Informatics and Decision Making, 19*(281). https://doi.org/10.1186/s12911-019-1004-8

van der Hulst, R. C. (2009). Introduction to Social Network Analysis (SNA) as an Investigative Tool. *Trends in Organized Crime, 12*, 103. https://doi.org/10.1007/s12117-008-9057-6

Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815478

Zhang, C., X. Liu and D. Biś, "An Analysis on the Learning Rules of the Skip-Gram Model," *2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8852182.

Zhou, T., Lü, L., & Zhang, Y.-C. (2009). Predicting Missing Links via Local Information. *The European Physical Journal B, 71*, 623–630. https://doi.org/10.1140/epjb/e2009-00335-8

Conclusions

Throughout this dissertation, I built on computational methods to understand the evolution of Mexico's organized crime, the violence related to it, and some of the consequences derived from such violence. Each chapter is structured as a single academic paper, but they are related in that they all use computational methods to study different aspects of organized crime and criminal organizations in Mexico, and they all follow a core argument: Mexico's war on drugs, initiated in the early 2000s, produced a process of fragmentation of criminal organizations that led to an increase of violence all over the country and that had pernicious consequences for the Mexican population. More specifically, I focus on the consequences of this fragmentation and violence on the hazards migrants face when crossing the U.S. – Mexico border. Although my focus is on Mexico, the debates raised have broader scope and are consequential to other countries and regions. For instance, in the first chapter I analyze how the network of violence amongst criminal organizations evolves and adapts given external shocks in the form of militarization and targeted attacks by the government. It provides a framework to understand some of the consequences of equivocated state policies in the face of organized crime and violence, which can be applied to many other case studies. The second chapter provides a frame to understand the state-like nature of organized crime given its capacity to control a territory, tax its population, and create protection rackets in a context of increased violence. Finally, the third chapter is more methodological and proposes the use of algorithms enhanced with artificial intelligence architectures to better predict the structure of networks and the future formation of links. The main argument is that these enhanced architectures are useful to study any type of network, not just criminal ones.

All three chapters point to important implications for policy makers: the first chapter shows that the kingpin strategy was a failure, and needs to be reconsidered, but the "how" might need a complex answer. During the 80s and 90s the federal government was highly centralized and was able to maintain a stable arrangement with organized. But given the current situation, going back to a structure where a few organizations controlled most of the territory and had balanced arrangements between them and the federal government is unlikely. Criminal organizations have established arrangements with local governments and might be too embedded in the local dynamics of power, which makes uprooting them increasingly difficult. Furthermore, criminal groups are multiple and heterogeneous, and it is unrealistic to think that they can all be tackled the same way. For instance, my study shows that some organizations are more violent than others, which can be observed by the geographical expansion of their conflicts and their central position in the network. That is the case of the Sinaloa cartel, the Zetas, and the Cartel Jalisco Nueva Generacion (CJNG). A practical, yet controversial, public policy approach could be to focus all military action mainly on the most violent organizations that represent an imminent threat to the safety and well-being of the population and the stability of the state. This would have to be accompanied by strict laws and actions that tackle the financial structure of criminal organizations, as well as all forms of corruption. Finally, the judicial system would have to be reformed to regain the trust of the population.

The second chapter of my dissertation also posits interesting insights and challenges for policy makers. If organized crime is structured and acts like a quasi-state, then how can the state approach it? If a frontal attack has not worked (as shown by my first chapter) then what are the options? Again, a multilayered approach would be needed. First the government would need to retake territorial and political control at the local level to better protect migrants and migration.

Furthermore, the government would need a more aggressive strategy to provide funding and protection for shelters and NGOs that work with migrants, particularly along the U.S-Mexico border. Furthermore, the fact that some migrants are even more vulnerable than other is not a new insight, but this study shows that such vulnerabilities are indeed subject to greater hazards when crossing to the United States, and the Mexican government needs acknowledge this. Finally, a judicial overhaul is much needed. Without a proper rule of law and legal protection for migrants, and the civil society in general, criminal organizations are essentially immune, which fosters their capacity to act like quasi-states and establish their own rule.

The last substantive chapter of my dissertation (Chapter 3) is intended mainly as a methodological contribution. It is an answer to the dearth of good quality data that is publicly available on illicit networks. Given these constraints, the proposed algorithms help better understand the structure of these networks and predict future alliances. In terms of policy making, one of the core arguments here is that as organized crime networks become increasingly enmeshed as a mechanism to expand their resilience, this type of analysis will play a critical role in combating them. These algorithms can help us to design better strategies to address complex illicit network structures, but these exercises also serve as a reminder of the acute missing data problem and to remember that illicit networks are likely denser than our data suggests.

In terms of some of the overall limitations of my research, despite promising results using machine learning and algorithms powered by artificial intelligence, the lack of good quality data that is publicly available is still an issue and can potentially hinder the scope of any research related to organized crime and criminal organizations. Second, the use of “black box” models limits the transparency and accountability of some of these methods and algorithms, which can be detrimental in terms of methodological robustness and reproducibility. Furthermore, although the high prediction scores of the proposed algorithms seem promising, more analysis is needed with additional criminal and non-criminal networks to confirm their effectiveness.

Finally, although my statistical models and machine learning techniques in Chapter 2 show robustness, more research would be necessary to strengthen the hypothesis about criminal organizations behaving like a state. Many of the smaller criminal groups that are part of the organized crime network analyzed throughout these chapters seem to be structured by relatively horizontal relationships that are more invested on quick economic gains than on controlling territories or creating protection rackets. A topic for future inquiries is understanding how some organizations can be more territorial than others and to what extent they behave more like modern firms than states or both.