Los Angeles Criminal Network and Analysis

Yuke Wu
Department of Electrical and Computer Engineering
Mountain View, CA
yukew@andrew.cmu.edu

Wenxi Yang
Department of Electrical and Computer Engineering
Mountain View, CA
wenxiyan@andrew.cmu.edu

Dongchen Cui
Department of Electrical and Computer Engineering
Mountain View, CA
dongchec@andrew.cmu.edu

***Abstract*—The research project focuses on analyzing the crime situation in Los Angeles, using network analysis. Los Angeles, as one of the largest and most diverse cities globally, presents a significant area of study with implications for the city's well-being. The dataset has 1,584,316 rows in total and the date of crimes spans from 2010-01-01 to 2017-09-09. The types of crimes in the dataset vary quite a lot, including "VEHICLE - STOLEN", "ATTEMPTED ROBBERY", "BURGLARY FROM VEHICLE", "BURGLARY, ATTEMPTED", "RAPE, ATTEMPTED", etc. The project aims to provide comprehensive insights to benefit government agencies and researchers interested in understanding and addressing crime issues. In the initial phase, an overview of the overall crime distribution in Los Angeles is investigated. Subsequently, the research examines various perspectives and characteristics of the crime data to derive meaningful insights.**

***Index Terms*—Criminology, Social Network, Temporal Transitions, Victim Analysis, Crime Analysis**

I. Introduction

Los Angeles, a sprawling metropolis with over 10 million inhabitants, presents an intricate tableau of crime, reflecting its complex socio-cultural mosaic. Traditional crime analysis, often siloed by precinct or neighborhood, fails to capture the nuanced interplay between Los Angeles's diverse communities and their unique criminal landscapes. This study seeks to transcend these boundaries using network analysis, a method yet to be fully exploited in this domain. Here, crime patterns refer to the frequency, location, and type of crimes committed, while sociological implications encompass the resultant social effects, ranging from economic shifts to communal well-being. By mapping these patterns onto the city's social fabric, the goal of this project is to identify not just the epicenters of crime but also the underlying socio-economic currents that shape them. Our objectives include pinpointing hubs of heightened criminal activity, examining vulnerabilities across varied demographics, and tracing the evolution of these patterns over time. The expected outcomes of this research extend beyond the precincts of policy-making and law enforcement; they promise to offer insights for urban planners, community activists, and public health officials, ultimately contributing to a safer, more cohesive Los Angeles.

II. Prior Work

The body of prior work in the domain of crime network analysis offers a valuable foundation for this project. Burcher sheds light on the common data challenges encountered in this field, such as the inevitable data incompleteness that our study must navigate. The instrumental approach advocated in the criminology literature suggests analyzing networks in terms of nodes and ties, a perspective that is integral to our analysis of Los Angeles crime networks.

Advanced Social Network Analysis (SNA) models, like the one introduced by Schwartz and Rouselle, which account for the strength of both actors and relationships, provide a nuanced understanding of network dynamics that is crucial for distinguishing different types of criminal activities in our research. Liu et al. underscore the significance of identifying key players in criminal networks, a concept that aligns with our objective to pinpoint influential individuals within LA's crime landscape.

Data mining techniques are another area that holds potential for enhancing crime analysis. The use of Self-Organizing Maps (SOM) and Artificial Neural Networks (ANN), as explored by Keyvanpour et al., presents an innovative approach to pattern identification, which could be adapted to our study to refine the analysis of crime relationships in Los Angeles. Rostami and Mondani offer a practical example of analyzing and comparing different types of criminal networks, providing a methodological template that can inform our comparative analyses.

Furthermore, practical applications of network analysis techniques, such as those discussed by Wheelar, provide concrete examples of how to utilize data within analytical software environments like Python and NetworkX, which can be leveraged to enhance our crime detection and matching processes.

In sum, the preceding studies contribute a range of analytical strategies and highlight potential pitfalls that will inform the methodological approach of our project. By integrating these insights, the goal of this project is to build upon the existing body of knowledge and contribute to the development of effective strategies for crime analysis in urban settings.

III. Introduction and Motivation

The study embarks on a multifaceted exploration of criminal activities in Los Angeles, employing network analysis to dissect these activities from three distinct perspectives.

This research is driven by three primary objectives: tracing the temporal patterns of crime, examining the demographic and spatial distribution of criminal activities, and scrutinizing the relationship between crime types and weapons used. These objectives are approached through the construction and analysis of three separate networks, each offering unique insights. The temporal network reveals the evolution of crime over time, the sex, age & area network sheds light on demographic and geographical trends, and the crime type & weapon used network illuminates the interplay between different crimes and their associated weapons. Network analysis techniques are employed to uncover underlying patterns and correlations that can inform strategic crime prevention and policy making.

The motivation for this multifaceted study stems from a growing need to understand the complex nature of criminal activities in urban environments. By dissecting crime data through these three innovative network lenses, our project aims to provide a holistic view of crime in Los Angeles, offering insights that transcend traditional, one-dimensional analysis. The findings from our temporal, demographic, and crime-weapon networks have profound implications for law enforcement strategies, policy formulation, and community safety initiatives. They may not only contribute to academic research in criminology but also serve as a possible practical guide for public authorities and organizations striving to create safer urban communities.

IV. Approach

The first step to start this project is data cleaning. The original dataset is imported and red into a Pandas data frame. Redundant or irrelevant columns were identified and dropped, and columns were renamed for clarity and consistency. Date and time columns were formatted, and additional temporal features (weekday, season) were extracted. The longitude and latitude information was extracted and missing values were handled for various columns. Mapping dictionaries were applied to standardize values in the “victim_sex” and “victim_descent” columns. Text formatting was applied to the “crime_code_description” column. An “ID” column was added for record identification, and the cleaned dataset was saved to a new CSV file. This meticulous data-cleaning process prepared the dataset for in-depth analysis and modeling.

After data cleaning, all temporal information (date, hour, month, weekday, season that crime occurred), crime details (crime description and weapon used, age, sex, and descent of the victim), and geographical information (address, longitude, latitude) are preserved and ready to be analyzed. The three networks derived from this network will be discussed independently.

A. Temporal Network

Dataset description: This network uses all crime incidents as specified in column “crime_code_description” with a “date_occured” column formatted as yyyy–mm–dd. This part of the data is pivotal in constructing a temporal network to analyze crime patterns over time.

Network Building: Each node represents a unique crime type, defined by the specific offense reported. These nodes are interconnected by edges that represent the chronological progression of crimes, where an edge from one node to another signifies that the crime corresponding to the originating node was followed by the crime of the subsequent node on the next date recorded in the dataset.

Assumptions & Task: The primary assumption guiding this approach is the belief that the sequence in which crimes occur can yield insight into potential causal or correlative relationships between different types of crimes. The main task is to identify and quantify these sequential crime patterns, thereby enabling a better understanding of how certain crimes lead to others within the fabric of urban life.

Analysis Plan: The analysis focuses on using a network to identify key crimes that act as hubs in the crime sequence, and community detection algorithms to recognize clusters of crimes that frequently occur in succession. The goal of this network is to uncover not only the key crimes in terms of their position within the daily crime sequence but also to detect distinct groupings of crimes that might suggest underlying structures in criminal behavior over time.

Main Research Question: What are the most common sequential crime patterns in Los Angeles, and what do these patterns suggest about the nature of crime progression in Los Angeles?

B. Network: Sex, Age, and Area

Dataset description: This network uses all crime incidents but drops data for *sex:Heterosexuality*, *sex:unknown*, *sex:Indeterminate*, and $age > 100$. There is quite a portion of recorded crimes have “unknown” for sex, “999.0” for age, leading to some extraordinarily strong edges and rest similar thin edges. Hence, such less useful data points are eliminated when building this network.

Network Building: This network is built with victim age, sex, and area name of which the crime is happened. Each node represents either age, sex, or area name. An edge is created when two nodes share one crime. While building the network with more refined filter rules, normalization is not applied because the number of data points of *sex: Female* is 660,448, *sex: Male* is 663,084, which are close to each other.

Assumptions & Task: One assumption is that within the categories of age and sex, victims are homogeneous enough that analyzing them as single groups (e.g., all 25-year-olds, all females) is meaningful. It also assumes that the data accurately represents crimes in LA, and there is no significant bias in reporting or recording that would skew the network analysis.

Analysis Plan: The analysis aims to utilize the constructed network of crime incidents to discern patterns of crime distribution across various Los Angeles areas, focusing on the demographic attributes of victims—specifically, their age and sex. The objective is to identify areas that may act as nodes with a higher incidence of crimes against particular demographic groups, suggesting potential hotspots for targeted criminal activity.

Main Research Question: How do the distributions of crime incidents vary across different areas of Los Angeles when considering the age and sex of the victims, and what might this indicate about the vulnerability of certain demographic groups within those areas?

C. Network: Crime type and weapon used

Dataset description: This network utilizes data from the column "crime_code_description" and "weapon_description". Columns lacking a description or having an unknown weapon are discarded.

Network Building: The network, representing crime types and weapons used, is constructed as a bipartite network with crime types and weapons as nodes. An edge is introduced between a crime type node and a weapon node when they are associated with the same crime activity. Even if an edge already exists between two nodes, a new edge is added for each additional shared crime activity. To simplify analysis, a projection method is employed, resulting in a weapon-weapon projection. This process involves iterating over each pair of weapon nodes in the previously mentioned multi-graph network of crime types and weapons. For each pair, the number of shared neighboring crime nodes is calculated, indicating the count of crimes involving both weapons. An edge is added between each weapon pair in the projected graph, with the weight of the edge corresponding to the number of these shared crimes.

Assumptions & Task: It is assumed that the type of weapon recorded in each crime case is accurate and definable. The primary task is to identify and quantify potential patterns in weapon usage.

Analysis Plan: The analysis aims to use the network to identify key weapons that exhibit high weighted degree. As well as finding those key weapons that with high clustering coefficient.

Main Research Question: What are the key weapons in the network of criminal activities? When analyzing the co-occurrence of weapons in criminal incidents, what clusters or communities of weapons are observed, and what insights does this provide into criminal weapon preferences and availability?

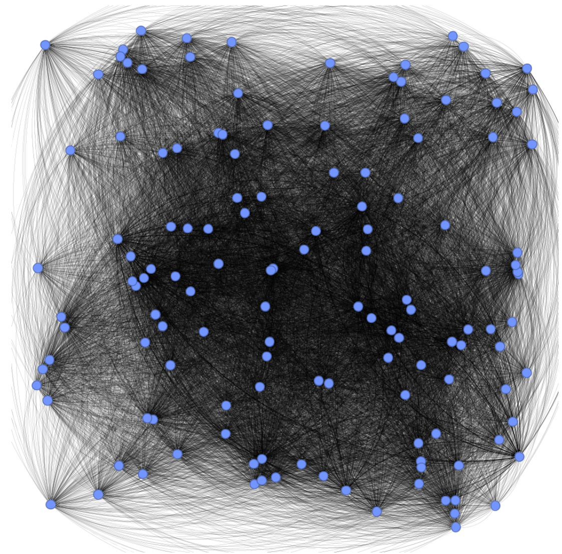

Fig. 1. Temporal Network of crime type and how one crime may lead to subsequent crimes

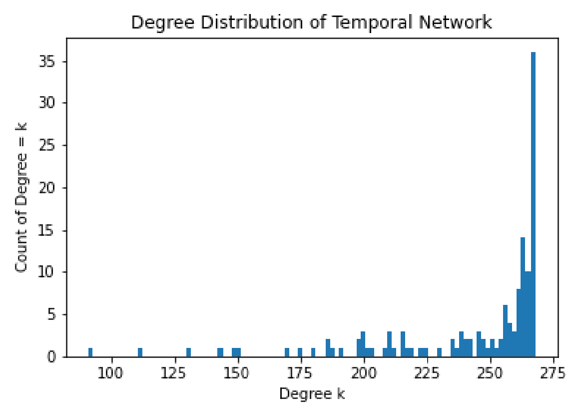

Fig. 2. Temporal Network Degree Distribution

V. Results

A. Temporal Network

Result Analysis: 1) The network has 135 nodes, representing distinct crime types, and 16,415 edges, indicating a vast array of sequential crime occurrences. The average degree is 243.19, which implies that each crime type can be followed by a considerable variety of other crimes, reflecting the interaction of criminal activities within Los Angeles.

2) The degree distribution in Figure 2 has a long-tail distribution, where a small number of nodes have a very high degree, while the majority have lower connectivity. This pattern is indicative of a scale-free network, suggesting the presence of highly influential crime types that could play a critical role in the structure and evolution of the network.

3) The visual representation of the network in Figure 1 is a directed network with weighted edges. It shows its dense connectivity, with a visible central core of nodes that are highly interconnected as nodes in the middle tend to have more edges. This is further supported by the high average clustering coefficient of 0.943 and a global clustering coefficient of 0.936, pointing to a robust network with tightly clustered groups of crime types.

4) The network's efficiency is emphasized by the average shortest path length of 1.09 and a diameter of 2, demonstrating that any crime type is closely linked to another, allowing for rapid progression within the crime sequence. The k-core analysis, which yielded the largest k-core of 113 nodes with a vertex connectivity of 102, suggests a strong backbone of the network comprising a subset of crime types with high connectivity.

Significance & Limitations: These results are significant because they help understand the crime dynamics in Los Angeles, revealing the potential for predicting crime sequences and understanding the conditions that caused these patterns. The high degree of connectivity and clustering within the network can be valuable in identifying key crime types that

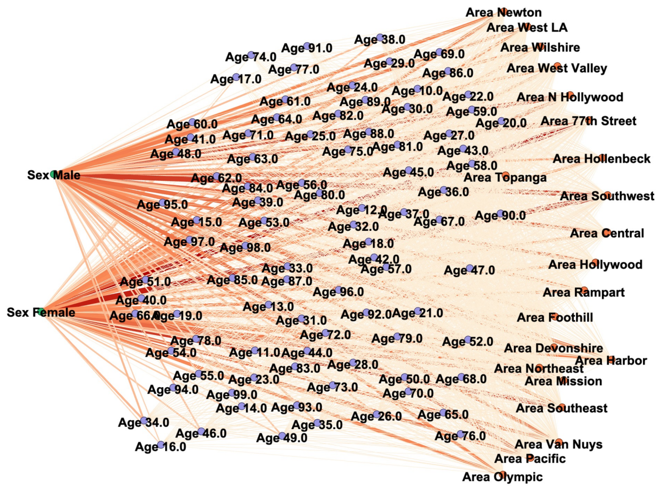

Fig. 3. Network of Sex, Age, and Area

could be targeted for intervention to disrupt the pattern of crime occurrences.

However, the study has its limitations. The network's complexity, with a high number of edges relative to nodes, may lead to a visualization that is difficult to interpret. Moreover, while the connectivity measures indicate a robust network, they do not provide information on the resilience of the network to targeted attacks on specific nodes or edges, which could be crucial in practical crime prevention strategies.

In summary, the network analysis provides a profound understanding of the interconnected nature of crime types. Further analysis is required to fully understand the implications of these patterns for law enforcement and urban policy-making.

B. Network: Sex, Age, and Area

Result Analysis: 1) The constructed network in Figure 3 comprises nodes that represent distinct demographic groups by age and sex, as well as the areas where crimes have occurred. The edges, totaling in number provided previously, signify the crime incidents linking these demographics to areas. An average degree of 37.34 indicates extensive interrelations within the network, suggesting that each demographic category is involved in a large number of incidents across different regions, indicating a widespread distribution of crime across the city's landscape.

2) The clustering coefficients with an average of 0.1458 and a global coefficient of 0.0923, though not as high as in tightly-knit networks, still indicate a moderate level of clustering. This suggests that there are indeed pockets within the network where crimes involving certain ages and sexes are more common, potentially pointing to targeted crime patterns or social vulnerabilities.

3) By iterating over the area nodes and their neighboring nodes representing age and sex, the weights of the edges—which signify the number of crimes—are summed to provide a total crime count for each demographic within each area. From this analysis, the area with the highest number of reported crimes for each sex is identified. The results indicate that "Area 77th Street" has the most reported crimes for both males and females, with a total of 53,995 crimes involving female victims and 41,000 involving male victims. This outcome suggests that "Area 77th Street" is a significant hotspot for crime within the network and may require special attention from law enforcement and public safety officials.

Significance & Limitations: The network analysis of crime incidents in Los Angeles, focusing on the interplay of victim age, sex, and geographic location, has significant implications for urban crime understanding. Metrics such as the average degree and clustering coefficients offer insights into the complex connections between demographics and crime locations. These findings are instrumental for law enforcement in resource allocation and for policymakers in crafting targeted crime prevention strategies.

However, the study's results are subject to limitations. The bipartite structure of the network might oversimplify the complex nature of crime interactions, reducing multifaceted relationships to binary links. The prominent degree of the sex nodes, a byproduct of their central role in the network's architecture, may conceal intricate patterns among age and area nodes. The analysis is also limited by its exclusion of unreported crimes, potentially underrepresenting the actual extent and patterns of crime within certain demographics or regions.

The static analysis presented here does not account for the dynamic nature of crime over time. As criminal activities evolve and are influenced by a multitude of temporal factors, this time-insensitive approach may not capture the full picture. Additionally, the lack of differentiation between various types of crime or their severity levels may limit the depth of understanding into criminal behavior and patterns.

In essence, this network analysis sheds light on the interconnected nature of crime relative to victim demographics and geographic distribution, but it serves as an initial step. To strengthen the findings, future research should incorporate dynamic temporal data, crime severity levels, and unreported crime statistics. Such comprehensive studies could provide more detailed implications for law enforcement actions and urban policy development.

C. Network: Crime type and weapon used and projection

Result Analysis: 1) Figure 4 displays the bipartite network representing crime type and weapon used. The network has 199 nodes, includes 121 nodes of crime type (pink nodes) and 78 nodes of weapon type (blue nodes). Figure 5 displays the weapon-weapon projection network (projection network building process can be found at section IV.C). The average degree of the weapon-weapon projection network is 70, which is notably high. It suggests that, on average, each weapon is connected to almost every other weapon in the network. In practical terms, this means that most weapons have been used together with most other weapons in different criminal incidents.

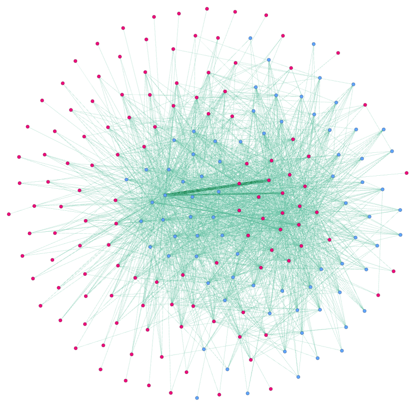

Fig. 4. Network of Crime type and weapon used

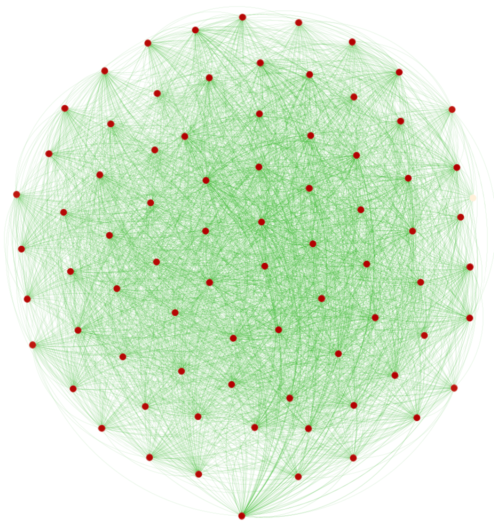

Fig. 5. Projection network of weapon used

2) The top 5 nodes with highest weight degree is calculated. STRONG-ARM, Weighted Degree: 1807; VERBAL THREAT, Weighted Degree: 1723; KNIFE WITH BLADE 6INCHES OR LESS, Weighted Degree: 1681; HAND GUN, Weighted Degree: 1637; OTHER KNIFE, Weighted Degree: 1560; This implies that these weapons are not only connected to many other types of weapons but also frequently appear alongside them in criminal activities.

3) The top 3 weapon pairs with highest weight is calculated. VERBAL THREAT with STRONG ARM, 86; VERBAL THREAT with HAND GUN, 64; VERBAL THREAT with KNIFE WITH BLADE 6INCHES OR LESS, 58. The co-occurrence of "Verbal Threat" with physical means or weapons suggests a pattern where intimidation and physical violence or the threat of violence go hand-in-hand.

Significance & Limitations: The network analysis from this weapon-weapon projection network indicates that there is a significant and diverse weapon usage in crime cases in LA. The presence of a variety of weapons with high weighted degrees indicates a diverse arsenal used in criminal activities, from knives and guns to physical force. The prominence of STRONG-ARM and VERBAL THREAT suggests that physical force and intimidation are key elements in many crimes, sometimes even more so than conventional weapons.

These findings can inform law enforcement and policymakers about the most common types of weapons and weapon combinations used in crimes. The high incidence of physical force and intimidation highlights the need for strategies that address not only conventional weapons but also crimes involving bodily force and verbal threats.

However, the projection from a bipartite network can certainly lose some information. This is primarily due to the shift in focus from the relationships between two different types of entities (crimes and weapons) to the relationships between entities of a single type (weapons).

Firstly, in the original bipartite network, each edge directly connects a specific crime type to a specific weapon, preserving the context of which weapons are used in which crimes. The projection only retains information about which weapons are used together, not in which crimes they are used together.

Secondly, some complex patterns, such as if certain weapons are used together only in specific types of crimes, are not discernible in the projection. The projection might also obscure nuances, like if a weapon is commonly used in many different types of crimes or only in a few specific ones.

VI. Conclusion and Short-Term Plans

Conclusion: 1) The temporal network analysis of Los Angeles crime data has revealed complicated patterns of crime sequencing and interconnectivity. With 135 nodes and over 16,000 edges, the network's structure is dense, indicating that each crime type is potentially followed by a diverse array of other crimes. The high average degree of 243.19 reflects this complexity. The clustering coefficients—average at 0.943 and global at 0.936—suggest a tightly clustered network, suggesting potential groupings of crime types that frequently occur in close succession. The short average path length of 1.09 and small network diameter of 2 reveal the close connections within the crime types, suggesting an efficient progression from one crime to another.

2) The investigation into crime incidents in Los Angeles using network analysis has revealed a nuanced interplay of demographic factors and crime across the urban expanse. The substantial average degree indicates a dense network, signifying that various demographics and neighborhoods are not isolated but part of a complex network of criminal activity. The network's considerable connectivity is further indicated by the modest clustering coefficients, the low average shortest path length, and the small diameter, all suggesting that crime incidents are interlinked throughout the city, affecting a broad spectrum of the population. This analysis provides a critical foundation for strategic, data-driven approaches to crime prevention and community safety, laying the groundwork for future in-depth research.

3) In addition in this study on Los Angeles crime, a meticulous development of a bipartite network linking crime types to weapons used was undertaken, and further projected into a weapon-weapon network to analyze weapon co-occurrences. The findings reveal a diverse and interconnected web of weapon usage, with STRONG-ARM, VERBAL THREAT, and various firearms and knives frequently appearing together in criminal activities, as indicated by the network's high average degree and the prevalence of certain weapon pairs. These insights are crucial for informing law enforcement and policy-making, highlighting the need for strategies that encompass both conventional and non-conventional weapons. While the projection method offers significant understanding into weapon relationships, it also abstracts specific crime-weapon contexts, underscoring the complexity and multifaceted nature of criminal activities. Therefore, this study contributes not only to criminological research but also aids in shaping more effective crime prevention and public safety policies.

Milestone 2 Plans: 1) For Milestone 2, the analysis of the **temporal network** will concentrate on exploring the scale-free nature of the crime network to identify key nodes that significantly influence the network's robustness. This will involve an in-depth analysis of the network's hubs and the role they play in crime dynamics. Sub-networks will be extracted for a focused study on significant crime patterns, and resilience testing will be conducted to evaluate the impact of targeted interventions. Community detection will also be utilized to identify clusters that represent interconnected crime activities. These efforts aim to enhance predictive capabilities and inform effective crime prevention strategies.

2) The analysis of the **Network: Sex, Age, and Area** will be extended to include the LA geographical data. The results derived from current Milestone will be visualized on the map of LA and hence gives people a better idea of how certain areas become more dangerous regarding people of certain sex and age.

3) The analysis of the **Network: Crime type and weapon used and projection** will be expanded to include a crime-crime projection network, offering an additional perspective for further investigation. This extension aims to uncover clusters of crime types that demonstrate close relationships in terms of weapon usage. Such findings may suggest similarities in underlying motives or methods across different crimes. Furthermore, there are plans to undertake a sub-network analysis, focusing on specific categories of crimes. This approach is anticipated to yield insights into the unique patterns of weapon preferences and associations within distinct crime categories.

Contributions: Until this milestone, the workload has been distributed evenly among the three members. During the preparation phase, Yuke focused on data cleaning and had columns ready to be analyzed. Meanwhile, Dongchen and Wenxi worked on finding prior works related to this project. During milestone 1 and the current milestone, Yuke worked on the temporal network. Wenxi worked on the network of sex, age, and area names. Dongchen worked on the network of crime types and weapons used and projection. All three members worked together on writing the joint part of this report.

References

[1] Burcher, M. (2020). Social network analysis and the characteristics of Criminal Networks. *Social Network Analysis and Law Enforcement*, 95–129. https://doi.org/10.1007/978-3-030-47771-4_4

[2] Campana, P. (2016). Explaining criminal networks: Strategies and potential pitfalls. *Methodological Innovations*, 9, 205979911562274. https://doi.org/10.1177/2059799115622748

[3] Keyvanpour, M.R., Javideh, M., & Ebrahimi, M.R. (2011). Detecting and investigating crime by means of data mining: a general crime matching framework. *Procedia Computer Science*, 3, 872-880. https://doi.org/10.1016/j.procs.2010.12.143

[4] Liu, X., Patacchini, E., Zenou, Y., & Lee, L.-F. (2012, June 22). Criminal Networks: Who is the Key Player? *FEEM Working Paper* No. 39.2012. https://ssrn.com/abstract=2089267 or http://dx.doi.org/10.2139/ssrn.2089267

[5] Rostami, A.,& Mondani, H. (2015). The complexity of Crime Network Data: A case study of its consequences for crime control and the study of networks. *PLOS ONE*, 10(3). https://doi.org/10.1371/journal.pone.0119309

[6] Schwartz, D.M., & Rouselle, T. (2009). Using social network analysis to target criminal networks. *Trends in Organized Crime*, 12, 188–207. https://doi.org/10.1007/s12117-008-9046-9

[7] Wheeler, A. (2014). Creating high crime sub-tours. Andrew Wheeler. https://andrewpwheeler.com/tag/networkx/