Exploring Crime Linkage Systems: Improving Criminal Investigation Efficiency by Network Science

You can access the code and results in the publicly accessible repository here:
[https://github.com/YueZheng2000/8880final_u7564091]

1. Introduction

1.1 What is Crime Linkage Systems?

Crime linkage systems, often visualized as networks where nodes represent individuals involved in crimes and edges represent the relationships between them, are sophisticated analytical tools employed by law enforcement agencies to identify connections across various criminal activities. These systems integrate and analyze data from diverse sources, including crime reports, forensic evidence, witness statements, and surveillance footage. By applying statistical and data mining techniques, crime linkage systems reveal patterns and correlations that typically elude traditional investigative methods. This functionality not only assists in identifying serial offenders but also in linking disparate cases that share similar characteristics or modi operandi. Consequently, these systems significantly enhance the ability to predict and prevent future crimes.

1.2 Why we need Crime Linkage Systems?

In the face of an increasingly complex crime landscape, traditional investigative methods often fall short, unable to cope with the sheer volume and intricacy of modern criminal activities. Crime linkage systems are essential as they leverage advanced data analytics to synthesize information from various sources—such as crime reports, forensic evidence, and surveillance data—thereby identifying patterns and connections between incidents that might initially appear unrelated.

These systems are crucial for several reasons. First, they enhance the efficiency of criminal investigations by allowing for the rapid detection and accurate resolution of crime, which is especially vital in a landscape where time is often of the essence. Second, by identifying and clustering related criminal activities, crime linkage systems enable law enforcement agencies to pinpoint core criminals and key crime series. This not only aids in immediate case resolution but also in strategic crime prevention, as recognizing patterns allows for the anticipation and thwarting of potential future crimes.

Ultimately, the implementation of crime linkage systems significantly reduces crime rates and boosts public safety by enabling proactive rather than reactive responses to criminal threats. Thus, these systems do not merely assist in solving crimes; they transform the approach to crime prevention and investigation, making it more strategic and effective.

1.3 The Crime Linkage System in the project.

The primary objective of this project is to construct a crime linkage system, beginning with the selection of an appropriate dataset. The chosen dataset for this project is the 'moreno_crime' dataset, published at http://konect.cc/networks/moreno_crime/. This is a bipartite, undirected network that is unweighted and contains no multiple edges, representing criminal records in a town call Moreno. It includes individuals who have appeared in at least one crime case, whether as a suspect, victim, witness, or in multiple roles. In this network, one type of node represents individuals, and the other represents crimes; edges indicate involvement in crimes. From this dataset, we aim to generate a new network: a weighted, undirected network where each node represents an individual, and the links (weighted by the number of crimes committed together) depict the criminal connections between individuals.

1.4 What are we trying to get in Crime Linkage Systems?

In addition to constructing criminal networks, our crime linkage system aims to derive several critical insights from the network analysis. These objectives include:

Analyzing Co-crime Relationships: By examining the attributes of the network, such as the strength and frequency of connections, we can identify and describe the nature of relationships between co-offenders. This analysis helps in understanding how criminals collaborate and the dynamics of their relationships.

Identifying Influential Criminals: Utilizing measures like centrality in the network, we aim to pinpoint influential criminals who are central to the network by virtue of participating in a high number of crimes. These individuals are often key nodes in the criminal network, whose detection might disrupt multiple criminal activities.

Mapping Close Associations: The network will also be used to explore close criminal associations from specific individuals. By tracing links from a given individual, we can uncover their direct and indirect connections, providing insights into the structure of their criminal networks.

Determining Criminal Involvement in Specific Crimes: From the data on specific crimes, the system will help determine which individuals are most central or pivotal to criminal activities, potentially distinguishing between major and minor offenders. This is crucial for prioritizing investigative resources and interventions.

Each of these outputs from the crime linkage systems not only aids in the immediate resolution of cases but also enhances strategic planning for future crime prevention and law enforcement initiatives. Through the detailed analysis of criminal networks, law enforcement agencies can gain a deeper understanding of criminal behaviors, anticipate potential threats, and efficiently allocate resources to mitigate crime.

2. Related Work and Innovation Point

2.1 Related Work

Crime linkage, as a methodological approach in criminal investigations, relies on the analysis of crime scene behaviors to link two or more crimes believed to be committed by the same offender (Woodhams, Hollin, & Bull, 2007). Over the past decade, this field has seen a significant increase in scholarly attention, with researchers focusing on validating the fundamental assumptions of behavioral consistency and distinctiveness (Bennell & Canter, 2002). These principles are critical as they require that an offender's behavior be consistent across crimes to be identifiable, yet distinctive enough to differentiate from other offenders (Woodhams & Bennell, 2014).

The practical application of crime linkage spans various crime types and is instrumental in police investigations of serious offenses such as sexual assaults and homicides, as well as in burglary, robbery, and vehicle theft (Burrell & Bull, 2011; National Crime Agency, n.d.).

Moreover, in certain jurisdictions, crime linkage also supports legal proceedings, adding a significant dimension to its utility (Labuschagne, 2006, 2012; Pakkanen, Santtila, & Bosco, 2014). The approach is particularly advantageous when traditional methods are untenable due to high costs, time constraints, or the absence of physical forensic evidence (Pakkanen, Zappalà, Grönroos, & Santtila, 2012; Grubin, Kelly, & Brunsdon, 2001).

Empirical studies support the application of these principles across a range of crime types, highlighting the versatility and breadth of crime linkage. Research has shown its effectiveness in contexts ranging from burglary and robbery to arson and homicide, employing various methodologies and drawing on data from multiple countries (Bennell & Canter, 2002; Salfati & Bateman, 2005; Santtila et al., 2008). However, these findings also caution that the principles do not uniformly apply to all offenders or in all contexts, suggesting a nuanced application of crime linkage (Woodhams & Labuschagne, 2012).

Importantly, while the theoretical support for crime linkage is robust, translating these theories into effective practice poses challenges. The use of crime linkage in legal contexts, for instance, requires meeting stringent standards of admissibility in some countries, highlighting the need for peer-reviewed, widely accepted practices within the scientific community (Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993).

Given the potential ramifications of erroneous crime linkage, such as misallocation of resources or undue public alarm, it is crucial that crime linkage practices are both efficient and accurate. This necessitates ongoing research focused not just on the theoretical underpinnings but also on the practical implementation of crime linkage across different policing contexts worldwide (Rainbow, 2014; Hazelwood & Warren, 2004).

In conclusion, the literature underscores the importance of crime linkage in modern policing, providing valuable insights into its capabilities and challenges. As crime continues to evolve, so too must the methodologies and technologies support crime linkage, ensuring they are as effective in practice as they are promising in theory.

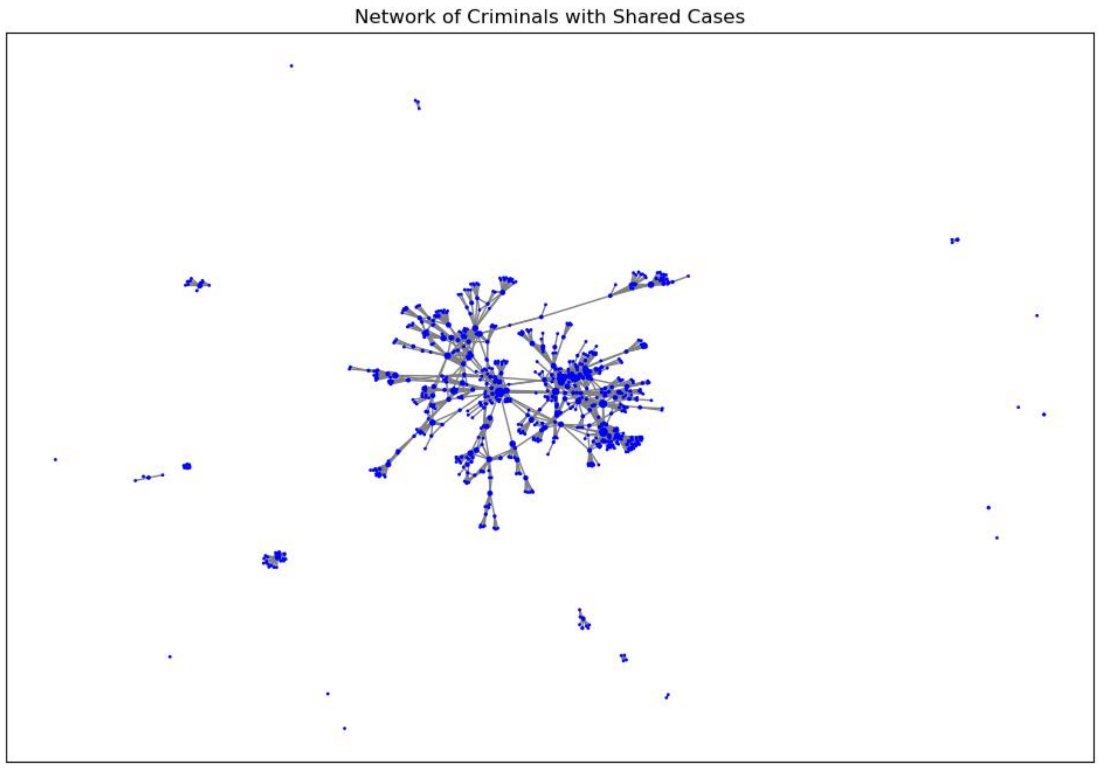
2.2 Innovation Point

In reviewing the literature on crime linkage systems, it is evident that most systems focus on establishing networks based on relationships between crime cases. However, this project proposes a novel approach by constructing a network based on the relationships between criminals. This approach shifts the focus towards understanding the motivations and the individuals involved in crimes. By doing so, it aims to identify key individuals within each crime and uncover central figures within the entire network. This method could potentially enhance our understanding of criminal networks by highlighting the roles and connections of the individuals within them.

3. Empirical Approach (Methods, Results and Findings)

3.1 The Criminal Relation Network

As outlined in the introduction, our project features a weighted, undirected network where each node represents an individual. The links, weighted by the number of crimes committed together, illustrate the criminal connections between these individuals. The network diagram below visualizes these relationships:

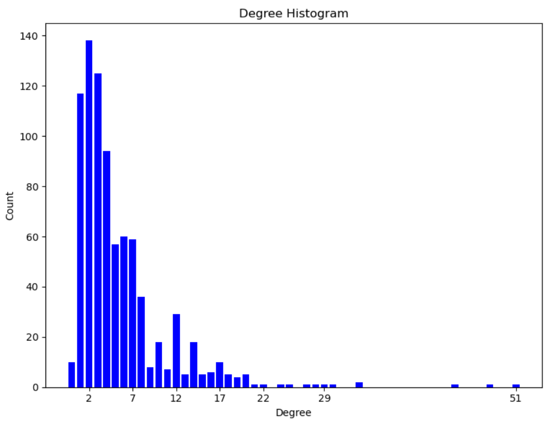


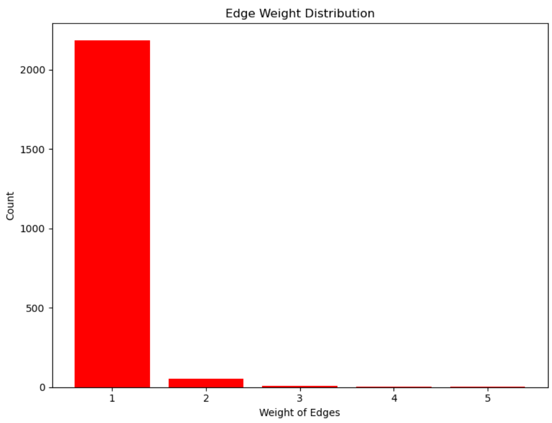
The basic attributes of the network are shown below:

Attribute	Value
Number of Nodes	829
Number of Edges	2253
Average Degree	5.435464414957781
Density	0.0065645705494659186
Average Clustering Coefficient	0.7190775294689993

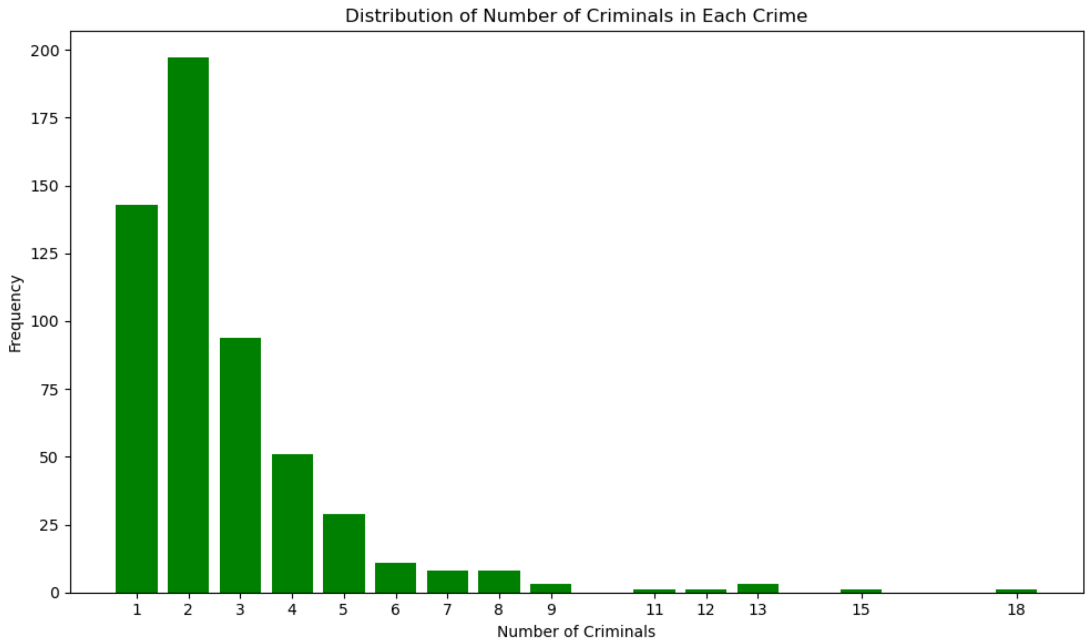
The network analysis reveals an average degree of approximately 5.435, indicating that each individual in the network is connected to about five others on average, which reflects a moderate level of connectivity. Despite this, the overall network density is quite low at 0.0065645705494659186, suggesting that the network is sparse and that these connections are not widespread across the entire network. Instead, they are likely concentrated within specific clusters or groups. This is further supported by the high average clustering coefficient of 0.719, indicating that these groups are tightly knit, with members within a cluster likely to be closely interconnected. This clustering suggests the presence of well-organized criminal cells or factions that collaborate closely, which might operate more effectively within their confined groups rather than across the broader network. This structure implies that while many criminals may only interact with a select few partners, these relationships are highly integrated, potentially facilitating more coordinated criminal activities within these clusters.

To get further understanding of the network, we compute the degree distribution and weight distribution of the network:





It is noticeable that most edges have a weight of 1, indicating that most criminals have only collaborated with another individual once. However, the degree distribution reveals that many criminals have a significantly higher degree, averaging 5.44. To investigate the cause of this discrepancy, we computed the distribution of the number of criminals involved in each crime, as shown below:



The average number of criminals per crime is approximately 2.68. Given that the number of potential connections (edges) that can be formed in a crime involving n criminals is $0.5 \times n \times (n-1)$, the discrepancy between the average degree and the weights may be attributed to crimes that involve many criminals, thereby significantly increasing the degree.
Based on the attributes of our network, we observe that although the criminal network in Moreno is somewhat sparse, most criminals have collaborated with at least one other individual. However, these collaborations tend not to be repeated—most result in just a single instance of cooperation.

3.2 Influential Criminals in the Network

To identify influential criminals in our network, we first need to define what makes a criminal "influential." In this project, a criminal is considered influential if they are involved in numerous crimes. Our goal, then, is to identify the fewest number of criminals

necessary to encompass as many of the crimes as possible.
To achieve this, we use a greedy algorithm with a threshold aiming to encompass at least 80% of the crimes. The main steps of our algorithm are outlined below:

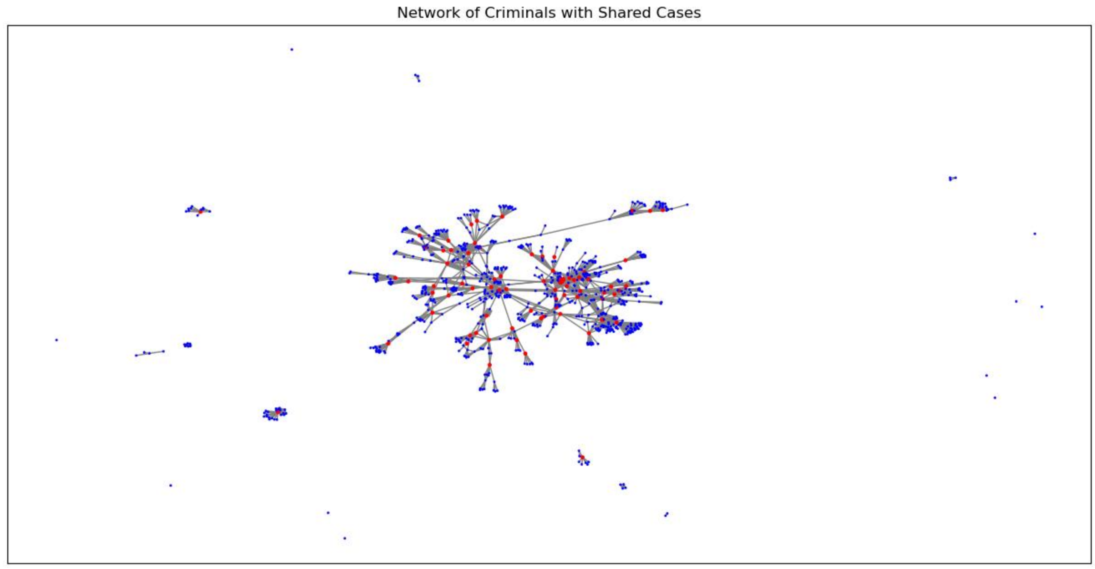
Step 1: Add the list of crimes as an attribute for each node.

Step 2: Select the criminal involved in the highest number of crimes.

Step 3: Remove the selected node and all associated crime IDs.

Step 4: Repeat step 2 until 80% of the crimes (equivalent to 440 out of 551 crimes) have been accounted for.

Using this algorithm, we found that 80 out of 829 criminals can encompass 80.04% of the crimes in the network. The influential criminals are marked in red in the network diagram below:



This leads to our second finding: a small group of criminals can encompass most of the crimes in the network, suggesting that targeting key criminals could be an effective strategy for crime prevention.

3.3 Close Association Mapping

In our project, it is crucial to determine the relationships between different criminals. To achieve this, we employ the Random Walk with Restart (RWR) algorithm to calculate the association between two nodes. This method is not only effective but also provides a nuanced understanding of network dynamics. The main steps of our RWR implementation are outlined below:

Step 1: Start from the node of interest (a specific criminal), initializing the walk probability entirely at this node.

Step 2: At each step, move to an adjacent node based on the edge weights (transition probabilities) or restart at the original node with a predefined probability. This ensures a blend of exploration and exploitation, maintaining focus around the starting node.

Step 3: Repeat the process until the distribution of probabilities across the nodes stabilizes (converges). This stationary distribution reflects the likelihood of each node being reached from the starting node, representing their association strength.

Step 4: Extract the probability values from the final distribution to determine the strength

of association between the starting node and all other nodes. Nodes with higher probabilities are considered to have a stronger association with the starting node.
As an example, the project selected the criminal with ID '100' and used the RWR algorithm to identify the top 5 nodes most closely associated with '100'. The result was: [522, 208, 759, 101, 244].

3.4 Key Criminal in a Crime

After establishing a method to measure association, we can calculate the average RWR for each node relative to every other node within the case. This approach allows us to identify which criminal is most associated with other criminals in the case, thus having a higher probability of being a key figure in the crime.
As an example, case 42 was processed using the algorithm. The resulting values for nodes [12, 13, 14, 556] were [0.1771, 0.1020, 0.1020, 0.1020], indicating that the criminal with ID 12 has the highest probability of being the key criminal.

4. Further Thinking

4.1 Broader Impact

The development and application of crime linkage systems utilizing network analytical methods might have some broader impacts in two main aspects below:
Enhanced Crime Solving Capabilities: By employing sophisticated techniques to identify relationships between criminals and their activities, law enforcement agencies can solve crimes more efficiently. This can lead to a higher rate of crime resolution, reducing the backlog of unsolved cases and providing justice more swiftly.
Data-Driven Policymaking: The insights gained from advanced crime linkage systems can inform data-driven policymaking. Policymakers can use the data to understand crime patterns and trends, leading to more effective laws and regulations aimed at combating crime.

4.2 Ethical Considerations

The implementation of crime linkage systems, while beneficial, raises several ethical considerations that must be addressed to ensure responsible use. Firstly, the collection and analysis of extensive data on individuals, including suspects, victims, and witnesses, necessitate strict adherence to privacy and data protection laws. It is essential to implement robust data security measures to prevent unauthorized access and misuse of sensitive information. Secondly, there is a risk of reinforcing biases within the criminal justice system. Algorithms used in crime linkage systems must be designed and monitored to avoid perpetuating existing biases against certain groups or communities. Thirdly, the potential for over-reliance on automated systems poses a concern; human oversight is crucial to interpret results accurately and to make fair, informed decisions. Finally, transparency in how these systems operate is vital for maintaining public trust. Law enforcement agencies should ensure that the methodologies and algorithms used are transparent and explainable to the public. Balancing the benefits of advanced crime linkage systems with these ethical considerations is imperative to maintain public confidence and ensure justice is served equitably.

4.3 Limitation

The dataset chosen causes the two main limitations in the project.
Firstly, the dataset lacks temporal information, which poses a significant challenge for implementing methods that rely on time steps. Temporal data is crucial in crime linkage systems as it helps in understanding the sequence and evolution of criminal activities over time.
Secondly, the dataset contains information from only a single town. This geographic limitation means that the conclusions drawn from the analysis may not be applicable to other regions or broader contexts. Despite this limitation, the findings are still valuable and provide insights that can be learned from and potentially adapted for use in different settings.

5. Future work

Future work on this project could focus on several key areas to enhance the effectiveness and applicability of the crime linkage system:

5.1 Utilize Larger and More Diverse Datasets

Expanding the dataset to include more diverse and larger datasets from multiple towns, cities, or even national databases would help in generalizing the findings. A more extensive dataset would allow for a more comprehensive analysis, potentially uncovering broader patterns and improving the robustness of the crime linkage system.

5.2 Incorporate Temporal Data

Integrating temporal data into the analysis is crucial. Future work should aim to include datasets that provide timestamps for criminal activities. This would enable the application of time-series analysis and temporal network models, offering deeper insights into the dynamics of criminal behavior over time.

5.3 Enhance Algorithmic Approaches

Exploring and incorporating more advanced algorithmic approaches could further improve the accuracy and efficiency of the crime linkage system. Techniques such as machine learning and deep learning could be used to identify more complex patterns and relationships within the data.

5.4 Improve Ethical and Bias Mitigation Strategies

Future work should also focus on enhancing ethical considerations and bias mitigation strategies. Developing more transparent and fair algorithms, alongside robust privacy protection mechanisms, is essential to ensure the ethical use of crime linkage systems.

Citation:

1. KONECT. (n.d.). Moreno Crime Network. Retrieved May 8, 2024, from http://konect.cc/networks/moreno_crime/
2. Woodhams, J., Hollin, C. R., & Bull, R. (2007). Case linkage. In R. N. Kocsis (Ed.), *Criminal profiling: International perspectives in theory, practice and research* (pp. 117–133). New Jersey: Humana Press.
3. Bennell, C., & Canter, D. V. (2002). Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. *Science & Justice, 42*, 153–164.
4. Woodhams, J., & Bennell, C. (Eds.) (2014). *Crime linkage: Theory, research, and practice*. Boca Raton, FL: CRC Press.
5. Burrell, A., & Bull, R. (2011). A preliminary examination of crime analysts' views and experiences of comparative case analysis. *International Journal of Police Science and Management, 13*, 2–15.
6. Labuschagne, G. N. (2006). The use of a linkage analysis as evidence in the conviction of the Newcastle serial murderer, South Africa. *Journal of Investigative Psychology and Offender Profiling, 3*, 183–191.
7. Labuschagne, G. N. (2012). The use of a linkage analysis as an investigative tool and evidential material in serial offenses. In K. Borgeson, & K. Kuehnle (Eds.), *Serial offenders: Theory and practice* (pp. 187–215). Sudbury, MA: Jones & Bartlett Learning.
8. Pakkanen, T., Santtila, P., & Bosco, D. (2014). Crime linkage as expert evidence: Making a case for the Daubert Standard. In J. Woodhams, & C. Bennell (Eds.), *Crime linkage: Theory, research, and practice* (pp. 225–250). Boca Raton, FL: CRC Press.
9. Pakkanen, T., Zappalà, A., Grönroos, C., & Santtila, P. (2012). The effects of coding bias on estimates of behavioural similarity in crime linking research of homicides. *Journal of Investigative Psychology and Offender Profiling, 9*, 223–234.
10. Grubin, D., Kelly, P., & Brunsdon, C. (2001). Linking serious sexual assaults through behaviour. UK: Home Office. https://doi.org/10.1037/e649492007-001
11. Salfati, C. G., & Bateman, A. L. (2005). Serial homicide: An investigation of behavioural consistency. *Journal of Investigative Psychology and Offender Profiling, 2*, 121–144.
12. Santtila, P., Pakkanen, T., Zappala, A., Bosco, D., Valkama, M., & Mokros, A. (2008). Behavioural crime linking in serial homicide. *Psychology, Crime & Law, 14*, 245–265.
13. Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993). 509 U. S 579 Supreme Court (US). Retrieved from: https://supreme.justia.com/cases/federal/us/509/579/case.html.
14. Rainbow, L. (2014). A practitioner's perspective: Theory, practice, and research. In J. Woodhams, & C. Bennell (Eds.), *Crime linkage: Theory, research, and practice* (pp. 173–194). Boca Raton, FL: CRC Press.
15. Hazelwood, R. R., & Warren, J. I. (2004). Linkage analysis: Modus operandi, ritual, and signature in serial sexual crime. *Aggression and Violent Behavior, 9*, 307–318.

Work Contribution:

Done by a single person, Yue Zheng u7564091.