Techniques to Detect Crime Leaders within a Criminal Network: A Survey, Experimental, and Comparative Evaluations

Kamal Taha, *Senior Member, IEEE* and Abdulhadi Shoufan, *Senior Member, IEEE*

This survey paper offers a thorough analysis of techniques and algorithms used in the identification of crime leaders within criminal networks. For each technique, the paper examines its effectiveness, limitations, potential for improvement, and future prospects. The main challenge faced by existing survey papers focusing on algorithms for identifying crime leaders and predicting crimes is effectively categorizing these algorithms. To address this limitation, this paper proposes a new methodological taxonomy that hierarchically classifies algorithms into more detailed categories and specific techniques. The paper includes empirical and experimental evaluations to rank the different techniques. The combination of the methodological taxonomy, empirical evaluations, and experimental comparisons allows for a nuanced and comprehensive understanding of the techniques and algorithms for identifying crime leaders, assisting researchers in making informed decisions. Moreover, the paper offers valuable insights into the future prospects of techniques for identifying crime leaders, emphasizing potential advancements and opportunities for further research. Here's an overview of our empirical analysis findings and experimental insights, along with the solution we've devised: (1) PageRank and Eigenvector centrality are reliable for mapping network connections, (2) Katz Centrality can effectively identify influential criminals through indirect links, stressing their significance in criminal networks, (3) current models fail to account for the specific impacts of criminal influence levels, the importance of socio-economic context, and the dynamic nature of criminal networks and hierarchies, and (4) we propose enhancements, such as incorporating temporal dynamics and sentiment analysis to reflect the fluidity of criminal activities and relationships, which could improve the detection of key criminal figures as their roles or tactics evolve.

***Index Terms*— Crime Leader Identification, Criminal Networks Analysis, Algorithmic Techniques in Criminology, Network Centrality Measures, Forensic investigation, digital forensic, Methodological Taxonomy of Crime Analysis.**

I. INTRODUCTION

Due to societal progress, organized crime has emerged as the primary type of criminal structure. Criminal organizations now operate within intricate social networks, making it challenging to distinguish between innocent individuals and members involved in criminal activities due to limited data availability [85]. In real-world investigations, while some conspirators are known and others are not, the goal is to ascertain the involvement of uncertain members and pinpoint the leaders prior to making arrests.

In criminal investigations, the challenge lies in mapping the criminal network's structure to identify its leaders and participants before proceeding with arrests. Criminal groups often share similarities through friend-of-a-friend relationships, co-offending experiences, referral chains, and the need for specialized expertise [49]. Connections between individuals can be categorized as strong or weak ties, representing different levels of interaction. Strong ties are close and trusted relationships, while weak ties are more distant, like co-workers [69].

Criminal groups often share similarities through friend-of-a-friend positions, co-offending experiences, referral chains, and the need for specialized expertise [28]. Individual connections range from strong, trust-based relationships with family and friends, to weaker ties like those with acquaintances or colleagues. Both strong and weak ties have their own pros and cons [69]. Crime data is classified by crime type [70]. Analyzing crime by category aids in crime prevention and reduction. Organizational structures and common locations affect crime frequency. Studying crime patterns over time highlights hotspots and helps predict and reduce future incidents. In-depth analysis of structured crime data enhances our grasp of criminal activities, with historical records pinpointing key areas of concern.

Within criminal networks, brokers play an even more critical role due to the absence of formal regulations and mechanisms governing transactions and conflicts in stateless environments [41, 42, 62, 63]. Success within criminal organizations relies heavily on social connections that provide access to profitable opportunities [45, 41]. In contemporary times, criminals must strike a delicate balance between efficiently managing illicit activities and ensuring the security of the group [50]. Criminal leaders act as brokers in their networks, with higher betweenness centrality scores indicating their strategic role [29].

In organized crimes, leaders serve as bridges between criminals, individuals in businesses, and politics, exploiting these connections for criminal opportunities [42, 47, 61, 79]. Identifying criminal leaders through wiretap data is limited due to cautiousness and minimized telecommunications usage by criminals [2, 9, 21]. Balancing efficiency and security, criminals limit information sharing to avoid detection, with leaders using telecommunications sparingly [48, 50]. Leaders may delegate risky activities to middle-level criminals.

K. Taha is with the Electrical Engineering and Computer Science Department, Khalifa University, UAE (e-mail: kamal.taha@ku.ac.ae).

A. Shoufan is with the Electrical Engineering and Computer Science Department, Khalifa University, UAE (e-mail: abdulhadi.shoufan@ku.ac.ae)

A. Motivations and Key Contributions

1. Main Challenge and Proposed Solution

1) **Current Issue:** Survey papers in the field of algorithms for identifying crime leaders and predicting crimes struggle with effectively categorizing these algorithms. They often use broad and non-specific groupings. This lack of specificity can lead to confusion when classifying unrelated algorithms and result in inaccurate evaluations using the same metrics.
2) **Proposed Solution:** This paper introduces a new methodological taxonomy. It hierarchically classifies algorithms for crime leaders prediction into specific and detailed categories and techniques, enabling a precise and systematic approach to categorization.

B. Comprehensive Survey and Enhanced Assessment

1) **Survey Goals:** The paper provides a detailed survey of algorithms, focusing on those that use *same* sub-techniques, techniques, sub-categories, and categories.
2) **Benefits of the Taxonomy:** Utilizing this taxonomy allows for more accurate assessments and comparisons of algorithms. This leads to a deeper understanding of their strengths and weaknesses and paves the way for future research.

C. Empirical and Experimental Evaluations

1) **Empirical Evaluation:** The paper includes an empirical evaluation, examining various techniques for identifying crime leaders based on four distinct criteria.
2) **Experimental Evaluation:** Through experimental evaluation, this study ranks algorithms, including those that utilize the same sub-technique, different sub-techniques within the same technique, different techniques within the same sub-category, different sub-categories within the same category, and categories.

D. Overall Contributions

1) **Comprehensive Understanding:** The integration of the methodological taxonomy with empirical and experimental evaluations offers researchers a thorough and nuanced understanding of available algorithms.
2) **Informed Decision-Making:** This approach aids researchers in making well-informed decisions about selecting appropriate techniques for specific needs.

B. Proposed Methodology-Based Taxonomy

We categorize crime leaders' identification algorithms into three main classes based on the techniques they use. These three broad classes are topology-based, clustering-based, and agent-based methods. Each of these methods is further subdivided into three tiers, with each tier being more specific than the previous one. Our methodology-based taxonomy is structured hierarchically as follows:

Methodology category → methodology sub-category → methodology techniques → methodology sub-techniques.

This hierarchy allows us to identify specific techniques or sub-techniques in the final level. Fig. 1 shows our methodology-based taxonomy. Our taxonomy offers the following benefits:

- *Enhanced organization:* It offers a well-organized framework for presenting the survey results. By grouping related approaches, the hierarchical structure helps readers to follow the logical flow of the paper.
- *Comprehensive coverage:* The taxonomy provides thorough coverage of all pertinent methods, and its hierarchical design helps identify research gaps and areas needing more exploration.
- *Comparison of techniques:* The taxonomy aids in comparing research techniques by grouping similar methods and highlighting their similarities and differences, allowing for an assessment of their strengths and weaknesses.
- *Improved reproducibility:* The taxonomy enhances research reproducibility by clearly describing approaches, making it easier for other researchers to replicate and build upon its findings.

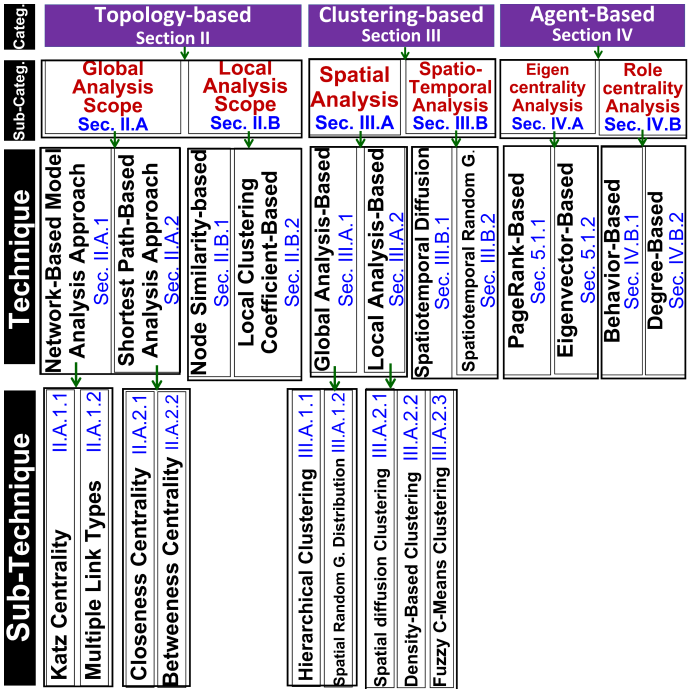

Fig. 1: Our methodology-based taxonomy that categorizes the algorithms for the identification of influential criminals and crime leaders into fine-grained classes in a hierarchical manner, as follows: methodology category → methodology sub-category → methodology technique → methodology sub-technique. For each category, sub-category, technique, and sub-technique, the figure also shows the section number in the manuscript that discusses it.

II. Topology-Based Analysis

A. Global Analysis Scope

Global topology analysis refers to the examination of the overall structure and properties of a criminal social network. Global topology analysis is a valuable approach for identifying influential individuals in a criminal social network. By examining the structural properties of the network, such as node centrality and community detection, we can gain insights into the key players and their roles within the criminal organization.

1. Network-Based Model Analysis Approach

1) Katz Centrality-Based Model

Katz centrality is a useful measure for identifying influential individuals in a criminal social network. It is a variant of centrality that considers both direct connections and indirect connections through a network. The Katz centrality score of a node is based on the sum of its immediate neighbors' centrality scores, weighted by a factor that decreases with the length of the path. Compute the Katz centrality scores for each node using the following equation: $C(v) = \alpha \sum(A(u, v) * C(u)) + \beta$, where $C(v)$ represents the Katz centrality score of node v, $A(u, v)$ is the element in the adjacency matrix at row u and column v, and $C(u)$ is the centrality score of node u. Iterate this equation until convergence, updating the centrality scores for each node. Fig. 2 depicts the procedure of Katz centrality.

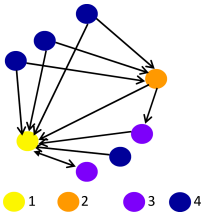

Fig. 2: The procedure of Katz centrality. Each node's legend number indicates the centrality rank of the node

Cavallaro et al. [16] used Katz centrality to identify influential individuals in Sicilian Mafia gangs. They aimed to understand the gangs' structure and organization using real-world datasets, focusing on their resilience to law enforcement. Two networks were created: one from phone call data and another from records of physical meetings within the gangs. Zhang et al. [86] integrated Katz centrality and betweenness centrality to evaluate node significance in networks, including criminal networks. This approach addressed limitations by combining the calculation of shortest paths using betweenness centrality with assigning varying weights to all paths using Katz centrality. This provided a more comprehensive measure of node importance. Calderoni et al. [13] found that Katz scores effectively utilized the entire graph structure and produced accurate results, even with limited network connectivity. They conducted experiments on networks based on meetings and recorded telephone calls among criminals. They concluded that Katz centrality was robust.

2) Multiple Link Types Model

A multiple link types model identifies influential criminals in a criminal social network by considering diverse connections and exchanges. Different link types, like drug exchanges or communication patterns, provide a comprehensive understanding of the network's structure and strategic roles. The model analyzes network data, categorizes link types, and examines both direct and indirect connections. Importance scores or centrality measures, such as eigenvector centrality, are assigned to each criminal based on involvement and influence. This approach reveals influential criminals with significant roles, as their impact extends beyond direct connections to involve multiple link types. Fig. 3 depicts a faction of a hypothetical criminal network with multi-link types.

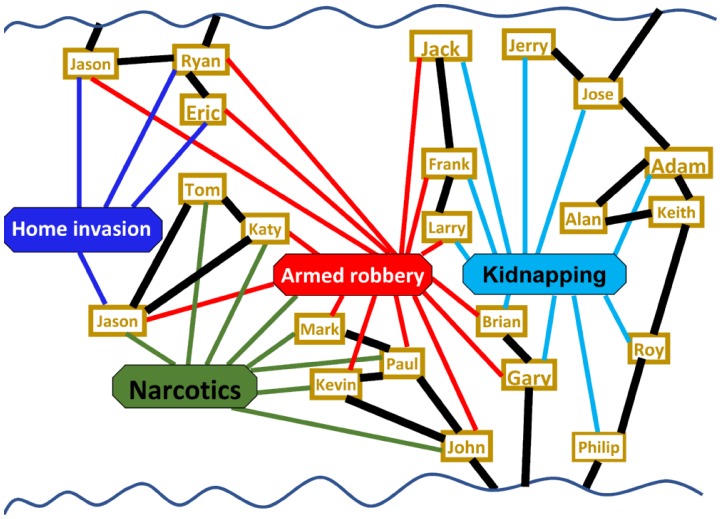

Fig. 3: A faction of a hypothetical criminal network with multi-link types

Bright et al. [9] studied different link types in a drug manufacturing and trafficking network. They found eight link types associated with specific resource exchanges, such as drugs. Examining multiple link types helped them understand the network structure and individuals' strategic roles. Ficara et al. [31] created a multilayer network by adding a third layer based on criminal activities committed collectively. The network had 226 actors, 454 edges within layers, and 3 layers: Meetings, Phone Calls, and Crimes. Analyzing actor and layer measures helped assess significance and dissimilarities. The multilayer approach revealed important actors not evident in separate layer examination. Schwartz and Rouselle [74] suggested choosing the indirect path with the highest indirect connection score in networks with multiple links. They showed that an influential actor's impact can be maximized through indirect connections involving multiple links, rather than a single direct connection. Maulana and Emmerich [55] proposed a method to examine network centrality in multiplex networks. They calculated Pareto fronts of node centrality, with each layer maximizing its own centrality. Dominance rank within a multiplex network indicated a node's significance. They focused on eigenvector centrality in their initial findings.

2. Shortest Path-Based Analysis Approach

2) Closeness Centrality-Based Model

Closeness centrality is a measure used in network analysis to identify influential nodes within a network based on their proximity and accessibility to other nodes. It quantifies how close a node is to all other nodes in the network, considering the shortest paths between them. In the context of a criminal social network, closeness centrality can be used to identify influential criminals based on their level of connectivity and potential control over information flow within the network. Closeness centrality for a node is defined as the inverse of the sum of the shortest path distances between that node and all other nodes in the network. The closeness centrality of a node "v" is defined as the reciprocal of the sum of its distances to all other nodes: $C(v) = 1 / \Sigma\, d(v, u)$, where: $C(v)$ represents the closeness centrality of node "v", $d(v, u)$ denotes the geodesic distance between node "v" and node "u", and Σ represents the summation over all other nodes in the network. Fig. 4 depicts the procedure of closeness centrality.

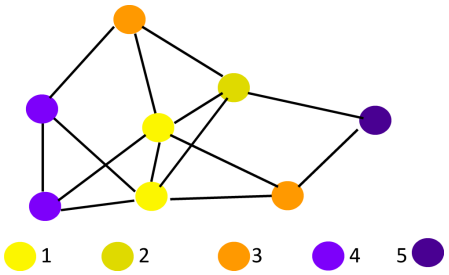

Fig. 4: The procedure of closeness centrality. Each node's legend number indicates the centrality rank of the node

Calderon [15] examined network analysis as a means to identify leaders within a large mafia network. The study focused on data gathered from an extensive investigation of the 'Ndrangheta, a Calabria-based mafia organization in Southern Italy. Operation Infinito successfully uncovered multiple mafia families and monitored their operational meetings. The author employed various metrics, including closeness centrality, degree centrality, and betweenness centrality, in the analysis. Shafia and Chachoo [66] studied the impact of social media platforms, particularly Facebook, on the spread of criminal propaganda. They formed smaller subnetworks to identify individuals and their connections, which required more in-depth examination. They utilized closeness centrality as a key concept. Yang [85] used Social Network Analysis (SNA) multiple times to extract crime networks and identify influential individuals within criminal organizations. The Fisher Discriminant Analysis Method was applied to determine a threshold for categorizing nodes into distinct groups. SNA facilitated crime network mining and the identification of key figures within the network. Memon [53] highlighted the importance of strong connections in evaluating node centrality. The author employed metrics like closeness and betweenness centrality, focusing on identifying the shortest paths and their lengths between nodes. This approach recognized the significance of strong ties.

2) Betweeness Centrality-Based Model

Betweenness centrality is a widely used measure for identifying influential individuals in a network, including a criminal social network. It quantifies the extent to which a node lies on the shortest paths between other nodes in the network. Nodes with high betweenness centrality act as bridges or intermediaries, connecting different parts of the network. The betweenness centrality of a node is calculated based on the number of shortest paths between pairs of nodes in the network that pass through that particular node. Fig. 5 depicts the procedure of betweenness centrality.

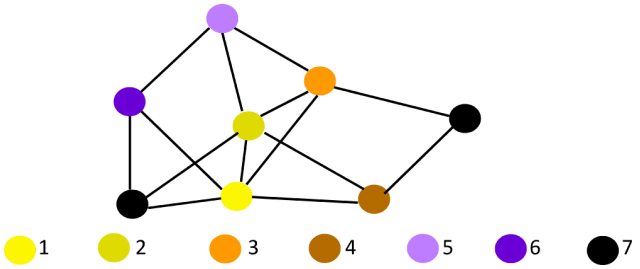

Fig. 5: The procedure of betweenness centrality. Each node's legend number indicates the centrality rank of the node

Malm et al. [56] analyzed co-offending in various criminal enterprise groups. They used data from multiple police systems and employed betweenness centrality to study the structure and composition. The study revealed distinct co-offending patterns across different crime groups. Décary-Hétu and Dupont [24] evaluated the effectiveness of SNA in enhancing information about cybercriminals and identifying subjects for further investigation. They demonstrated that SNA, including betweenness centrality and other measures, provides scientific and unbiased metrics for identifying key actors.

Fidalgo et al. [30] conducted a study to explore the identification of potentially significant nodes in fraud networks. Their approach involved assessing the relationship between control and influence by using betweenness centrality to measure bridging centrality. The bridging centrality metric of a node is derived from its betweenness centrality and bridging coefficient. Taha and Yoo [75] introduced a forensic analysis approach to identify influential criminals in a criminal network using edge betweenness centrality. Their method calculates the shortest-path edge betweenness for each edge. A Minimum Spanning Tree (MST) is constructed for the network based on these weights. Each node, denoted as u, is assigned a score representing the number of nodes in the MST that rely on u for their existence. Influential nodes are identified by ranking them according to their score. Calderoni and Superchi [17] conducted a study that investigated the characteristics of criminal leadership by examining the involvement of leaders in meeting and telephone communications. They also compared the meeting and wiretap networks to identify leaders. The findings revealed a significant correlation between high betweenness centrality and the likelihood of being a criminal leader. Taha and Yoo [76] proposed a forensic analysis method to detect influential criminals in a network. The method focuses on

critical communication pathways and involves calculating betweenness centralities of nodes to estimate their impact on information flow. The method also considers the betweenness centralities of nodes connected to the path, providing a comprehensive evaluation of path importance. Grassi et al. [39] studied betweenness centrality to identify criminal leaders in a meeting participation network. Despite expected correlations, different forms of betweenness centrality yielded distinct rankings for nodes. Dual projection methods were generally more effective than traditional approaches in identifying criminal leaders.

B. *Local Analysis Scope*

1. *Node Similarity-Based Model*

Node Similarity centrality is a metric that quantifies the similarity of a node to other nodes in the network based on their shared neighbors. It measures how well-connected a node is to other nodes that are also well-connected. In the context of a criminal social network, it can help identify individuals who have connections to other influential criminals. The method defines a similarity metric to measure the similarity between nodes in the network. This metric should capture the characteristics or attributes that make a criminal influential. For example, you can consider factors such as the type and severity of criminal activities and the number and importance of connections. There are several common node similarity metrics such as Jaccard Similarity, Adamic-Adar Similarity, Preferential Attachment, Cosine Similarity, Pearson Correlation Coefficient, and Euclidean Distance.

Berlusconi et al. [10] contended that intelligence and investigation activities could suffer adverse consequences as law enforcement agencies might overlook certain individuals and connections. They showcased how node similarity can detect potential missing links within criminal networks, even when the available information is inherently noisy or incomplete. Tundis et al. [78] utilized node similarity to examine the resemblances and connections between criminals involved in Organized Crime and Terrorist Networks. Their objective was to uncover both similarities and clusters of users associated with illegal activities, such as drugs, weapons, and human trafficking. Additionally, the approach aided in identifying group leaders and mediators within these networks.

Calderoni et al. [14] utilized community analysis techniques to investigate the arrangement of a criminal network, which depicted the extent of individuals' joint participation in meetings. In this pursuit, they employed a node similarity measure, operating under the assumption that nodes exhibiting higher similarity are more likely to share the same label. Kumari et al. [43] presented two intelligent techniques for deceiving community detection algorithms (CDAs) with the aim of concealing nodes within a network. They employed node-based matrices, persistence scores, and safeness scores to define optimization problems that would confuse the CDAs.

2. *Local Clustering Coefficient-Based Model*

The local clustering coefficient is a measure used in network analysis to quantify the level of clustering or interconnectedness of nodes in a network. In the context of a criminal social network, it can be used to identify influential criminals or individuals who have a strong influence within their local neighborhood. The local clustering coefficient of a node is a measure of how closely connected its neighbors are to each other. The local clustering coefficient, denoted as C_i, measures the extent of interconnectedness among the nodes within the neighborhood of a given node v_i. It is calculated by dividing the number of links present among the neighboring nodes by the maximum number of possible links that could exist among them. Several algorithms can be used to compute the local clustering coefficient, such as the Watts-Strogatz algorithm or the triangle counting algorithm. The procedure of local clustering coefficient is illustrated in Fig. 6.

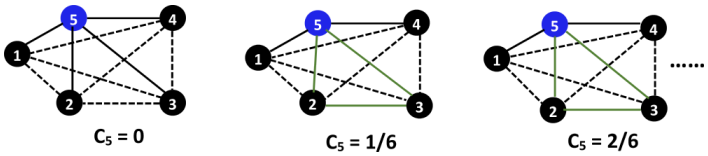

Fig. 6: The procedure of local clustering coefficient is illustrated in the figure.

Agreste et al. [2] examined the network structure of a Mafia organization, documenting its development over time and emphasizing its adaptability to interventions aimed at membership targeting. Furthermore, they highlighted its ability to withstand disruptions caused by police operations. The researchers proposed a two-stage approach, where the criminal network was initially divided into subgroups using a clustering algorithm. They then calculated the Average Clustering Coefficient for each vertex in relation to its degree and consistently found it to be greater than 0.6.

Ozgul and Erdem [57] introduced a measure of resilience for criminal networks, which they applied to two actual criminal networks. We examined the resilience outcomes in relation to various factors such as their activities, recruitment methods, network growth, survival strategies, and the level of secrecy maintained following legal prosecution. To assess the resilience, the authors utilized two measures: the Average centrality of leaders and the Clustering coefficient. Catanese et al. [20] introduced LogAnalysis, a forensic system designed to assist forensic investigators in comprehending the hierarchical structures within criminal organizations. This system enables the identification of central members within such organizations. LogAnalysis offers the capability to calculate both the global clustering coefficient of a given phone call network and the local clustering coefficient of nodes.

Song et al. [68] examined two distinctive features of small-world networks: the local clustering coefficient and the global characteristic path length. Their findings revealed that individuals within the fake review group exhibited lower friend counts and were more inclined to provide negative ratings, particularly with ratings of 1 or 2.

III. Clustering-Based Analysis

A. Spatial-Based Analysis

1. Global-Based Analysis

1) Hierarchical-Based Clustering

Hierarchical-based clustering can be used to identify influential criminals in a criminal social network. It allows us to group individuals based on their similarities and differences, enabling us to identify clusters of influential criminals within the network. Spatial global-based hierarchical clustering considers the geographical aspect of the criminal network, allowing us to identify spatially cohesive clusters of influential criminals. It creates a hierarchical structure of clusters by successively merging or splitting clusters at different levels. Fig. 7 illustrates the general procedure of the approach.

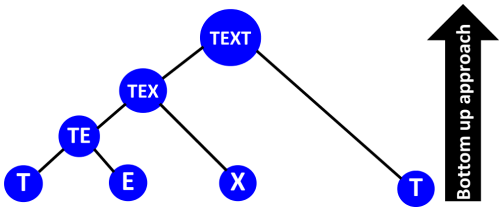

Fig. 7: The general procedure of the hierarchical-based clustering approach is illustrated in the figure

A forensic system called CrimeNet Explorer, developed by Xu and Chen [83], aids law enforcement agencies in uncovering criminal networks and their leaders. It utilizes techniques like the concept space approach, hierarchical clustering, social network analysis, and multidimensional scaling. Hierarchical clustering divides the network into subgroups based on relationship strength, facilitating the identification of significant criminals. Kazmi et al. [44] compared ten methods to detect roles and influential leaders in terrorist networks, finding hierarchical-based clustering to be the most effective, especially in real-time situations. Afra and Alhajj [3] used crime incident reports to construct a criminal graph, connecting criminals based on co-occurrence in the reports. They applied hierarchical clustering to uncover influential criminals, reveal hidden relationships, and identify hierarchical criminal groups.

2) Spatial Random Graph Distribution-Based Clustering

Spatial Random Graph Distribution-Based Clustering is a clustering algorithm that combines spatial information with network analysis to identify clusters or communities within a criminal social network. This approach can be used to detect influential criminals or key players within the network. One of the approaches for incorporating spatial information into the network is by assigning spatial coordinates to each node based on their associated locations. This step helps capture the geographical proximity or spatial relationships among individuals within the network. The algorithm will group nodes into clusters based on their spatial and network proximity using an algorithm, such as k-means or DBSCAN, to the network.

Agarwal and Toshniwa [4] identified influential local leaders in hazard networks, including criminal networks. Their algorithm assigned ranking scores to weakly connected neighbors, considering random spatial information. Factors like outer degree, inner degree neighbors, epsilon, and damping factors strongly influenced the ranking relationship. Ficara et al. [31] conducted simulations to disrupt Mafia networks, using centrality metrics (degree, betweenness, closeness) to identify influential criminals. The intervention strategy randomly eliminated one actor at each step, assessing network integrity with three measures. Duijn et al. [23] studied resilience in criminal networks when disrupted. They found that targeting influential leaders weakens the network. Three recovery mechanisms were used to simulate network resilience.

2. Local-Based Analysis

1) Spatial Diffusion-Based Clustering

Local Spatial Diffusion Clustering is a technique used for identifying influential criminals in a social criminal network based on spatial diffusion patterns. It leverages both the social relationships among individuals and their spatial proximity to detect clusters of criminal activity. It performs local spatial diffusion by propagating the cluster labels based on the density-based clustering results and the spatial constraints.

Calderoni et al. [15] proposed a methodology to detect and understand criminal organization structures. They used geometric space diffusion and temporal perspectives to analyze power dynamics and visualize structural changes over time, with implications for law enforcement. Meneghini et al. [52] introduced a methodological approach for estimating a criminal trafficking network across different local spatial geographical levels. This methodology focuses on identifying the most probable routes within the criminal network that can be targeted and exploited by criminals. Taha and Yoo [77] introduced CLDRI, a forensic analysis system to identify key individuals in a criminal organization. CLDRI quantifies their influence by considering factors like local spatial relationships and information diffusion among connected nodes.

2) Density-Based Clustering

Local Density-Based Clustering (LDBC) is a methodology proposed for identifying influential criminals within a criminal social network. LDBC utilizes local density information to cluster individuals and identify those with significant influence in the network. By analyzing the density of connections surrounding each node, LDBC can identify clusters of influential criminals in the criminal social network. It starts by selecting an arbitrary core point and expands it by finding all density-reachable points from that core point. This process continues recursively, connecting core points and expanding the clusters until no more density-reachable points can be found. Clusters that contain a high concentration of influential criminals are likely to represent groups of influential criminals.

Everton et al. [26] analyzed the evolution of the Noordin Top terrorist network and its changing structure over time. They investigated how the network's goals and strategies employed by authorities affected its interconnectedness and local spatial density. The authors observed that some networks tend to become more internally dense and centralized as time goes on

Gunnell et al. [38] proposed a method to identify individuals associated with gangs, comprehend gang activities, and gain insights into gang structure and organization. They found that the "gang links" sub-network had the highest density, indicating higher levels of involvement compared to other networks.

Ficara et al. [31] analyzed the Montagna multiplex network using real criminal data from the Montagna anti-mafia operation. They examined the network's layers separately and measured various metrics such as density, connected components count, size of the largest connected component, and average clustering coefficient. Additionally, they investigated the characteristics of the top 20 influential actors, including degree, degree deviation, neighborhood, exclusive neighborhood, relevance, and exclusive relevance.

David et al. [15] explored the changes in both structure and function of a criminal network over time. They observed that the network's local density remained stable throughout the studied period but became more decentralized towards the end. The centrality scores of individual nodes indicated that influential nodes varied over time, reflecting the evolving priorities and objectives of the network.

3) Fuzzy C-Means Clustering

The Local Fuzzy C-Means Clustering approach can be utilized to identify influential criminals within a criminal network. This method employs a fuzzy clustering algorithm that considers the local characteristics of the network to determine the influential individuals. By considering the connectivity and relationships among criminals, the algorithm assigns membership degrees to each individual, indicating their level of influence within the network. The higher the membership degree, the more influential the criminal is. Each data point can belong to multiple clusters with varying degrees of membership.

Premasundari and Yamini [59] utilized a local Fuzzy C-Means approach to perform clustering on crime rates. They developed a novel multiple clustering model and assessed its effectiveness using the USArrests dataset. The outcomes were then employed to forecast the likelihood of crime occurrence by visually analyzing the crime patterns across different states in the United States. Sivanagaleela and Rajesh [70] introduced a fuzzy C-Means algorithm, which was suggested as a suitable method for clustering crime data related to various cognizable crimes like Kidnapping, Murder, Theft, Burglary, and Robbery. By employing the fuzzy clustering technique, this algorithm effectively identifies areas with higher crime rates. conducted a study on factor clustering analysis concerning violent crimes. They utilized a Fuzzy c-means algorithm that incorporated information entropy to address the issue of overlapping data.

B. Spatiotemporal-Based Analysis

1. Spatiotemporal Diffusion-Based Analysis

Spatiotemporal diffusion-based clustering is a method that combines spatial and temporal information to identify clusters or communities within a network. In the context of identifying influential criminals in a social crime network, this approach can be useful for understanding the propagation patterns of criminal activities and the influence exerted by certain individuals. Common clustering algorithms, such as k-means, DBSCAN, or spectral clustering, can be employed to partition the data points into clusters based on their diffusion characteristics.

Zhao and Tang et al. [87] provided a comprehensive overview of urban crime. Their study examined environmental and social criminal theories and utilized analysis techniques to gain insights from geospatial and temporal crime data. They also identified influential criminal leaders by studying criminal spatiotemporal diffusion patterns. Park and Tsang [60] proposed a framework to identify and visualize influential individuals in co-offending networks. They considered the temporal aspect by using centrality measures to detect key players in criminal networks. Their framework captures temporal changes and provides insights into network dynamics. Siriwat and Nijman [71] studied crime rates in Thailand to uncover spatial and temporal patterns. Their analysis aimed to identify and analyze patterns related to the origin, transit, and destination of crimes, as well as temporal patterns over time.

2. Spatiotemporal Random Graph-Based Analysis

The spatiotemporal random graph-based models consider the spatial and temporal aspects of criminal activities and analyze the connections between individuals involved in criminal behavior. They identify clusters of criminal activity that are geographically and temporally connected. This analysis can provide insights into how criminal activities propagate and evolve over space and time. In the context of identifying influential criminals in a social crime network, this approach can be valuable in understanding the dynamics of criminal activities and the roles played by different individuals.

Berlusconi [6] introduced the spatiotemporal Random Graph Distribution to analyze temporal changes in structure. It combines qualitative analysis of wiretapped conversations with a quantitative element, using network statistics and exponential random graph models. This helps identify influential criminals. Win et al. [81] proposed an algorithm using fuzzy clustering and random initialization to identify potential criminal patterns in extensive spatiotemporal datasets. It detects patterns from large-scale datasets that encompass criminal activities. The study assessed crime rates for different locations. Griffiths et al. [40] studied the spatial and temporal behavior of UK-based Islamist terrorists. They analyzed the frequency and timing of their visits to different locations, aiming to assess if their movement patterns differed from the ordinary criminals.

IV. Agent-Based Analysis

A. Eigen Analysis-Based

1. PageRank-Based Analysis

PageRank is a graph algorithm that can be applied to identify influential criminals in a social crime network. It assigns importance scores to nodes in a graph based on the structure of the network. The basic idea behind PageRank is that a node is considered more important if it is connected to other important nodes. In the context of a crime network, this can be interpreted as individuals who are connected to other influential criminals being more likely to be influential themselves. Fig. 8 illustrates the general procedure of PageRank.

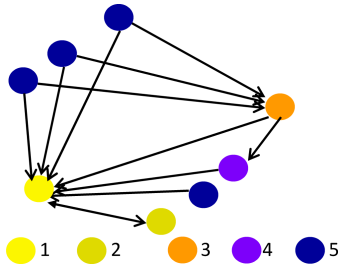

Fig. 8: Illustration of the procedure of PageRank. Each node's legend number indicates the centrality rank of the node

Isah et al. [35] conducted a study to understand connections and community patterns in crime data, including conventional and cyber crimes, and predicting organized criminal networks. They used PageRank to analyze networks, identify influential nodes, detect subgroups, and assess network interconnectedness. Budur et al. [8] used probabilities generated by their model as weights for current edges to identify influential nodes. They calculated weighted PageRank scores from the weighted network and computed classical PageRank scores based on unweighted edges. Calderoni et al. [14] studied link prediction's effectiveness in different relationships within a social group, particularly in a mafia organization. They assessed various algorithms, including the PageRank score.

2. Eigenvector-Based Analysis

Eigenvector-based analysis is a powerful technique that can be used to identify influential individuals within a social network, including criminal networks. In the context of a criminal social network, influential criminals can be identified based on their centrality and their ability to control or influence the flow of information or resources within the network. It computes the eigenvector centrality for each node in the network. Eigenvector centrality quantifies the influence of a node by considering both its direct connections and the influence of its connected nodes. It assigns higher centrality scores to nodes that are connected to other influential nodes. It calculates the centrality scores using the eigenvector associated with the largest eigenvalue of the adjacency matrix. This procedure is illustrated in Fig. 9.

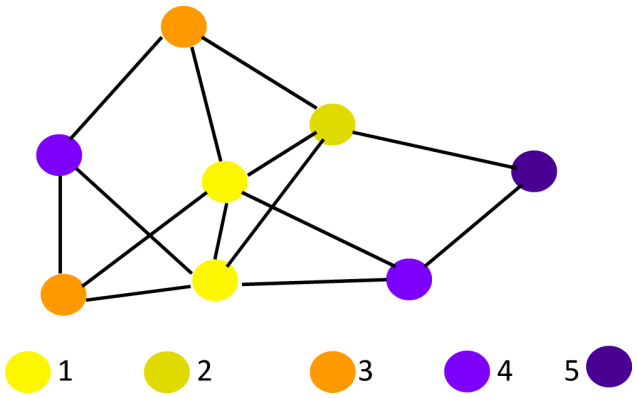

Fig. 9: Illustration of the procedure of Eigenvector. Each node's legend number indicates the centrality rank of the node

Ferrara et al. [32] introduced LogAnalysis, an expert system designed to enable semi-supervised identification of criminal communities within networks created from phone call records. This system assists forensic investigators in comprehending the hierarchical structures within criminal organizations, uncovering influential members who facilitate connections among sub-groups. Shang and Yuan [69] conducted an assessment on the efficacy of three distinct techniques to classify an unfamiliar network into three categories: terrorist, cocaine-related, or noncriminal. The computation of eigenvector centrality for the entire network followed a similar approach as that of closeness and betweenness centralities. They also calculated the average eigenvector centrality by taking the mean value across all network nodes.

Calderoni et al. [14] presented a forensic system that examines the roles of individuals within a criminal organization's hierarchical structure and predicts crimes by analyzing spatiotemporal patterns of criminal activities. It operates on the principle that if a node has numerous central neighbors, it should also be considered central. It utilizes eigenvector centrality to determine the significance of a node by considering the importance of its neighbors.

B. Role Centrality-Based Analysis

1. Behavior-Based Analysis

Behavior-Based Analysis is an approach used to identify influential criminals within a social criminal network. By analyzing the behavior patterns of individuals within the network, this method aims to determine the level of influence or importance a criminal holds within the network. Rather than relying solely on traditional measures such as hierarchical positions or organizational roles, Behavior-Based Analysis focuses on observing and analyzing the actual behaviors and actions of individuals. This can include studying their involvement in criminal activities, their interactions with other criminals, their decision-making processes, and their ability to persuade or influence others. This analysis identifies impactful criminals in the social criminal network by examining behavioral aspects.

Rodrigueza and Estuar [65] conducted a study on human behavior in disasters. They created models of perceived behavior using networks such as Agent x Agent, Agent x Knowledge, Agent x Task, and Agent x Belief. These models were analyzed across the three phases of a disaster. Additionally, they employed SNA to identify influential agents within a simulated disaster behavior network. Hutchins and Benham-Hutchins [34] conducted a study where they investigated how intelligence analysts, along with network analysis software and methodologies, utilized a combination of measures to analyze the behavior of criminal organizations. The researchers presented data from three networks to demonstrate the findings obtained through organizational risk analysis, both quantitatively and qualitatively.

Easton and Karaivanov [27] explore the network structures that naturally emerge as a consequence of the interplay between a deterrence policy and the responses of networked agents as they adapt the crime network itself. The objective of the study was to gain insight into criminal behavior. Within any network, the "key player" policy identifies the individual agent whose elimination would lead to the largest decrease in total crime.

Wang et al. [82] investigated user search behavior and internet information foraging by analyzing user search sessions. The study utilized a set of search logs from a large search engine. User sessions were identified using hierarchical agglomerative clustering. Based on information foraging theory, the researchers proposed a model that predicts the probability distribution of the number of queries and clicks in a search session. The model assumes that users make sequential decisions, continuing the search as long as the expected value of continuing exceeds a threshold. A machine learning-driven tool for detecting and evaluating cyber threats was developed by Wang et al.[80]. The tool employs a two-stage analysis approach, incorporating both unsupervised and supervised learning techniques, and operates on a dataset of 822,226 log entries obtained from an AWS cloud-based web server. By leveraging unsupervised learning, the tool can uncover patterns, anomalies, and potential threats.

2. Degree-Based Analysis

Degree-based analysis is an approach that can be used to identify influential individuals within a social criminal network. It focuses on examining the degrees of connectivity or centrality of nodes (individuals) within the network. In this analysis, the degree of a node refers to the number of connections or links it has with other nodes in the network. By calculating and comparing the degrees of all nodes in the network, one can identify individuals with high degrees or those who are highly connected. These individuals are considered influential within the criminal network because they have a greater potential to spread information, resources, or criminal activities. Moreover, targeting or eliminating these high-degree individuals may have a significant impact on the overall structure and dynamics of the criminal network. This process is depicted in Fig. 10.

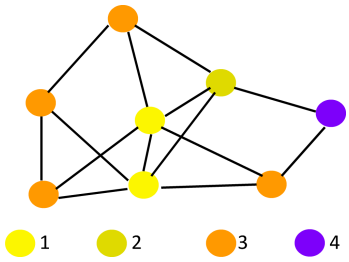

Fig. 10: Illustration of the Degree-based analysis procedure. Each node's legend number indicates the centrality rank of the node

Duijn and Klerks [23] explored the utilization of social network theory within the context of Dutch law enforcement. The paradigm of Intelligence-Led Policing has led to the accumulation of substantial information regarding habitual lawbreakers and criminal networks. To simulate various control strategies, the researchers applied five algorithms for network disruption to a criminal cannabis network. The disruption strategies aimed to target social capital through the use of degree and betweenness metrics, which were associated with attacks on the network. Bright et al. [11] examined the efficacy of five different law enforcement interventions in disrupting and dismantling criminal networks. The study focused on evaluating three measures at each step of the interventions, which included: (1) network degree centralization, (2) the number of active components within the network, and (3) the size of the largest active component. By analyzing them, the researchers assessed the impact and effectiveness of the law enforcement interventions on the targeted criminal networks.

Agarwal et al., [1] collected a dataset from Twitter, which included 3.2 million distinct users and over 12 million tweets. The data was categorized into three awareness groups in order to identify influential individuals within each category. The researchers analyzed various dynamic features of each user, such as their in-degree (number of incoming connections), out-degree (number of retweets), favorite count, and activity on social media platforms (SMPs). These features were used to rank the users across the entire dataset. Bright and Delaney, [7] investigated how a criminal network's structure and function evolved over time. They computed centrality measures, specifically degree and betweenness, for each member of the network at different time intervals.

Colladon and Remondi [19] conducted research on the significance of social network metrics and introduced novel network mapping techniques that are not typically used in anti-money laundering practices. They analyzed variables such as in-degree, out-degree, and their combined all-degree measures. The researchers discovered that there are typically strong correlations between the in-degree, out-degree, and all-degree variables, indicating a normal relationship. Petersen et al. [58] introduced a knowledge management strategy for visualizing potential secondary effects following the removal of a node, enabling investigators to explore "what if" scenarios in criminal network analysis. The authors proposed a node removal algorithm that utilizes various measures, including degree centrality, to identify important nodes in the network.

V. COMPARATIVE EVALUATIONS

In this section, we analyze the different techniques addressed in this paper. We assess each technique based on four criteria: the core idea behind the technique, the rationale behind its implementation, the necessary conditions for achieving its best performance, and its limitations. Table 1 illustrates the techniques based on ***topology***, while Table 2 showcases the techniques based on ***clustering***. The ***agent-based techniques*** are outlined in Table 3. Our aim is to provide a comprehensive understanding of the benefits and drawbacks of each technique, as well as its suitability for specific tasks.

Table 1: Evaluating each **topology**-based technique in terms of the following four criteria: its underlying principle, its justification, its conditions for optimal performance, and its limitations.

Tech.	Papers	Technique Essential Concept	Rationale Behind the Usage of the Technique	Conditions for the Optimal Performance of the Technique	Limitations of the Technique
Katz-Based Model	[16, 86, 13]	The network is represented as an adjacency matrix. The elements of the matrix indicate the presence or strength of connections between nodes. Then, the appropriate values for the parameters α and β in the Katz centrality equation are selected. Once the Katz centrality scores are computed for each node in the network, the nodes are ranked based on these scores. To identify influential criminals, a threshold value is chosen as a cutoff point. This threshold serves as a criterion for determining which nodes surpass a certain influence level	In a criminal network, individuals don't always interact directly with everyone else, but their influence can still spread through connections. Katz centrality considers both direct and indirect connections to measure influence in the network. Criminal networks often have hierarchies or chains of command, where top-ranking individuals control lower-ranking members. Katz centrality captures influence propagation by assigning higher scores to nodes connected to highly influential nodes. It considers multiple paths, network structure integration, and varying levels of influence.	The technique can be improved by: (1) including relevant nodes beyond the main criminals to analyze the influence within the network comprehensively, such as their associates, subordinates, and other individuals involved in criminal activities, (2) accurately representing the direction of relationships in the network to capture the flow of influence correctly, (3) adapting the analysis to align with the specific characteristics of the criminal context, (4) experimenting with different damping factor values to emphasize various levels of influence and accurately capture the dynamics of the network.	The limitations are: (1) Katz centrality considers both direct and indirect connections, but direct connections may be more important in determining influential criminals, (2) Katz centrality may struggle to identify influential criminals who operate discreetly or maintain hidden connections, leading to an underestimation of their true influence, (3) Katz centrality doesn't account for important contextual factors like criminal expertise, reputation, or hierarchical structures, which are crucial for accurately identifying influential criminals, (4) Katz's accuracy depends on the initial selection of influential individuals
Multiple Link Types Model	[9, 31, 74, 55]	It starts by characterizing different connections in the social criminal network. It assigns weights or strengths to these connections based on factors like intensity, frequency, or significance. The analysis occurs on a multilayer network, where each layer represents a unique connection type. Metrics/indices are then created to measure individuals' influence, considering the multiple connection types. The approach enables the examination of individuals' roles across different network connection types	Criminal activities involve complex relationships and interactions. Considering multiple types of connections provides a more comprehensive understanding of network dynamics and individual roles. In criminal networks, interactions like partnerships, hierarchies, resource sharing, and communication patterns are present. Each connection type represents a unique aspect of these interactions. Examining different link types reveals connections that may not be apparent when focusing on a single relationship type.	The technique can be enhanced by: (1) understanding and representing each link type's specific meanings and implications in the network data (differentiate between link types by using varied edge labels, weights, or attributes), (2) evaluating the importance and significance of each link type concerning its potential influence on the network, (3) assigning suitable weights or strengths to each link type based on their relevance and impact on influence within the network, (4) utilizing algorithms capable of effectively handling multiple link types	The limitations are: (1) the lack of standardized definitions and criteria for link types can result in subjective interpretations and categorizations, (2) analyzing multiple link types increases complexity, requiring larger and more diverse datasets, (3) emphasizing structural aspects of the network overlooks crucial contextual information, such as motivations and individual characteristics (understanding criminal behavior and influence requires considering these factors alongside network structure), (4) multiple link types contribute to complex network structures.
Closeness Centrality Model	[85, 53, 12, 66]	The network's closeness centrality is utilized to assess the proximity of each criminal, employing algorithms like Dijkstra's or Floyd-Warshall. Criminals with greater closeness centrality are deemed more influential in the network. Criminals with higher closeness centrality are also more likely to have access to critical information, resources, or connections within the network. They may serve as crucial intermediaries or decision-makers, facilitating the flow of illicit activities, coordinating criminal operations, or exerting control over other members.	Criminal networks rely on efficient information exchange. Individuals with high closeness centrality quickly connect with others, facilitating information dissemination. Identifying such criminals reveals key intermediaries controlling the flow. They effectively communicate and coordinate with many network members. Removing influential criminals with high closeness centrality has a profound impact on the structure and functionality of the network in which they operate. Closeness centrality is a measure that quantifies an individual's accessibility within a network, highlighting their ability to reach other individuals.	The technique can be enhanced by: (1) a connected network is crucial (fragmented or isolated nodes/groups hinder closeness centrality's accuracy in assessing overall criminal influence), (2) relying solely on closeness centrality may not sufficiently identify influential criminals, necessitating consideration of specific criminal activities and roles, (3) efficient communication channels among criminals are vital for using closeness centrality to identify influential individuals (barriers undermine the centrality's ability to capture influence), (4) depending on network's characteristics, an appropriate centrality threshold is needed	The limitations of closeness centrality are: (1) it overlooks the nature and significance of criminal behavior, failing to differentiate between influential criminals involved in high-impact activities and others, (2) it assumes equal and rapid information flow across the network, disregarding individuals with more control or unique skills that amplify their influence, (3) it neglects indirect influence transmitted through intermediaries, who play crucial roles in information flow and influence transmission, (3) it assumes a connected network, but social criminal networks can exhibit fragmentation or isolated subgroups.

Betweeness Centrality Model	**[56, 24, 30, 75, 17, 76, 39]**	The betweenness centrality of a node is established based on the count of shortest paths that traverse it. To systematically calculate this, algorithms like Brandes or the Girvan-Newman can be employed to compute the shortest paths between every pair of nodes and determine the proportion of these paths that pass through each node. After calculating the betweenness centrality values for each node, nodes with higher scores can be interpreted as having a more significant influence within the social network. This measure identifies nodes that serve as crucial connectors, acting as intermediaries between other nodes, and having control over the flow of information.	In criminal networks, interconnected subgroups or communities are common. Betweenness centrality helps identify individuals who connect or mediate between these groups. In such networks, effective information dissemination is crucial for operational success. Individuals with high betweenness centrality regulate information flow and influence decision-making. Their removal or arrest can fragment the network. High betweenness centrality is attributed to individuals who appear frequently on the shortest paths within the criminal network, indicating their crucial roles within it. This metric holds relevance in criminal investigations, empowering analysts to prioritize their attention to individuals who exert influence in the network.	The technique can be enhanced by: (1) Networks should have significant interconnectivity among members (fragmented networks hinder accurate betweenness centrality measurement), (2) reliable network information is crucial (incomplete data compromises betweenness centrality accuracy), (3), analyzing within a specific timeframe and aligning data for betweenness centrality is important (criminal networks evolve over time), (4) setting appropriate thresholds based on network characteristics and objectives is crucial (improper thresholds affect measurement effectiveness), and (5) combining betweenness centrality with qualitative analysis and expertise enhances identifying influential criminals.	The limitations of Betweenness Centrality are: (1) criminal networks use hidden or indirect communication channels, making it difficult for it to measure criminal influence accurately, (2) it favors intermediaries and network bridges, but influential criminals in tightly-knit groups or specific subgroups may be overlooked, (3) it assumes a single path of information flow, while criminal networks have multiple paths and redundant connections that can bypass high-betweenness individuals, reducing their influence, (4) it treats all interactions equally, disregarding the diverse activities, dynamics, hierarchies, and varying roles within criminals, (5) it may overlook important information about node influence not captured by the shortest path criterion.
Node Similarity-Based Model	**[14, 43, 10, 78]**	The connections between nodes determine their structural similarity. Network centrality measures can assess the significance of each node based on its position in the network. High centrality scores indicate influential nodes with strong connections. The characteristics of individuals in the social criminal network, including demographics, criminal history, and affiliations, are examined. Similarity metrics like cosine or Jaccard similarity are used to measure attribute profile similarity between nodes. Typically, both structural and attribute similarity measures are combined for a comprehensive understanding. The combination of these measures allows for a more comprehensive analysis.	Structural equivalence suggests that nodes with similar connections have similar roles in the network. Individuals who are structurally equivalent to influential criminals may hold influential positions in the criminal network. This helps to identify nodes resembling influential criminals based on their connections. By systematically examining connections and relationships, we can identify individuals who exhibit similar behavioral patterns and characteristics to those of known influential criminals. Within the network of connections that we evaluate, certain nodes emerge as potential influential criminals. These individuals exhibit a level of influence that is comparable to known influential criminals within the criminal activities.	The criminal network representation should accurately capture interactions and relationships while reflecting its specific context and dynamics. Choosing an appropriate similarity metric is vital, as different metrics yield different results based on the network structure and criminal connections. A threshold may be needed to determine significant connections, enhancing accuracy in identifying influential criminals. Comparing node similarity centrality scores against relevant baselines is crucial. Validating results with additional information and considering temporal dependencies can enhance our understanding of network influence, leading to a more precise analysis. We can uncover the underlying mechanisms hat drive network influence over time.	The limitations Node similarity centrality are: (1) analyzing criminal networks requires considering influence flows and identifying who gives/receives orders (it overlooks this, limiting its effectiveness), (2) criminal networks involve different illegal activities (it lacks information on activity nature and severity, crucial for identifying influential criminals), (3) it focuses on network structure, disregarding contextual factors (e.g., socio-economic) impacting criminal network dynamics, and (4) criminal networks hide covert connections not visible in the network data (it relies on observable structure, potentially missing hidden relationships), (5) Node similarity-based models typically do not account for temporal dependencies and the dynamic nature of networks.
Local Clustering Coefficient Model	**[57, 20, 68, 2]**	The method calculates the local clustering coefficient for each person in the criminal network, analyzing their relationships and gauging interconnectedness. It identifies criminals with high local clustering coefficients, indicating strong connections among their neighbors and potential clusters. By examining these clusters, it investigates the interactions, activities, and roles of criminals, aiming to understand their importance in the network. The examination of the clusters provides investigators with a comprehensive understanding of the inner workings of criminal networks, allowing them to identify key targets, dismantle the infrastructure, and neutralize the influential individuals	The local clustering coefficient identifies criminal clusters with strong interconnections. These clusters foster cohesive relationships and collaboration among criminals, providing insights into influential individuals. High clustering coefficients indicate close connections between specific criminals, suggesting collaboration in criminal activities. Also, the coefficient highlights bridge nodes that connect different clusters, playing crucial roles in linking isolated parts of the network. These bridge nodes act as critical intermediaries, facilitating the exchange of illicit goods, services, or knowledge across different criminal groups or activities. They possess a unique position within the network.	To enhance the technique: (1) identify nodes with high local clustering coefficients, indicating potential influence due to their highly interconnected neighbors, (2) consider nodes with high degree centrality, suggesting influential criminals with extensive connections, (3) pay attention to nodes connecting multiple high-degree nodes, acting as bridges between distinct clusters or groups, enabling influence across the criminal network, and (4) analyze the clustering patterns of a node's neighbors to identify tightly knit criminal subnetworks where the node's presence signifies influence within its immediate vicinity, (5) the network should have a more uniform distribution of connections across nodes.	The limitations of the technique are: (1) the global network structure and broader impact are not considered, even if influential criminals have connections to different non-clustered parts of the network, (2) it focuses on network structure and overlooks important attributes like criminal activities, reputation, and leadership skills that determine influence within a criminal network. (3) it ignores directional influence or control in the network, disregarding individuals who control information or resources, (4) it is a static measure that overlooks the dynamic evolution of the network over time, (5) The clustering coefficient measures the presence of triangles, but it does not provide information about longer cycles.

Table 2: Evaluating each **clustering**-based technique in terms of the following four criteria: its underlying principle, its justification, its conditions for optimal performance, and its limitations.

Tech.	Papers	Technique Essential Concept	Rationale Behind the Usage of the Technique	Conditions for the Optimal Performance of the Technique	Limitations of the Technique
Hierarchical-Based Clustering	[83, 44, 3]	It assesses the likeness between nodes. It employs hierarchical clustering via agglomerative or divisive approaches. In agglomerative clustering, nodes start as separate clusters and merge similar clusters based on a linkage criterion until a single cluster represents the whole network. In divisive clustering, the network begins as one cluster and gradually splits the most dissimilar cluster (e.g., using dissimilarity thresholds) into two until individual nodes are reached. Nodes' hierarchy positions reveal their influence.	Hierarchical clustering identifies patterns in a criminal network by grouping similar individuals, revealing clusters that represent criminal organizations, factions, or hierarchies. This method captures hierarchical structures where influential individuals hold higher positions. Organizing clusters into a tree-like structure helps identify key players at different influence levels. Through the analysis of connections between clusters and the identification of individuals bridging different parts of the network, hierarchical clustering aids in pinpointing central figures.	The technique is effective when the network has moderate to high connection density, but sparse networks present challenges in identifying influential criminals due to limited information. Choosing an appropriate similarity or distance measure (e.g., Euclidean distance or Jaccard index) is crucial. Linkage methods (e.g., single linkage) yield varying results, so selecting the right method aligns with the network's characteristics is crucial. Defining the optimal number of clusters requires a clear criterion (e.g., using a validation measure such as silhouette coefficient)	Hierarchical clustering has limitations in dealing with extensive datasets due to computational demands. It produces a tree-like structure that can be challenging to interpret. Determining the level of influential criminals is difficult, hindering actionable insights. Parameter choices greatly affect clustering outcomes and require domain expertise. The temporal dimension of criminal networks is not considered, and inappropriate distance metrics lead to suboptimal identification of influential criminals. Non-structural aspects are overlooked.
Spatial Random Graph Distribution	[31, 4, 23]	The model considers network connections and spatial distances to represent the network structure. It uses distribution-based clustering to identify clusters of influential criminals by incorporating spatial and network information. It assumes influential criminals are part of tightly connected groups with strong spatial clustering. Metrics like cluster size, density, and spatial extent are used to understand the spatial distribution and network structure of influential criminals.	Criminal activities often show spatial patterns, with criminals operating in specific regions. This approach considers spatial relationships to detect clusters of criminals in proximity, indicating hotspots. Criminal networks have complex structures with various relationships like co-offending and communication. This method analyzes connections between individuals to identify tightly connected subgroups, revealing cohesive criminal groups. Influential criminals exhibit distinct behaviors, and this approach can detect anomalies within the network.	To enhance the method: (1) incorporate spatial attributes into the criminal social network to capture accurate spatial patterns and connections, (2) collect sample data to ensure meaningful results, (3) account for spatial relationships of criminals accurately, as the method assumes a decrease in interaction probability with increasing spatial distance, (4) carefully select parameter values (e.g., spatial decay function) based on the network characteristics, (5) select appropriate statistical methods (e.g., maximum likelihood estimation)	Limitations include: (1) difficulty in identifying influential criminals in localized clusters due to assuming a random spatial distribution, (2) lack of incorporation of temporal information, ignoring the evolving nature of criminal activities, (3) disregarding hierarchies and types of connections in networks, (4) neglect of individual behavioral characteristics, such as behavior, skills, and reputation, which can affect a criminal's influence, and (5) oversight of contextual factors with significant impacts on criminal networks.
Spatial Diffusion-Based Clustering	[15, 52, 77]	The method examines how criminal behavior spreads in a network by simulating its dissemination among connected individuals. Factors like criminal past associations are considered during the diffusion process. Metrics such as diffusion metrics quantify influence, and clustering methods group individuals based on their influential attributes. Spatial clustering accounts for influence and spatial proximity, identifying influential criminals in clusters	Spatial diffusion clustering focuses on local interactions/connections, considering geographical or network proximity. This helps understand how criminal behavior disseminates, especially in networks where physical location and proximity are crucial. By analyzing diffusion patterns and network centrality, influential individuals can be identified. This helps identify "hotspots" of criminal activities, offering insights into the Key criminal hubs can be detected and the direction of criminal flow can be discerned.	To enhance the method: (1) ensure the network has clear community structure, with densely connected criminals sharing similar characteristics or roles within communities and sparser connections between communities, (2) employ various models (e.g., threshold model, cascade model) based on the characteristics of the network, (3) select seed nodes that cover different communities, (4) carefully tune parameters, such as the propagation threshold and activation probability (likelihood of adopting influence).	(1) reliance on the assumption of spatial proximity for influential criminals, which may not always hold true for those operating at higher levels or with decentralized networks, (2) lack of consideration for factors like social connections and organizational hierarchy, resulting in the oversight of influential criminals without strong spatial relationships, (3) don't consider temporal changes in influence within a network due to arrests or power struggles, (4) overestimation or underestimation of criminals due to inappropriate parameter selection.
Density-Based Clustering	[38, 26, 31, 15]	It uses density reachability to determine if a data point belongs to a cluster based on proximity to other points. It considers two parameters: epsilon (ε), the radius defining the point's neighborhood, and minPts, the minimum number of points for a dense region. A data point is density-reachable from another if it falls within the epsilon radius and the other point is a core point. It establishes connectivity between core points and forms the basis for cluster expansion. Data is considered outlier if unreachable	Density-based clustering, like DBSCAN, is adept at handling non-linear and irregular cluster shapes, enabling the detection of intricate patterns in networks. This makes it suitable for identifying influential criminals in cohesive groups. It can handle noise and outliers by treating them as separate, allowing focus on dense clusters likely to contain influential criminals. It enables analyzing large criminal social networks, making it practical to identify influential criminals in real-world scenarios	To enhance the method: (1) carefully select parameters for the clustering algorithm, (2) select the distance metric based on the data characteristics (e.g., using Jaccard similarity if the criminal network is represented as a graph, and Euclidean distance if the data is in vector form), (3) decide whether to include outliers in the clustering results or treat them separately (4) incorporate domain knowledge and expertise of the criminal social network to significantly enhance the clustering performance	The method has limitations in recognizing influential criminals in sparsely populated areas with low population density, possibly resulting in their misclassification as noise or outliers and their exclusion from analysis. Computational cost is high in large networks. The sequencing of data points during processing can affect the clustering outcomes, making it less reliable in identifying the same influential criminals across multiple runs. As the data's dimensionality increases, the method's effectiveness decreases

Fuzzy C-Means Clustering (FCM)	**[59, 70]**	FCM assigns membership degrees to criminals, indicating their association with clusters and influence within them. Higher degrees indicate greater influence. Cluster centroids, calculated using weighted averages of data points and their membership degrees, represent the central points of clusters and their defining attributes. FCM iteratively adjusts membership degrees and centroids to find the optimal solution. The fuzziness parameter (m) in FCM controls assignment ambiguity, with higher values enabling flexible assignments across multiple clusters. FCM is particularly useful when data points exhibit overlapping or uncertain boundaries between clusters.	Criminals can have different impacts based on connections, criminal activities, and social reputation. FCM clustering assigns membership degrees to capture this variability in influence. In a criminal network, some criminals can have varying degrees of influence in multiple clusters. FCM allows criminals to belong to multiple clusters, reflecting their influence in different criminal activities. This helps identify influential criminals spanning across various groups, crucial for understanding their overall impact in a network. FCM is robust against noise and outliers as it considers membership degrees rather than relying solely on distance measures. FCM provides a valuable approach for clustering tasks that involve uncertainty and overlapping clusters	The method can be enhanced by: (1) select an appropriate membership function to match data characteristics, with options like Gaussian, exponential, and sigmoid functions, (2) choose a reasonable number of clusters based on prior knowledge or techniques like elbow method or silhouette analysis, (3) select network features that capture factors like connections, propagation, and social attributes play a role, (4) set convergence criteria (e.g., maximum iterations, membership value threshold) carefully (too few iterations yield suboptimal results, while excessive iterations increase computation time without significant clustering improvement), and (5) cluster validity indices to help assess clustering quality.	The method has limitations: (1) subjective membership assignment and initial algorithm parameters can affect the accuracy and reliability of identifying influential criminals, (2) FCM may assign outliers to clusters, potentially mixing influential criminals with other criminals and impacting identification accuracy, (3) incorrectly choosing the number of clusters can lead to inaccurate identification or grouping of influential individuals, (4) FCM is a static clustering algorithm that doesn't account for the dynamic nature of criminal social networks, (5) interpreting fuzzy clustering results, especially for identifying influential criminals, can be challenging, (6) FCM may struggle to handle datasets with clusters of significantly unequal sizes.
Spatiotemporal Diffusion Analysis	**[60, 71, 87]**	A diffusion model explains how criminal activities spread in a network (e.g., Hawkes) can be used based on network characteristics. Network centrality measures (e.g., degree, betweenness) quantify an individual's influence. Analyzing spatiotemporal patterns reveals how criminal behavior spreads across locations and time. It involves studying diffusion patterns, identifying hotspots, and temporal trends. Individuals with high influence scores across multiple locations and time periods are influential criminals in the network. The method often incorporates spatial and temporal data, such as geographic coordinates, timestamps, and attributes of the diffusion process, to investigate the interactions and interdependencies between spatial and temporal dimensions	Spatiotemporal diffusion analysis can unveil patterns in the spread of criminal activity, offering valuable insights into the structure, organization, and dynamics of criminal networks. By studying historical data on criminal behavior, this analysis can help identify potential future hotspots or areas where criminal activities are likely to emerge or increase. Criminal networks are complex systems with intricate relationships and dynamics. Understanding spatiotemporal diffusion enables us to grasp how criminal behaviors propagate through these networks, including their speed and direction. This comprehension is crucial for developing effective strategies to disrupt or dismantle the network. Overall, Spatiotemporal Diffusion Analysis is a valuable tool for understanding, predicting, and managing the spread of phenomena across space and time.	Enhancements can be achieved by: (1) ensuring accurate network representation for identifying influential criminals, (2) refining spatial and temporal resolution to capture the geographical and temporal context of criminal activities, (3) selecting an appropriate propagation model (e.g., epidemic models, influence maximization models, or information cascades models) to accurately simulate diffusion within the criminal network, and (4) employing techniques like network centrality (e.g., degree centrality) and diffusion-based measures (e.g., influence scores, cascading probabilities) to quantify individual influence in the network. It is important to validate and evaluate the results of Spatiotemporal Diffusion Analysis. Comparison with ground truth data, external validation measures, or sensitivity analyses can help assess the accuracy.	The methods have limitations in recognizing important connections and assessing individual impact due to network representation constraints. Criminal networks being decentralized, fragmented, and constantly evolving makes it difficult to establish clear boundaries and capture relevant interactions. The presence of leaders, influential figures, or external factors can bias diffusion patterns within the network. Hidden connections and activities may not be accounted for, limiting the identification of influential criminals operating discreetly. The methods may overlook contextual factors influencing criminal networks despite focusing on spatiotemporal diffusion. The selection of analytical techniques, parameters, or models in Spatiotemporal Diffusion Analysis involves subjective decisions.
Spatiotemporal Random Graph Analysis	**[6, 81, 40]**	The approach utilizes a mathematical depiction of the social network of criminals, encompassing both space and time. It comprises nodes to represent individuals and edges to signify their relationships or interactions, accompanied by corresponding spatiotemporal characteristics. The method utilizes various analytical techniques to study the spatiotemporal criminal network. These techniques may include network centrality measures (e.g., degree centrality, betweenness centrality), community detection algorithms, diffusion models, or statistical methods for spatiotemporal data analysis.	The method enables us to comprehensively comprehend the structure of criminal networks by capturing both the spatial proximity and temporal dynamics of criminal interactions. By employing spatiotemporal random graph analysis, we gain insights into the evolution of the criminal social network, enabling us to detect emerging patterns like subgroup formation, the emergence of influential individuals, and changes in criminal information flow. Also, spatiotemporal random graph analysis can be utilized to develop predictive models for criminal behavior, enhancing proactive law enforcement measures.	Spatial and temporal resolutions must capture criminal interactions and network dynamics accurately. Enhanced resolutions aid in precise analysis, identifying patterns, subgroups, and influential individuals. Real-world data has uncertainties and missing information. Robust statistical methods, network algorithms, and imputation techniques address these issues and ensure reliable results. Metrics for identifying influential criminals should align with network objectives and characteristics, capturing various aspects of influence and power dynamics. Comparison with ground truth data, external validation measures, or sensitivity analyses can help assess the accuracy.	The method's limitations are: (1) Inadequate representation of complex criminal networks as spatiotemporal graphs may oversimplify and overlook crucial connections. (2) Assumptions about criminal behavior may not hold true in all scenarios, impacting accuracy. Selecting appropriate model parameters is challenging and subjective. (3) The method primarily focuses on structural properties of networks, neglecting factors like social skills and manipulation abilities. (4) Qualitative understanding is essential to grasp criminals' motivations and dynamics, beyond network analysis. (5) Effectiveness may be limited in rapidly changing criminal environments.

Table 3: Evaluating each **agent**-based technique in terms of the following four criteria: its underlying principle, its justification, its conditions for optimal performance, and its limitations.

Tech.	Papers	Technique Essential Concept	Rationale Behind the Usage of the Technique	Conditions for the Optimal Performance of the Technique	Limitations of the Technique
PageRank-Based Analysis	[35, 8, 14]	Initially, individuals have equal importance scores, representing a uniform distribution of influence with a sum of 1. Importance scores are iteratively recalculated based on connections and their importance. During each iteration, an individual's importance is shared with neighbors, considering the individual's own importance and connection strength. This process continues until importance scores reach a stable state, identifying the network's influential criminals.	The method considers the network structure to capture influence flow, identifying individuals with strong connections to other influential individuals. These well-connected individuals receive higher influence scores, suggesting that being linked to influential individuals enhances their own influence. This is crucial for identifying key players. PageRank provides an objective and quantitative measure of influence, enabling prioritization and ranking of individuals within the network. It is scalable and adept at handling large networks.	To improve the method: (1) assign appropriate weights to network edges to represent interactions between individuals and enhance analysis accuracy, (2) select the damping factor carefully based on the network's characteristics to ensure meaningful results, (3) set appropriate convergence threshold to balance computational efficiency and accuracy, considering the trade-off between smaller thresholds for higher precision, (4) implement strategies to handle sink nodes for unbiased results, (5) assign weights to nodes based on attributes like criminal records or affiliations	Limitations are: (1) PageRank ignores temporal dynamics, failing to consider the evolution of criminal networks over time, (2) incomplete or missing criminal connections can lead to biased or incomplete analysis, not accurately reflecting the network dynamics, (3) PageRank treats all nodes equally, disregarding unique characteristics or roles of individuals in a criminal network, overlooking influential criminals with important qualities beyond connectivity, (4) PageRank provides influence scores but lacks detailed explanations for their influence, (5) It is computationally costly.
Eigenvector-Based Analysis	[69, 32, 14]	To capture the relationships among criminals, an adjacency matrix is created. The values within the matrix indicate the strength of connections between the criminals. Each node is assigned a centrality score, which is determined by considering the significance of its neighboring nodes. The eigenvalues in the matrix indicate the influence of the eigenvectors. Large eigenvalues, corresponds influential nodes.	Influential criminals may have affiliations with other influential individuals. By considering the quality and importance of these connections, eigenvector analysis can identify individuals who have significant influence. Eigenvector centrality encompasses the concept of influence diffusion, wherein the influence of a criminal is propagated through the network if they are connected to highly influential individuals. The method can uncover the hierarchical structure of influence	To enhance the method: (1) consider directional relationships in the network. Incorporate directionality using algorithms like HITS or personalized PageRank to improve accuracy, (2) use efficient algorithms (e.g., power iteration or PageRank) to expedite the computation of eigenvector centrality for each node, (3) normalize the network size to ensure fair comparisons when comparing influences, (4) perform sensitivity analysis to evaluate the reliability of the method.	The method assumes a fixed network structure, not capturing the dynamic nature of criminal relationships. This reduces the effectiveness of real-time identification of influential criminals as new individuals enter or leave the network. Important factors like situational dynamics impacting criminal behavior may be overlooked, leading to an incomplete understanding of individual influence within the network. The method assumes a uniform network, whereas criminal networks vary in influence and interactions.
Behavior-Based Analysis	[65, 34, 27, 82, 80]	It involves defining and calculating influence metrics (e.g., node importance metric) to measure the impact and importance of individuals within the criminal social network. It analyzes the behavioral patterns of individuals (e.g., recurring behaviors and modus operandi). It also focuses on identifying anomalies or deviations from typical behaviors within the network. Behavior-based analysis can be used to develop predictive models that forecast future criminal activities or identify influential criminals.	Analyzing behavior can offer valuable insights into the future conduct of individuals involved in a criminal network. Through the examination of their previous behaviors, connections, and engagements, analysts can forecast potential future criminal acts and identify emerging patterns that influential criminals might employ. Criminal networks frequently exhibit hierarchical structures. Behavior-based analysis aids in deciphering these hierarchical dynamics by observing individuals' interactions and the roles they undertake in the network. Proactive methodology empowers investigators to take preventive actions .	To improve the method: (1) access to advanced data analytics tools is crucial for handling complex data in behavior analysis (these tools should process diverse data types, detect patterns, and generate insights), (2) combine expertise in criminology, data analysis, and social network analysis enables pattern identification and assessment of individual influence, (3) consider geography criminal activities to provide insights into behavior patterns and network dynamics, (4) act promptly in behavior analysis to allow for better identification and response to emerging patterns.	(1) Often, it lacks contextual information necessary for understanding the underlying motivations and relationships between individuals, (2) Some influential criminals may operate covertly, making it difficult to gauge their true level of influence through behavioral analysis alone, (3) Criminal networks constantly adapt and evolve over time and this method may struggle to keep pace with these changes as it relies on historical data, (4) potential for false positives and false negatives in identifying influential individuals, stemming from incomplete data or limited analytical capabilities.
Degree-Based Analysis	[23, 11, 1, 19, 58]	First, it calculates the degree of each node in the criminal network by counting their connections or associations. Nodes with higher degrees of connectivity are considered influential as they have more connections, allowing them to control information flow and criminal activities. The degree distribution reveals if the network follows certain pattern, like power law distribution, where a few individuals have significantly more connections. This helps identify influential criminals who play crucial roles in the network's operations	Criminal networks have complex social structures with diverse relationships and interactions. Analyzing degrees helps understand the network's structure by quantifying connections and identifying key players in central positions. Information flow is crucial in criminal networks for coordination, resource sharing, and control. Individuals with high degrees of connections have access to extensive information and engage in multiple interactions. Examining their degree centrality reveals influential individuals who facilitate information dissemination, activity coordination, and network control.	To enhance the method: (1) supplement degree analysis with contextual understanding of the specific criminal context, including the nature of activities, individual roles, and social factors (this avoids misinterpretations), (2) compare the analyzed network with baseline or reference networks created using random or hypothetical data (this helps identify influential criminals by establishing a benchmark for comparison), (3) integrate degree analysis with other network analysis techniques for a comprehensive understanding of the network's structure and dynamics, enhancing the insights gained.	(1) Degree-Based Analysis focuses solely on connections, neglecting their quality, strength, and nature, (2) it overlooks the nuances of power and influence in hierarchical criminal networks, potentially missing key players in significant roles, (3) Degree-Based Analysis doesn't consider changes in positions and influence within the network, leading to outdated assessments, and (4) it fails to incorporate contextual factors like reputation, trust, expertise, and resource access, limiting its ability to comprehensively understand influential individuals in the network

VI. EXPERIMENTAL EVALUATIONS

Within this section, we conduct experimentation to evaluate and rank the different techniques outlined in this paper. To represent each group of algorithms that share the same underlying technique, we selected a single algorithm as a representative. Subsequently, we assessed and ranked these chosen algorithms. We ran the algorithms on a Windows 10 Pro machine, which was equipped with an Intel(R) Core(TM) i7-6820HQ processor operating at 2.70 GHz and had 16 GB of RAM.

A. Datasets

The evaluations were conducted using the following datasets:

- *Chicago Crime Dataset:* The Chicago Police Department dataset, publicly accessible and dating from 2001, details reported city crimes. Updated regularly from the Department's CLEAR system, it includes information like crime type, location, timing, and arrest records. Organized by police district, it allows for citywide crime pattern analysis and includes fields like dates, addresses, coordinates, FBI codes, and location types. The dataset is downloadable as the Chicago Crime Dataset [22].
- *San Francisco crime dataset:* The dataset from San Francisco, featuring 39 crime categories, shows larceny/theft as the most common. The dataset was sourced from the San Francisco crime dataset [73].

B. Evaluation Setup

- *Katz centrality:* We assigned a value of $\alpha = 0.1$ to the attenuation factor (α), which regulates the impact of remote nodes on the centrality score. Moreover, we specified the maximum number of iterations, denoted as max_iterations = 100, for computing the centrality scores.
- *Closeness and Betweeness Centralities*: We utilized the conventional closeness centrality metric, which accounts for the shortest distance between nodes. We disregarded nodes that cannot be reached from other nodes. Then, we standardized the closeness centrality scores within the range of 0 to 1.
- *Node Similarity-Based Model:* We took the following into consideration: (1) we employed Cosine Similarity as the metric to measure similarity, (2) TF-IDF vectors were used to represent the features, (3) for each node, we identified the top 4 nearest neighbors, and (4) we applied a similarity threshold of 0.5 to remove connections that were considered weak.
- *Local Clustering Coefficient-Based Model:* For neighborhood size, we considered a larger neighborhood (degree-3 neighbors). We considered 0.001 for Learning Rate, 0.01 for Regularization Strength, 2 for number of layers, 64 for number of hidden units, 32 for batch size, 0.5 for dropout rate, and 100 for number of epochs.
- *Hierarchical-Based Clustering:* We employed the average linkage criterion and Euclidean distance as the distance metric for our hierarchical-based clustering. To achieve a desired number of clusters, we opted to truncate the dendrogram at a fusion coefficient of 0.5. Additionally, we established a distance threshold of 0.5, whereby the merging process halts if the dissimilarity between two clusters exceeds this value. Furthermore, we imposed a minimum cluster size of 50 data points, terminating the merging if a cluster falls below this threshold. For the dissimilarity percentage change, we set a termination threshold of 5%, causing the algorithm to cease merging if the dissimilarity between two merge steps decreases by less than this percentage.
- *Spatial Random Graph Distribution-Based Clustering*: We have defined the following parameters for our clustering algorithm: (1) Radius (R): We have set R to 100 meters, which determines the spatial proximity of points in the dataset, (2) k-nearest neighbors (k): We have set k to 5, which determines the number of nearest neighbors considered when constructing the spatial graph, (3) Distribution Threshold (DT): We have set DT to 0.6, which determines the threshold value used to decide if an edge should exist between two data points based on their distribution similarity, (4) Minimum Cluster Size (min_cluster_size): We have set it to 10, specifying the minimum number of points required for a cluster to be considered valid, (5) Spatial Density Threshold (SDT): We have set SDT to 0.4, which determines the threshold used to determine if a cluster is spatially dense enough, (6) Number of Iterations: We will perform 10 iterations to refine the cluster assignments, and (7) Random Seed (seed): We have set the seed to 42, which ensures the reproducibility of the clustering results by using a specific random seed value.
- *Spatial Diffusion-Based Clustering:* We establish these following values: (1) DT = 0.1 as the Distance Threshold (DT), dictating the maximum allowable distance between two data points to qualify them as neighbors, (2) TT = 5 minutes as the Time Threshold (TT), the maximum time difference between two data points to classify them as neighbors, 5 as the Minimum Cluster Size, and DR = 0.5 as the Medium Diffusion Rate.
- *Density-Based Clustering:* We establish the following values: (1) $\varepsilon = 0.5$ as the Epsilon (ε), indicating the radius within which the algorithm scans for neighboring points, and (2) MinPts = 5 as the Minimum Points (MinPts), defining the minimum number of points needed within the ε radius to constitute a dense region or cluster.
- *Fuzzy C-Means Clustering:* We established the following values: (1) m = 2.0 as the Fuzziness parameter (m) to regulate the level of fuzziness in the clustering process, (2) max_iter = 100 as the Maximum number of iterations (max_iter) to set the upper limit on how many times the algorithm should iterate before terminating, and (3) epsilon = 0.01 as the termination criterion (epsilon) to determine the convergence threshold (if the difference in membership values between successive iterations falls below epsilon, the algorithm stops).
- *Spatiotemporal Diffusion-Based Analysis:* We have defined the following parameters for our analysis: (1) Time Step (Δt): We have set Δt to be 1 hour. This value determines the duration between consecutive time points in the analysis, (2) Spatial Step (Δx, Δy, Δz): The values of Δx, Δy, and Δz have been set to 1 kilometer each. These values determine the distance between adjacent spatial grid points in each dimension, (3) Diffusion Coefficient (D): The value of D is 0.1 square kilometer per hour. This coefficient measures the speed at which the diffusing substance spreads and determines the rate of diffusion in the

analysis, (4) Initial Concentration (C_0): The initial concentration of the diffusing substance at the beginning of the analysis is set to $C_0 = 1$, and (5) Simulation Duration (T): The total duration of the simulation is set to T = 24 hours.

- *PageRank-Based Analysis:* We used the following settings: (1) damping factor (d): We chose a value of 0.9 for the damping factor. This factor determines the likelihood that a random surfer will continue clicking on links instead of jumping to a random node, and (2) convergence threshold for PageRank scores: We specified a small change in PageRank scores as the convergence threshold. This threshold determines when the PageRank has reached a stable solution and converged.
- *Eigenvector-Based Analysis*: We established the algorithm's termination condition to be a maximum of 100 iterations. Additionally, we determined that the algorithm should stop iterating when the convergence threshold reaches 1e-6.

C. Model for Simulating the Spreading Ability and Metrics for Evaluating the Performance of the Algorithms

To assess the accuracy of the ranked list of influential nodes generated by one of the selected algorithms, we conducted a comparison between the list and an actual propagation process involving the nodes. This evaluation was carried out using a widely recognized procedure outlined in Chen et al. [18]:

i. Recording the list of nodes ranked by each algorithm.
ii. Utilizing the SIR model (Chen et al. [18]) to simulate the spreading ability of nodes. Within this model, every node is assigned to one of three states: susceptible, infected, or recovered. In each state, only one node is considered infected. Infected nodes can infect their susceptible neighbors with a specific probability of spreading. In the experiments, we set the spreading probability β = 0.1-0.2, and the recovery probability μ = 1. Initially, we set the top-*k* ranked nodes to be infected, where *k* = 1% * *n* (*n* is the number of nodes). Thereafter, the number of infected nodes increases based on the SIR model.
iii. Using the nodes ranked by one of the algorithms and the corresponding one ranked by the SIR model, we recorded the pair scores in a list.

This survey utilizes the following three metrics to assess the performance of the algorithms:

- *Kendall's tau correlation coefficient τ (Kendall [46]):* It gauges the resemblance of data rankings between two quantities. The value of τ is in the range {+1, -1}. It is defined as shown in Equation 1.

$$\tau = \frac{N_1 - N_2}{0.5N\ (N-1)} \quad (1)$$

where N_1 and N_2 are the number of concordant and discordant pairs, respectively.

- *Monotonicity index (Zareie [88]):* It is a metric used for quantifying the resolution of different indices. It is defined as shown in Equation 2:

$$M(L) = \left[1 - \frac{\sum_{i \in L} |V|_i * (|V|_i - 1)}{|V| * (|V| - 1)}\right]^2 \quad (2)$$

where $|V|_i$ is the number of nodes that have the same index rank *l* in the ranked list *L;* and *V* is the number of nodes.

- *Percentage average absolute error (Saxena [72]):* It is a numerical amount of the discrepancy between an exact value and the corresponding estimated one. The absolute error (AE(*v*)) for a node *v* is defined as in Equation 3:

$$AE(v) = |EST(v) - ACT(v)| \quad (3)$$

The percentage average absolute error PAAE(*v*) for the node v is defined as shown in Equation 4:

$$PAAE(v) = \frac{\text{Average absolute error}}{\text{network size}} 100\% \quad (4)$$

D. Methodology for Selecting a Representative Algorithm for Each Technique and Ranking the Various Techniques

The following approach was employed for conducting the experimental evaluations:

- **Evaluating each sub-technique:** Upon conducting a thorough review of papers documenting algorithms that make use of a specific sub-technique, we successfully pinpointed the paper with the highest influence. The algorithm described in this paper was selected as the representative of the sub-technique. In order to ascertain the most noteworthy paper among all those reporting algorithms employing the same sub-technique, we evaluated several factors, including its level of innovation and publication date.
- **Ranking the sub-techniques that belong to the same overall technique:** We computed the average scores of the selected algorithms that employed a common sub-technique. Subsequently, we ranked the sub-techniques belonging to the same technique based on their scores.
- **Ranking the various techniques that belong to the same sub-category:** The average scores of the selected algorithms that employed a shared technique were calculated. Subsequently, we ranked the techniques falling under the same sub-category based on their scores.
- **Ranking the various sub-categories that belong to the same category:** We computed the average scores of the selected algorithms that utilized a common sub-category. Subsequently, we ranked the sub-categories falling under the same category based on their scores.

We searched for publicly accessible codes corresponding to the algorithms we chose to represent their respective techniques. Unfortunately, we were only able to acquire codes for the following two papers: (Cavallaro et al., [16]; Berlusconi et al., [10]). The codes for these papers are provided below:

- [16]: https://github.com/lcucav/ criminal-nets/tree/master/disruption
- [10]: https://figshare.com/articles/dataset/Oversize_network/3156067

For the remaining papers, we created our own implementations using TensorFlow, as described by Sinaga and Yang [67]. We trained these implementations using the Adam optimizer, as suggested by Sinaga and Yang [67]. TensorFlow's APIs offer users the ability to develop their own algorithms (Morselli and Giguere, [54]). Python 3.6 served as our development language, and we utilized TensorFlow 2.10.0 as the models' backend.

Tables 4-6 and Figs. 11-13 present the experimental results.

Table 4: Kendall's Coefficient scores of the algorithms. The table also shows the following rankings: (1) the various sub-techniques that belong to the same technique, (2) the various techniques that belong to the same sub-category, (3) the various sub-categories that belong to the same category, and (4) the categories.

Cat	Sub-Cat	Technique	Sub-Technique	Selected Papers	Data sets	Score	Sub-Tech. Rank	Tech Rank	Sub-Cat. Rank	Cat. Rank
Topology-Based Analysis	Global Analysis	Network-Based Analysis	Katz Centrality	Cavallaro [16]	Chi	0.851	1	2	1	1
					SFO	0.858				
			Multiple Link Types	Bright [9]	Chi	0.746	2			
					SFO	0.774				
		Shortest Path Analysis	Closeness Centrality	Calderon [15]	Chi	0.819	2	1		
					SFO	0.826				
			Betweeness Centrality	Taha [75]	Chi	0.829	1			
					SFO	0.832				
	Local Analysis	Node Similarity-Based	N/A	Berlusconi [10]	Chi	0.804	N/A	3	2	
					SFO	0.815				
		Local Clustering Coefficient	N/A	Agreste [2]	Chi	0.784	N/A	4		
					SFO	0.792				
Clustering-Based Analysis	Spatial Analysis	Global Analysis	Hierarchical Analysis	Xu [83]	Chi	0.772	1	1	1	3
					SFO	0.783				
			Spatial Random G. Distribution	Agarwal [4]	Chi	0.668	2			
					SFO	0.674				
		Local Analysis	Spatial Diffusion Clustering	Taha [77]	Chi	0.665	2	2		
					SFO	0.708				
			Density-Based Clustering	Everton [26]	Chi	0.687	1			
					SFO	0.710				
	Spatio-Temporal	Spatiotemporal Diffusion	N/A	Zhao [87]	Chi	0.649	N/A	1	2	
					SFO	0.693				
		Spatiotemporal Random G.	N/A	Berlusconi [6]	Chi	0.637	N/A	2		
					SFO	0.644				
Agent-Based	Eigenvector Centrality	PageRank-Based	N/A	Isah [35]	Chi	0.882	N/A	1	1	2
					SFO	0.911				
		Eigenvector	N/A	Ferrara [32]	Chi	0.879	N/A	2		
					SFO	0.899				
	Role Centrality	Behavior-Based	N/A	Hutchins [34]	Chi	0.611	N/A	1	2	
					SFO	0.623				
		Degree-Based	N/A	Bright [11]	Chi	0.573	N/A	2		
					SFO	0.587				

Table 5: Monotonicity scores of the algorithms. The table also shows the following rankings: (1) the various sub-techniques that belong to the same technique, (2) the various techniques that belong to the same sub-category, (3) the various sub-categories that belong to the same category, and (4) the various categories.

Cat	Sub-Cat	Technique	Sub-Technique	Selected Papers	Data sets	Score	Sub-Tech. Rank	Tech Rank	Sub-Cat. Rank	Cat. Rank
Topology-Based Analysis	Global Analysis	Network-Based Analysis	Katz Centrality	Cavallaro [16]	Chi	0.992	1	1	1	1
					SFO	0.994				
			Multiple Link Types	Bright [9]	Chi	0.989	2			
					SFO	0.992				
		Shortest Path Analysis	Closeness Centrality	Calderon [15]	Chi	0.991	1	2		
					SFO	0.990				
			Betweeness Centrality	Taha [75]	Chi	0.984	2			
					SFO	0.988				
	Local Analysis	Node Similarity-Based	N/A	Berlusconi [10]	Chi	0.981	N/A	1	2	
					SFO	0.985				
		Local Clustering Coefficient	N/A	Agreste [2]	Chi	0.977	N/A	2		
					SFO	0.982				
Clustering-Based Analysis	Spatial Analysis	Global Analysis	Hierarchical Analysis	Xu [83]	Chi	0.965	1	1	1	3
					SFO	0.973				
			Spatial Random G. Distribution	Agarwal [4]	Chi	0.961	2			
					SFO	0.969				
		Local Analysis	Spatial Diffusion Clustering	Taha [77]	Chi	0.962	1	2		
					SFO	0.970				
			Density-Based Clustering	Everton [26]	Chi	0.958	2			
					SFO	0.967				
	Spatio-Temporal	Spatiotemporal Diffusion	N/A	Zhao [87]	Chi	0.951	N/A	2	2	
					SFO	0.968				
		Spatiotemporal Random G.	N/A	Berlusconi [6]	Chi	0.955	N/A	1		
					SFO	0.963				
Agent-Based	Eigenvector Centrality	PageRank-Based	N/A	Isah [35]	Chi	0.992	N/A	1	1	2
					SFO	0.995				
		Eigenvector	N/A	Ferrara [32]	Chi	0.983	N/A	2		
					SFO	0.987				
	Role Centrality	Behavior-Based	N/A	Hutchins [34]	Chi	0.972	N/A	1	2	
					SFO	0.973				
		Degree-Based	N/A	Bright [11]	Chi	0.957	N/A	2		
					SFO	0.962				

Table 6: % average absolute error scores of the algorithms. The table also shows the following rankings: (1) the various sub-techniques that belong to the same technique, (2) the various techniques that belong to the same sub-category, (3) the various sub-categories that belong to the same category, and (4) the categories.

Cat	Sub-Cat	Technique	Sub-Technique	Selected Papers	Data sets	Score	Sub-Tech. Rank	Tech Rank	Sub-Cat. Rank	Cat. Rank
Topology-Based Analysis	Global Analysis	Network-Based Analysis	Katz Centrality	Cavallaro [16]	Chi	0.783	1	1	1	2
					SFO	0.537				
			Multiple Link Types	Bright [9]	Chi	0.847	2			
					SFO	0.658				
		Shortest Path Analysis	Closeness Centrality	Calderon [15]	Chi	1.573	2	2		
					SFO	0.873				
			Betweeness Centrality	Taha [75]	Chi	1.097	1			
					SFO	0.724				
	Local Analysis	Node Similarity-Based	N/A	Berlusconi [10]	Chi	2.107	N/A	1	2	
					SFO	0.566				
		Local Clustering Coefficient	N/A	Agreste [2]	Chi	2.647	N/A	2		
					SFO	0.941				
Clustering-Based Analysis	Spatial Analysis	Global Analysis	Hierarchical Analysis	Xu [83]	Chi	4.860	2	1	2	3
					SFO	1.467				
			Spatial Random G. Distribution	Agarwal [4]	Chi	4.491	1			
					SFO	1.142				
		Local Analysis	Spatial Diffusion Clustering	Taha [77]	Chi	5.853	1	2		
					SFO	2.073				
			Density-Based Clustering	Everton [26]	Chi	5.907	2			
					SFO	2.651				
	Spatio-Temporal	Spatiotemporal Diffusion	N/A	Zhao [87]	Chi	2.964	N/A	1	1	
					SFO	3.132				
		Spatiotemporal Random G.	N/A	Berlusconi [6]	Chi	3.482	N/A	2		
					SFO	3.292				
Agent-Based	Eigenvector Centrality	PageRank-Based	N/A	Isah [35]	Chi	0.408	N/A	1	1	1
					SFO	0.325				
		Eigenvector	N/A	Ferrara [32]	Chi	0.432	N/A	2		
					SFO	0.338				
	Role Centrality	Behavior-Based	N/A	Hutchins [34]	Chi	0.695	N/A	2	2	
					SFO	0.454				
		Degree-Based	N/A	Bright [11]	Chi	0.616	N/A	1		
					SFO	0.409				

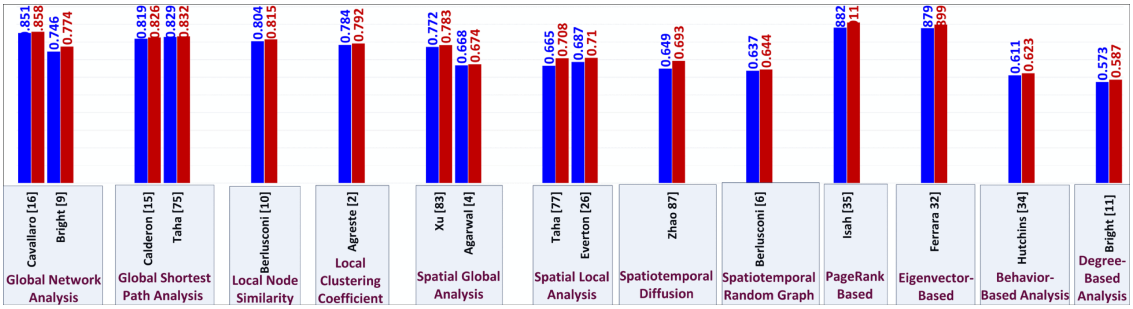

Fig. 11. Kendall's tau correlation coefficient τ scores of the algorithms. The algorithms are grouped based on the common underlying techniques they employ.

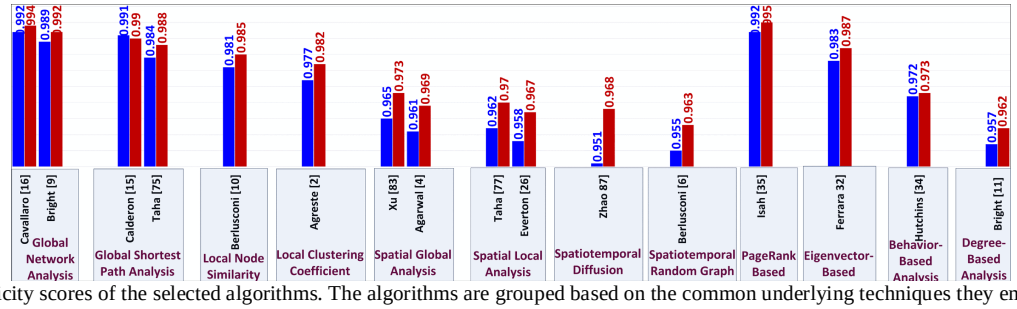

Fig. 12. Monotonicity scores of the selected algorithms. The algorithms are grouped based on the common underlying techniques they employ.

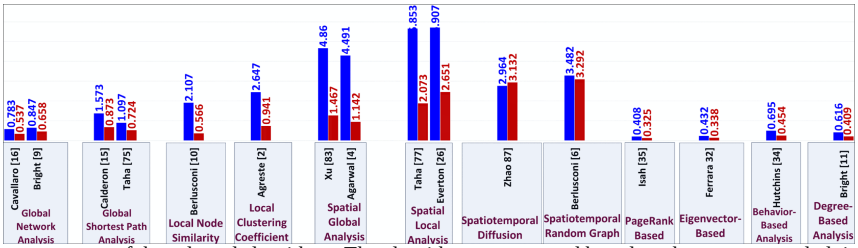

Fig. 13. % average absolute error scores of the selected algorithms. The algorithms are grouped based on the common underlying techniques they employ.

E. Discussion of the Experimental Results

1. Shortest Path-Based Analysis (Closeness-Based Centrality and Betweeness-Based Centrality)

The algorithms applied to the SIR model performed well overall. Algorithms based on betweenness centrality, however, were less effective in networks where information doesn't follow the shortest paths and failed to predict the influence of low-connectivity nodes. The inclusion of weak ties decreased accuracy, especially during initial stages, but performance improved in later stages and in well-connected networks. Closeness centrality-based algorithms efficiently computed node rankings, with a sigmoid pattern observed between closeness centrality and reverse ranking. Minor edge additions significantly boosted node centrality and ranking, identifying potential key players in criminal networks due to their proximity to other nodes. Algorithms using betweenness centrality excelled in blocking information flow and identifying influential nodes, especially when considering the ego-betweenness of top-ranked nodes. These results were efficient and comparable to general betweenness outcomes. This approach emphasized the role of individual connections in linking subgroups and spreading information, showing that individuals with high betweenness centrality are often involved in various criminal activities across different subgroups.

2. Spatial Global Analysis-Based (Spatial Random Graph and Distribution, Hierarchical Clustering)

The tested algorithms effectively propagated through networks, outperforming the single-point contact SIR model in generating infected nodes and identifying top-10 nodes with high accuracy. Random-based filtering in these algorithms showed superior performance compared to other approaches using a similar strategy, particularly effective in selecting high-degree nodes as seed nodes. However, instability issues arose due to the randomness in selection. The algorithms excelled in large-scale networks, and experiments varying the parameters of the spatial random graph distribution algorithm indicated robust performance in identifying crime leaders. This approach, emphasizing geographic proximity, was more effective in pinpointing crime leaders by leveraging spatial patterns in criminal activities and social connections. The influential nodes identified by these algorithms had a greater impact on spreading information compared to those ranked highly by degree, betweenness, eigenvector, or PageRank metrics. In terms of efficiency, these algorithms surpassed many popular Influence Maximization algorithms, maintaining comparable accuracy.

3. Spatiotemporal Clustering (Spatiotemporal Diffusion and Spatiotemporal Random Graph)

The node ranking correlation between the SIR model and the algorithms showed moderate agreement, with modest accuracy across different β values. The algorithms performed slightly better with β between 0.13 and 0.15, but less effectively for the top 3% of nodes. However, they successfully identified influential low-degree nodes near network cores. As the number of core nodes per community was adjusted, the algorithms' performance approached that of the SIR model, suggesting potential for optimizing diffusion performance. The Spatiotemporal Diffusion algorithms revealed temporal trends and identified influential individuals during specific periods, aiding law enforcement in resource allocation and targeting. By incorporating spatial data, these algorithms pinpointed crime clusters and key local nodes, suggesting strategies for disrupting localized criminal networks. Similarly, Spatiotemporal Random Graph algorithms detected temporal shifts in node centrality, highlighting nodes with fluctuating influence and bridge nodes connecting different network areas. Monitoring these nodes could disrupt cross-regional criminal activities and limit crime spread. These findings offer strategic insights for law enforcement in tackling criminal networks.

4. Eigen centrality Analysis (Eigenvector, PageRank)

The algorithms effectively handled varying propagation probabilities in networks, excelling in identifying structurally important nodes. They captured local dynamics through bridge-like structures and evaluated global roles based on key bridge connections. Infection extent depended on a node's neighbor count and their propagation capabilities, efficiently pinpointing central nodes in clusters and bridge nodes. In weighted social networks, these algorithms surpassed other centrality measures, highlighting an inverse relationship between sub-graph density and node centrality. They remained stable against random network perturbations and excelled in distinguishing nodes with different spreading abilities, balancing accuracy with computational efficiency. Eigenvector-Based algorithms effectively identified influential individuals in criminal networks, distinguishing crime leaders with high centrality scores. These leaders, often involved in multiple criminal activities, formed influence clusters within the network, with the algorithms accurately capturing the hierarchical structure.
PageRank-Based algorithms also successfully identified key crime leaders by assessing interconnectedness and influence. They excelled in ranking individuals by applying PageRank principles to analyze criminal network dynamics.

5. Role centrality Analysis (Degree and Behavior-Based)

These algorithms, focusing on local structural information, failed to consider the global network structure, leading to the identification of high-degree but low-influence nodes. Their accuracy declined with increased spreading degrees, performing poorly at 4-hop spreading. Additionally, as the value of k (number of top nodes) increased, the propagation capability of these top-k nodes diminished. The algorithms' performance worsened with larger datasets, particularly in networks with nodes of smaller degrees. In networks with unclear community structures, the algorithms underperformed, though they had acceptable propagation range and transmission rates in other networks. They efficiently summarized node connectivity without needing to analyze the entire network topology. However, the propagation ability of top-k nodes identified decreased with higher k values, and these influential nodes often connected different communities. Experimentally, the Degree-Based centrality algorithm showed some success in identifying crime leaders by focusing on individuals with numerous connections, yet it only achieved moderate performance, suggesting a limited ability to pinpoint true crime leaders. The behavior-based centrality algorithm, on the other

hand, effectively identified crime leaders by combining behavioral attributes with network structure, distinguishing leaders from other network members more accurately.

6. Network-Based Model Analysis (Katz Centrality and Multiple Link Types)

The analysis revealed that individuals with high Katz centrality scores, indicating many direct and indirect connections, are potential criminal leaders. This centrality measure identified influential individuals who may not have numerous direct connections but are strongly connected to other central figures, underscoring the role of indirect connections in understanding influence within the criminal network. High Katz centrality scores also helped uncover cohesive clusters or subgroups, representing distinct criminal organizations or factions led by central figures. The Multiple Link Types Model further differentiated the roles of criminal leaders based on various relationship types, like co-offending, communication, and financial transactions. For instance, high centrality from co-offending relationships indicated an individual's ability to coordinate criminal activities, while centrality from communication patterns pointed to their role as information hubs. Financial transaction-based centrality highlighted control over criminal finances. By analyzing individuals with high overall centrality scores in this model, key subgroups or clusters were identified, representing different criminal factions led by central figures. This approach provided valuable insights into the structure and dynamics of the criminal social network.

7. Spatial Local Analysis (Density-Based Clustering, Fuzzy C-Means Clustering, Spatial diffusion Clustering)

By applying the Density-Based Clustering (DBSCAN) algorithm with optimized parameters to the criminal social network dataset, clusters representing various criminal groups were identified. The algorithm effectively pinpointed potential crime leaders in each cluster based on their network positions and connections. The DBSCAN algorithm showed proficiency in identifying crime leaders, as evidenced by the strong cohesion within clusters and clear separation between them, indicated by the silhouette coefficient. Similarly, the Fuzzy C-Means (FCM) algorithm demonstrated effective performance in identifying crime leaders within the same network. The algorithm's membership values quantified individuals' associations with different criminal groups, facilitating the identification of influential members. Additionally, centrality measures indicated that these identified crime leaders held prominent positions within the network, exerting significant influence over other members.

VII. Potential Future Perspectives on Techniques for Identifying Crime Leaders

1) Katz Centrality

Katz centrality measures the influence of a node based on the number and importance of its neighboring nodes. In the future, Katz centrality could be enhanced by incorporating additional factors such as temporal dynamics, sentiment analysis, or multi-layer networks. By considering the evolving nature of criminal activities, sentiment analysis can help identify key individuals who exhibit patterns of involvement or influential behavior.

2) Multiple Link Types

Criminal social networks often have different types of relationships between individuals. Extending centrality measures to accommodate multiple link types could provide more nuanced insights. For example, a criminal network may have connections based on financial transactions, personal relationships, or shared locations. Incorporating such diverse link types into centrality calculations can reveal different dimensions of influence and identify individuals with varying roles in criminal activities.

3) Closeness Centrality

Closeness centrality quantifies the accessibility of a node within a network, based on the shortest paths to other nodes. Future perspectives for closeness centrality in criminal social networks could involve incorporating geographic factors, such as proximity to crime scenes or hotspots, as well as time-dependent factors, such as the frequency of interactions. By considering spatiotemporal aspects, closeness centrality can identify influential criminals who are geographically well-positioned and actively engaged in criminal activities.

4) Betweenness Centrality

Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes. In the future, betweenness centrality could be enhanced by considering the context of criminal activities and the flow of information or resources within the network. By incorporating additional information, such as the nature of criminal transactions or the exchange of illegal goods, betweenness centrality can identify individuals who act as intermediaries, controlling the flow of resources, or those who bridge groups within the network.

5) Node Similarity-Based Analysis

Future advancements in node similarity-based analysis could integrate graph neural networks (GNNs) to better capture complex patterns in criminal networks. GNNs can learn rich node embeddings, considering nodes and their neighborhoods, thus enhancing the identification of influential criminals by their similarity to known influential individuals.

6) Local Clustering Coefficient

Future use of the local clustering coefficient in criminal networks could include dynamic aspects of criminal activities, focusing on temporal patterns where clusters form or dissolve over time. Analyzing evolving local clustering coefficients can help identify influential criminals central to these temporal cluster changes.

7) Hierarchical Analysis

Hierarchical analysis aims to identify hierarchical structures within a network, such as nested clusters or levels of influence. In the future, advancements in hierarchical analysis for identifying influential criminals in a criminal social network could involve the integration of multi-resolution techniques. These techniques would allow for the identification of influential individuals at different scales, capturing both macro-level structures and micro-level dynamics. By uncovering hierarchical patterns of influence, law enforcement agencies can better understand the organization and power dynamics within criminal networks.

8) Spatial Random Graph Distribution

In the future, spatial random graph distribution for identifying

influential criminals could integrate geographic information systems (GIS) and spatial analytics. This approach would analyze the distribution of criminal activities and individual connectivity within geographical areas, identifying criminals with significant local presence or those linking different regions. Combining spatial random graph distribution with GIS can offer insights into the dynamics of criminal networks.

9) Spatial Diffusion Clustering

Spatial diffusion clustering focuses on the spread of information or activities within a network over space and time. Future perspectives for this approach in identifying influential criminals could involve the incorporation of machine learning algorithms capable of modeling and predicting the spatial diffusion of criminal activities. By understanding the patterns of how criminal activities spread and identifying individuals who are central to these diffusion processes, law enforcement agencies can effectively target and disrupt criminal networks.

10) Density-Based Clustering

Density-based clustering algorithms aim to identify clusters in a network based on the density of connections. In the context of identifying influential criminals, future perspectives for density-based clustering could involve the integration of multiple data sources, such as social media data, telecommunications records, or financial transactions. By combining network data with external data sources, law enforcement agencies can gain a more comprehensive understanding of criminal networks and identify influential individuals based on their connectivity patterns and the richness of information available.

11) Spatiotemporal Diffusion

Spatiotemporal diffusion analysis focuses on understanding the spread of information, behaviors, or activities over both space and time. In the context of identifying influential criminals, future perspectives for spatiotemporal diffusion analysis could involve the integration of advanced machine learning techniques and big data analytics, as follows:

1. *Predictive Modeling:* Machine learning advancements, particularly spatiotemporal forecasting models, enable prediction of future criminal activity using historical data. These models consider factors like spatial proximity, temporal patterns, social dynamics, and environmental influences, helping law enforcement identify and prioritize individuals likely to be central in criminal activities.
2. *Real-Time Monitoring:* Incorporating real-time data like social media, surveillance footage, and sensor data improves spatiotemporal diffusion analysis. Law enforcement can use this for real-time monitoring of criminal activities, identifying key players in criminal operations and the diffusion process. This enables proactive disruption of criminal networks.

12) Spatiotemporal Random Graph

Spatiotemporal random graph analysis considers both the spatial and temporal dimensions of a network, incorporating the interactions between nodes over time and across geographical locations. Future perspectives for spatiotemporal random graph in identifying influential criminals could involve the following:

1. *Network Evolution Analysis:* Criminal networks, characterized by evolving strategies, alliances, and activities, change over time. Analyzing their spatiotemporal evolution helps identify criminals with stable influence or those rapidly gaining power. Understanding these dynamics offers insights into the stability and resilience of criminal organizations.
2. *Community Detection:* Community detection algorithms in spatiotemporal random graphs reveal clusters of individuals with strong spatiotemporal connections. Identifying these densely connected criminal groups enables law enforcement to disrupt these communities and target influential criminals linking different clusters.

13) PageRank-Based Analysis

PageRank, which gauges node importance in a network through connectivity and neighbor importance, can be enhanced for identifying influential criminals by incorporating temporal dynamics, criminal activity patterns, and individual attributes. Adapting PageRank-based algorithms to consider the evolving nature of criminal networks and individual behaviors over time can more accurately capture the influence of criminals who change strategies or engagement patterns.

14) Eigenvector Centrality

Eigenvector centrality, which assesses a node's influence based on its connections and its neighbors' influence, can be improved for identifying influential criminals by integrating multi-layer or multi-modal networks. Considering various types of interactions in criminal networks, like financial transactions, communication, or shared locations, will allow eigenvector centrality to offer a more holistic assessment of criminal influence across multiple dimensions.

15) Behavior-Based Analysis

Future approaches to behavior-based analysis for identifying influential criminals could integrate machine learning to analyze complex behavioral patterns. Leveraging advanced analytics, like anomaly detection algorithms, this method can pinpoint influential criminals with unique or abnormal behaviors, uncovering hidden key individuals not easily detected through network structure alone.

16) Degree Centrality

Future applications of degree centrality in identifying influential criminals could involve using weighted networks to account for the strength or importance of connections. This approach would allow for a more nuanced understanding of influence by considering the intensity or significance of relationships, helping to identify criminals with not only numerous connections but also influential ties in the network.

VIII. CONCLUSIONS

This survey paper tackles the issue of vague and generalized categorizations in algorithmic approaches to crime leader identification and prediction. Traditional surveys often use broad classifications, leading to misalignments and imprecise evaluations. In response, our work introduces a novel, detailed methodological taxonomy, specifically for predicting crime leaders. We divide crime leader identification algorithms into three main classes: topology-based, clustering-based, and agent-based methods. Each class is further subdivided into three increasingly specific tiers, refining categorization and improving the precision and assessment of algorithms. Our key contributions are threefold as followed:

1. Our survey provides a detailed analysis of crime prediction

algorithms, focusing on their sub-techniques, techniques, and categories. This taxonomy aids in accurate assessments, enhancing understanding of these algorithms' strengths and limitations, crucial for future research.

2. We conducted an empirical evaluation of techniques for identifying crime leaders, using four criteria to offer insights into their practical efficacy and applicability.
3. Our experimental evaluation compares and ranks numerous algorithms across different levels: sub-techniques, techniques, sub-categories, and categories. This comprehensive analysis gives a nuanced view of their performance and appropriateness in various scenarios.

Below, we present the main discoveries from our experimental outcomes:

- *The techniques yielded the best results (PageRank-Based and Eigenvector):* The algorithms excelled across varying propagation probabilities, effectively mapping structural dependencies in dense network regions. They proficiently identified both central nodes in clusters and connectors between network segments. An inverse relationship was noted between sub-graph density and node centrality. Resilient to random network perturbations, these algorithms surpassed others in assigning distinct rankings to nodes based on their spreading capabilities, balancing sorting accuracy with computational efficiency. Their accurate portrayal of hierarchical structures in networks highlights their potential in pinpointing key players and provides deep insights into the structural aspects of criminal networks.
- *The technique that achieved the second highest performance (Katz Centrality and Multiple Link Types):* Katz centrality effectively identified influential individuals in a criminal network, emphasizing the role of indirect connections. This measure revealed distinct clusters or subgroups, likely representing different criminal organizations or factions led by central figures, demonstrating its ability to uncover the network's hierarchical structure. The model highlighted the roles and characteristics of criminal leaders. By analyzing individuals with high centrality scores in this model, key subgroups or clusters associated with various criminal organizations were identified, further indicating the presence of central figures or leaders in these groups.
- *The technique that achieved the lowest performance (Spatiotemporal Random Graph):* The agreement between node rankings in the SIR model and the algorithms was moderate, varying with different β values, indicating modest accuracy in node ranking. As the number of ranked nodes increased, the algorithms' ability to improve rankings diminished. While there was a slight improvement within the β range of 0.13 to 0.15, the algorithms performed poorly among the top 3% of ranked nodes.

References

[1] Agarwal, A., Uniyal, D., Toshniwal, D. and Deb, D. "Dense Vector Embedding Based Approach to Identify Prominent Disseminators From Twitter Data Amid COVID-19 Outbreak," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(3):308-320, 2021.

[2] Agreste, S., Catanese, S., & Fiumara, G. Network structure and resilience of mafia syndicates. Information Sciences, 351(C), 30–47, 2016.

[3] Afra, S., & Alhajj, R. (2021). Integrated framework for criminal network extraction from Web. *Journal of Information Science*, *47*(2), 206–226.

[4] Agarwal, A., Toshniwal, D. Identifying Leadership Characteristics from Social Media Data during Natural Hazards using Personality Traits. *Sci Rep* 10, 2624 (2020).

[5] Bright, D. A., Hughes, C. E., & Chalmers, J. (2012). Illuminating dark networks: A social network analysis of an Australian drug trafficking syndicate. Crime, Law and Social Change, 57(2), 151–176.

[6] Berlusconi, G. Come at the king, you best not miss: criminal network adaptation after law enforcement targeting of key players, Global Crime, 23:1, 44-64, 2022.

[7] Bright, D., Delaney, J. Evolution of a drug trafficking network: mapping changes in network structure and function across time. Global Crime 14(2–3), 238–260 (2013).

[8] Budur, E. and Lee, S. Kong, V. "Structural Analysis of Criminal Network and Predicting Hidden Links using Machine Learning", ArXiv, 2015.

[9] Bright, D., Greenhill, C., Morselli, C. Networks within networks: using multiple link types to examine network structure and identify key actors in a drug trafficking operation. Global Crime 16(3), 1–19 (2015).

[10] Berlusconi, G., Calderoni, Piccardi, C. Link Prediction in Criminal Networks: A Tool for Criminal Intelligence Analysis. PLoS ONE 2016.

[11] Bright, D. Greenhill, C., Britz, T., Ritter, A., Morselli, C. (2017) Criminal network vulnerabilities and adaptations, Global Crime, 18:4, 424-441.

[12] Calderoni, F. (2014). Identifying Mafia Bosses from Meeting Attendance. In: Masys, A. (eds) Networks and Network Analysis for Defence and Security. Lecture Notes in Social Networks. Springer, Cham.

[13] Calderoni, F., et al. "Robust link prediction in criminal networks: A case study of the Sicilian Mafia," Expert Systems with Applications, 2020.

[14] Calderoni, F. Brunetto, D., Piccardi, C. "Communities in criminal networks: A case study, Social Networks", V. 48, 2017, Pages 116-125.

[15] Calderoni, F., David B., Zheng, Q. "Inductive Discovery Of Criminal Group Structure Using Spectral Embedding", Inform. & Security, 2014.

[16] Cavallaro L, et al. (2020) Disrupting resilient criminal networks through data analysis: The case of Sicilian Mafia. PLoS ONE 15(8).

[17] Calderoni, F., Superchi, E. The nature of organized crime leadership: criminal leaders in meeting and wiretap networks. *Crime Law Soc Change* 72, 419–444 (2019).

[18] Chen, D. et al. Identifying influential nodes in complex networks, Physica a: Statistical mechanics and its applications, 391(4):1777–1787, 2012

[19] Colladon, A. and Remondi, E. Using social network analysis to prevent money laundering. Expert Syst. Appl. 67, 49–58 (2017).

[20] Catanese, S., Ferrara, E. & Fiumara, G. Forensic analysis of phone call networks. *Soc. Netw. Anal. Min.* 3, 15–33 (2013).

[21] Campana, P., & Varese, F. (2011). Listening to the wire: Criteria and techniques for the quantitative analysis of phone intercepts. Trends in Organized Crime, 15(1), 13–30.

[22] Chicago Crime Dataset (2023): https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2

[23] Duijn, P.A.C., Klerks, P.P.H.M. (2014). Social Network Analysis Applied to Criminal Networks: Recent Developments in Dutch Law Enforcement. In: Masys, A. (eds) Networks and Network Analysis for Defence and Security. Lecture Notes in Social Networks. Springer, Cham.

[24] Décary-Hétu, D., Dupont, B. The social network of hackers. Global Crime 13(3), 160–175 (2012).

[25] Duijn, P., Kashirin, V. & Sloot, P. The Relative Ineffectiveness of Criminal Network Disruption. *Sci Rep* 4, 4238 (2014).

[26] Everton, S. et al. Dark network resilience in a hostile environment: optimizing centralization and density. Cri. Just. Law Soc. 16(1):1–20, 2015

[27] Easton, S. and Karaivanov, A. "Understanding optimal criminal networks", Global Crime Journal, 2009, volume{10},pages={41 - 65}.

[28] Felson, M. The natural history of extended co-offending, Trends in Organised Crime (12): p. 159–165, 2009.

[29] Freeman, L. C. (1979). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215–239.

[30] Fidalgo, P., et al. "Star-Bridge: a topological multidimensional subgraph analysis to detect fraudulent nodes and rings in telecom networks," *IEEE Inter Conf on Big Data (Big Data)*, Osaka, Japan, 2022, pp. 2239-2242.

[31] Ficara, A. et al. "The Whole Is Greater than the Sum of the Parts: A Multilayer Approach on Criminal Networks" *Futur. Int.* 14(5): 123, 2022.

[32] Ferrara, E. *et al.*. Detecting criminal organizations in mobile phone networks. Exp. Syst. Appl. 2014, 41, 5733–5750.

[33] Ficara, A., Curreri, F., Fiumara and De Meo, P. "Human and Social Capital Strategies for Mafia Network Disruption," in *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1926-1936, 2023

[34] Hutchins, C., Benham-Hutchins, M. Hiding in plain sight: criminal network analysis. Comput. Math. Org. Theor. 16(1), 89–111 (2010).

[35] Isah, H., Neagu, D., Trundle, P. "Bipartite network model for inferring hidden ties in crime data," *IEEE/ACM ASONAM*, Paris, France, 2015.

[36] Javed, A., Ahmed, W., Kifayat, K., Gadekallu, T. "A Comprehensive Survey on Computer Forensics: State-of-the-Art, Tools, Techniques, Challenges, and Future Directions," *IEEE Access*, 10:11065-11089, 2022.

[37] Guo, Z., Cho, J., Hong, M. and Mitra, T. "Online Social Deception and Its Countermeasures: A Survey," *IEEE Access*, 9:1770-1806, 2021.

[38] Gunnell, D., Hillier, J., Blakeborough, L. Social network analysis of an

urban street gang using police intelligence data (2016).
[39] Grassi, R. et al. “Betweenness to assess leaders in criminal networks: New evidence using the dual projection approach” Soc. Net., 56:23-32, 2019.
[40] Griffiths, G., et al. UK-based terrorists’ antecedent behavior: A spatial and temporal analysis. Appl. Geogr. 2017, 86, 274–282.
[41] Kleemans, E. et al. (2008). Criminal careers in organized crime and social opportunity structure. European Journal of Criminology, 5(1), 69–98.
[42] Kleemans, E. R., & Van Koppen, M. V. (2014). Careers in organized crime. In G. Bruinsma & D. Weisburd (Eds.), Encyclopedia of criminology and criminal justice (pp. 285–295). Springer New York.
[43] Kumari, S., et al. Intelligent deception techniques against adversarial attack on the industrial system. Int. J. Intell. Syst. 36, 5, 2021, 2412–2437.
[44] Kazmi, F., Butt, W., Saeed, A. Evaluation of Role Detection Approaches in Terrorist Networks. International Conf. on Management Engineering, Software Engineering and Service Sciences (ICMSS). 245–249, 2018.
[45] Kleemans, E. R. (2014). Theoretical perspectives on organized crime. In L. Paoli (Ed.), The Oxford handbook of organized crime (pp. 32–52). London: Oxford University Press.
[46] Kendall, M. “The treatment of ties in ranking problems," *Biometrika*, vol. 33, no. 3, 1945, pp. 239_251
[47] Kleemans, E. R., & Van De Bunt, H. (1999). The social embeddedness of organized crime. Transnational Organized Crime, 5(1), 19–36 [13z].
[48] Kleemans, E. R., Brienen, M. E. I., & Van De Bunt, H. (2002). Dutch organized crime monitor: Georganiseerde criminaliteit in Nederland. Tweede Reportage Op Basis van de WODC-Monitor. Den Haag: WODC.
[49] Kawthalkar, I., et al. A. "A Survey of Predictive Crime Mapping Techniques for Smart Cities," *National Conf on Emerg. Trends on Sust. Tech. & Eng. Applications (NCETSTEA)*, Durgapur, India, pp. 1-6, 2020.
[50] Morselli, C., Giguère, C., & Petit, K. (2007). The efficiency/security trade-off in criminal networks. Social Networks, 29(1), 143–153.
[51] Morselli, C. (2009). Inside criminal networks. New York: Springer.
[52] Meneghini, C., Aziani, A. & Dugato, M. Modeling the structure and dynamics of transnational illicit networks: an application to cigarette trafficking. *Appl Netw Sci* 5, 21 (2020).
[53] Memon, B. "Identifying Important Nodes in Weighted Covert Networks Using Generalized Centrality Measures," *2012 European Intelligence and Security Informatics Conference*, Odense, Denmark, 2012, pp. 131-140.
[54] Morselli, C. and Giguere, C. “Legitimate strengths in criminal networks,” *Crime, Law Social Change*, 45(3):185–200, 2006
[55] Maulana, A. and Emmerich, M. "Towards many-objective optimization of eigenvector centrality in multiplex networks," *Inter. Conf. on Control, Dec. & Info. Tech. (CoDIT)*, Barcelona, Spain, 2017, pp. 0729-0734.
[56] Malm, A., Bichler, G., Nash, R. Co-offending between criminal enterprise groups. Global Crime 12(2), 112–128 (2011).
[57] Ozgul, F. and Erdem, Z. "Deciding resilient criminal networks," *IEEE/ACM ASONAM*, Paris, France, 2015, pp. 1368-1372.
[58] Petersen, R., et al. "Node Removal in Criminal Networks," *European Intell. & Sec. Informatics Conference*, Athens, Greece, 2011, pp. 360-365.
[59] Premasundari, M. and Yamini, C. “A violent crime analysis using fuzzy c-means clustering approach”, Journal on Soft Computing, 2019, 9(3).
[60] Park, A. and Tsang, H. "Detecting Key Players in Criminal Networks Using Dynalink," *2013 European Intelligence and Security Informatics Conference*, Uppsala, Sweden, 2013, pp. 208-211.
[61] Papachristos, A. et al. The embedded and multiplex nature of Al Capone. In C. Morselli (Ed.), Crime and networks, pp. 97–115, 2014. New York.
[62] Paoli, L. (2002). The paradoxes of organized crime. Crime, Law and Social Change, 37(1), 51–97.
[63] Reuter, P. (1983). Disorganized crime: The economics of the visible hand. Cambridge: MIT Press.
[64] Reiss, J. Albert J. and Farrington, D. “Advancing knowledge about co-offending: Results from a prospective longitudinal survey of london males,” Criminal Law and Criminology (1973), vol. 8(2); 360–395, 1991.
[65] Rodrigueza, R. and Estuar, M. "Social Network Analysis of a Disaster Behavior Network: An Agent-Based Modeling Approach," *IEEE/ACM Adv Soc, Net. Min (ASONAM)*, Barcelona, Spain, 2018, pp. 1100-1107.
[66] Shafia and Chachoo, M. "Social Network Analysis Based Criminal Community Identification Model with Community Structures and Node Attributes," *ICSSIT Conference*, Tirunelveli, India, 2022, pp. 334-339.
[67] Sinaga. K. and Yang, M. "Unsupervised K-Means Clustering Algorithm," in *IEEE Access*, vol. 8, pp. 80716-80727, 2020.
[68] Song, C., et al.Graphic model analysis of frauds in online consumer reviews. Intern. Conf Int things, Data Clou Comp (ICC). Article 47, 1–7
[69] Shang, X. and Yuan, Y. “Social Network Analysis in Multiple Social Networks Data for Criminal Group Discovery”, Inter. Conf. on Cyber-Enabled Distributed Computing Knowledge Discovery, 2012, pp. 27-30.
[70] Sivanagaleela, B. and Rajesh, S. "Crime Analysis and Prediction Using Fuzzy C-Means Algorithm," *International Conference on Trends in Electronics & Informatics (ICOEI)*, Tirunelveli, India, 2019, pp. 595-599.
[71] Siriwat, P. and Nijman, v. “Quantifying the illegal high-value rosewood trade and criminal trade networks in the Greater Mekong Region,”, Biological Conservation, Volume 277, 2023, 109826.
[72] Saxena, A., Gera, R., and Iyengar, S. "Fast estimation of closeness centrality ranking.", in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,* 2017, pp. 80-85.
[73] San Francisco Crime Dataset (2023): https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnf-yvry
[74] Schwartz, D.M. and Rouselle, T. Using social network analysis to target criminal networks. *Trends Organ Crim* 12, 188–207 (2009).
[75] Taha, K., and Yoo, P. "Using the Spanning Tree of a Criminal Network for Identifying Its Leaders,". *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 2, 2017, pp. 445-453.
[76] Taha, K., and Yoo, P. "Shortlisting the Influential Members of Criminal Organizations and Identifying Their Important Communication Channels,". *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 8, 2019, pp. 1988-1999.
[77] Taha, K. and Yoo, P. "A system for analyzing criminal social networks," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, France, 2015, pp. 1017-1023
[78] Tundis, A. et al. "Similarity Analysis of Criminals on Social Networks: An Example on Twitter," *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, Valencia, Spain, 2019, pp. 1-9.
[79] Vargas, R. (2014). Criminal group embeddedness and the adverse effects of arresting a Gang’s leader: A comparative case study. Criminology, 52(2), 143–168.
[80] Wang, S., et al. "Cyber Threat Analysis and Trustworthy Artificial Intelligence," *Inter. Conf. Cry., Sec. Priv. (CSP)*, China, 2022, pp. 86-90.
[81] Win, K.N., Chen, J., Chen, Y. *et al.* PCPD: A Parallel Crime Pattern Discovery System for Large-Scale Spatiotemporal Data Based on Fuzzy Clustering. *Int. J. Fuzzy Syst.* 21, 1961–1974 (2019).
[82] Wang, J., et al. "An Investigation of Cybercrime-Related Online Search Behavior vs General Search Behavior," *Inter. Conf. Internet Monitoring and Protection (ICIMP 2007)*, San Jose, CA, USA, 2007, pp. 4-4.
[83] Xu J, Chen H (2005) CrimeNet explorer: a framework for criminal network knowledge discovery. Trans Inf Syst 23(2):201–226.
[84] Yang, C., et al. "A Rough-fuzzy C-means using information entropy for discretized violent crimes data," *13th International Conference on Hybrid Intelligent Systems (HIS 2013)*, Gammarth, Tunisia, 2013, pp. 23-27.
[85] Yang, L. "Based on social network crime organization relation mining and central figure determining," *IEEE International Conference on Computer Science and Automation Engineering*, Beijing, 2012, pp. 55-58.
[86] Zhang, Y., et al. "Identifying Node Importance by Combining Betweenness Centrality and Katz Centrality," *Inter. Conference on Cloud Computing and Big Data (CCBD)*, Shanghai, China, 2015, pp. 354-357.
[87] Zhao, X. and Tang, J. “Crime in Urban Areas: A Data Mining Perspective”, SIGKDD} Explor. Volume 20, number 1, pp. 1-12, 2018.
[88] Zareie, A., Sheikhahmadi, A., Jalili, M., Fasaei, K. “Finding influential nodes in social networks based on neighborhood correlation coefficient”. *Knowl.-Based Syst,* vol. 194, 2020, 105580.

Kamal Taha is an Associate Professor in the Department of Electrical Engineering and Computer Science at Khalifa University, UAE, since 2010. He received his Ph.D. in Computer Science from the University of Texas at Arlington, USA. He has over 100 refereed publications that have appeared in prestigious top ranked journals, conference proceedings, and book chapters. Over 30 of his publications have appeared in IEEE Transactions journals. He was as an Instructor of Computer Science at the University of Texas at Arlington, USA, from August 2008 to August 2010. He worked as Engineering Specialist for Seagate Technology, USA, from 1996 to 2005 *(Seagate is a leading computer disc drive manufacturer in the US)*. His research interests span information forensics & security, bioinformatics, information retrieval, data mining, databases, and defect characterization of semiconductor wafers,, with an emphasis on making data retrieval and exploration in emerging applications more effective, efficient, and robust.

Abdulhadi Shoufan received his PhD from the Technische Universitaet Darmstadt, Germany in 2007. From 2008 to 2010, he led the Security Hardware group at the Center of Advanced Security Research Darmstadt and developed hardware solutions for post-quantum cryptography. He is currently an Associate Professor at Khalifa University. He is interested in zero-trust architecture, embedded security, cryptographic hardware, UAVs' safe and secure operations, unmanned traffic management (UTM), learning technology, and engineering education. He worked on different projects for Boeing, Lockheed Martin, PSEG, TII, and the Ministry of Education for Information Security in Germany.