2023

Community and key player detection for disrupting illicit drug supply networks in social media platforms – especially on Instagram

Akassi Rachel Niamke Aman
West Virginia University, arn0025@mix.wvu.edu

Community and key player detection for disrupting illicit drug supply networks in social media platforms – especially on Instagram.

Akassi Rachel Niamke Epse Aman

Thesis submitted to the Statler College of Engineering and Mineral Resources

at West Virginia University

in partial fulfillment of the requirements for the degree of

Master in

Electrical Engineering

Xin Li, Ph.D. Chair

Brian Woerner, Ph.D.

Omid Dehzangi, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia

2023

Keywords: Social Media, Instagram, Illicit Drug traffic, Community Detection, Key player

ABSTRACT

Community and key player detection for disrupting illicit drug supply networks in social media platforms – especially on Instagram.

Akassi Rachel Niamke Epse Aman

This thesis focuses on the pressing issue of illicit drug trafficking and its impact on public health and safety at a global level. With the advent of digital technologies and social media platforms, combating drug trafficking has become increasingly challenging for law enforcement and researchers alike. Among these platforms, Instagram, a popular photo and video-sharing social networking platform, has emerged as a prominent hub for drug trafficking activities.

In this study, we delve into the effectiveness of community and key player detection algorithms in identifying and disrupting illicit drug supply networks on Instagram. To conduct our research, we collected real Instagram data spanning from June to August 2022. We examined several community detection algorithms, including Louvain, Newman-Girvan, Infomap, Label Propagation, and Hierarchical Clustering. Additionally, we explored key player algorithms, namely CDKPE, TopRank, K-core, and KPEI, the latter being a novel algorithm introduced in this study. Our objective was to assess the performance of these algorithms in accurately identifying and targeting drug trafficking networks.

Our findings reveal that the Louvain and Newman-Girvan algorithms outperformed others in terms of community detection, demonstrating their effectiveness in identifying cohesive groups involved in drug trafficking on Instagram. In terms of key player detection, the CDKPE and KPEI algorithms emerged as the most effective, highlighting the individuals who play pivotal roles within these networks.

These algorithms offer practical applications for law enforcement agencies seeking to disrupt drug trafficking operations on Instagram. By emphasizing the importance of leveraging advanced analytical tools, our study underscores the significance of combating drug trafficking on social media platforms.

Acknowledgements

I would like to express my sincere gratitude to Dr. Xin Li for taking me in as a member of his social media research team and providing me with a graduate assistant position, which enabled me to complete my master's degree in electrical engineering. I am also grateful to Dr. Muhammad Choudhry for his guidance and support during the period without an academic advisor. I would also like to thank Dr. Chuanbo Hu for his invaluable support during my research and the committee members, Dr. Brian Woerner and Dr. Omid Dehzangi, for their constructive feedback and suggestions. Additionally, I would like to extend my appreciation to all my family members for their unwavering support and encouragement throughout this journey.

Table of Contents

List of Tables

List of Figures

Chapiter 1. Introduction

Illicit drug supply networks continue to operate through social media platforms, with Instagram being a particularly popular platform for drug dealers to advertise and sell their products. To combat this problem, community and key player detection techniques can be used to identify individuals and groups involved in these activities. By understanding the structure and dynamics of these networks, law enforcement agencies can disrupt and dismantle them more effectively. In this discussion, we will explore some methods for community and key player detection to disrupt illicit drug supply networks on Instagram.

1.1.Background

Illicit drug use has become a growing concern for society, with increasing trends over the years. According to the National Center for Drug Abuse Statistics (NCDAS) [1], approximately 37.31 million Americans aged 12 or older (13.5% of the population) were current illegal drug users (used within the last 30 days) in 2020, up from 3.8% from the previous year. The rise in drug use reflects the growth of illicit drug trade in the US, which has shifted from traditional offline methods to online platforms due to the proliferation of smartphones and social media growth [2]. Recent research has shown that popular social media platforms such as Instagram, Twitter, and Facebook are now utilized by drug dealers as communication and marketing tools [3]. The estimated global drug market revenue was about $426-$652 billion in 2017 [4], which has led to drug trafficking evolving with the advancement of modern technologies, creating new challenges for law enforcement agencies and researchers in disrupting drug trafficking networks.

1.2.Motivation

Drug trafficking through social media platforms such as Instagram has become a significant challenge for law enforcement agencies and public health agencies, as it can exacerbate drug abuse, particularly among young people, who make up a large proportion of Instagram's user base [5][6]. In recent years, there has been growing interest in using machine learning algorithms to automatically detect drug dealer accounts on Instagram. One promising approach is the use of multimodal machine learning systems that can identify drug dealer accounts and posts by analyzing a range of content, including image content [7]. However, detecting the entire community around these drug dealers, as well as identifying key players within these communities, remains a major challenge.

Community detection in social networks involves identifying groups of nodes that are densely connected to each other and weakly connected to other groups [8]. In the context of illicit drug trade on Instagram, communities can be thought of as groups of users who are involved in the same drug supply network. By detecting and analyzing these communities, law enforcement agencies can gain a better understanding of how drugs are being sold on the platform and develop more effective strategies to disrupt these networks.

Identifying key players within these communities is also crucial. Key players are individuals who are well-connected and influential within their communities and can have a significant impact on the beliefs and behaviors of other members of the community [9]. By targeting key players within illicit drug supply networks, law enforcement agencies can disrupt and dismantle these networks more effectively. This approach has the potential to change the beliefs and behaviors of other community members, leading to a reduction in illicit drug trade on the platform.

While there has been some research on community detection and key player identification in other domains, there is a lack of research on identifying drug supply networks and key players on social media platforms like Instagram [10]. Therefore, developing effective algorithms for community detection and key player identification is an urgent research need. Such algorithms can assist law enforcement agencies in combating the sale of drugs online and reduce the negative impact of drug abuse on society.

1.3.Overview

In this study, we evaluated the performance of different algorithms for detecting drug dealer communities and key players on social network data. Furthermore, we introduced a novel algorithm for this purpose. The study utilized the "Identifying Drug Dealers on Instagram (IDDIG)" database, which contained 4,000 user accounts, including 1,400 drug dealers [11]. Our analysis focused on 115 out of these 1,400 confirmed dealer accounts, with 14,034 posts and 31,324 related comments collected between June 2022 and August 2022.

The paper is structured as follows: Section 2 provides a review of related research on social media data mining for disrupting illicit drug trade, community detection, and key player identification on social media. In Section 3, we describe our proposed method for community and key player detection. Section 4 presents our data collection and results, while Section 5 offers conclusions and discussion. Finally, Section 6 includes the references cited in this study.

Chapter 2. Literature survey

Social media platforms have become a ubiquitous source of data, offering insights into a wide range of topics, including drug use and trafficking. Mining social media data has the potential to disrupt illicit drug activity by identifying and tracking key players, uncovering patterns of behavior, and facilitating targeted interventions. In this literature survey, we examine recent research related to the social media data mining for illicit drug disruption as well as detection of community and key player.

2.1. Social Media Data Mining for illicit drug disruption

In recent years, social media data mining has gained popularity due to the abundance of data available on these platforms. Several studies have been conducted to mine social media data for various purposes, including recommendation systems, social affairs, and public health. However, limited work has been done on tracking drug abuse and dealing on social media. Choudhury et al. [12] proposed a framework for analyzing social media data to identify patterns of drug use and trafficking, while Buntain and Golbeck [13] analyzed the time and location patterns of drug use by mining Twitter data. Yiheng Zhou and Luo [14] described a method for analyzing drug use patterns using Instagram data. Recently, Hu et al. [11] introduced a multimodal approach to identify illicit drug dealers on Instagram using text, image, and network data. They collected a dataset of 20,000 Instagram posts related to drugs and achieved an F1-score of 0.87 for identifying drug dealers. In this study, we will use some of the more than 1,400 identified drug dealer accounts to conduct our research.

2.2. Community detection in social media

Social media networks are complex, and detecting communities within them can provide insights into user behavior, social dynamics, and the spread of information.

Several algorithms have been proposed for community detection in social media.

Wu et al[15] provided a review of community detection algorithms in social media and discusses their applications in various domains, such as information retrieval, recommendation systems, and social media marketing. The authors also provided a critical analysis of the limitations and challenges of community detection in social media. Mahmood et al [16] presented a comparative study of different community detection algorithms in social media, including modularity-based methods, hierarchical clustering, and label propagation. The authors evaluate the performance of the algorithms on different datasets and discuss their strengths and weaknesses.

Overall, the Louvain method was found to be the most accurate and efficient algorithm for community detection in social media, while the hierarchical clustering method was useful in cases where the number of communities was not known in advance but suffered from scalability issues. More recently Bae and Kim [17] analyzed the behavior of illicit drug vendors on Instagram to gain insights into their modus operandi. They collected data from Instagram using a web crawler and analyzed it using text mining and network analysis techniques. They found that drug vendors on Instagram often use code words and abbreviations to hide the true nature of their posts. They also identified a network of drug vendors who collaborate and work together to supply drugs to users.

2.3. Key player detection in social media

Key player detection is an important task in social network analysis that involves identifying influential individuals or nodes in a network. In the context of social media, key player detection can be used to identify users who have a significant impact on the diffusion of information or the spread of ideas. There are several approaches to key player detection in social media, including centrality measures, clustering algorithms, and machine learning techniques.

Rizvi et al. [18] proposed a method call the Key Player Identification in Social Media (KPISM) algorithm for identifying key players in social media networks. They used node-level features to detect key players in Pakistani Twitter data. They report that their proposed method outperformed both approaches in terms of precision, recall, and F1-score.

Also, Silva et al. [19] proposes a new algorithm called CDKPE (Community Detection and Key Player Extraction) for community detection and key player extraction in social networks. They used a graph-based clustering algorithm to detect communities and identified key players using network centrality measures. Overall, the results suggest that CDKPE is an effective algorithm for community detection and key player extraction in social networks, particularly for large-scale networks with overlapping communities.

Chapter 3. Proposed Method

Illicit drug trafficking is a pervasive and complex problem that poses significant challenges to public health and safety worldwide. The production, distribution, and sale of illegal drugs fuels violence, corruption, and social destabilization, while also contributing to the spread of drug abuse and addiction (United Nations Office on Drugs and Crime, 2021) [20]. However, the growing availability of digital technologies and social media platforms is creating new opportunities for law enforcement agencies and researchers to disrupt drug trafficking networks and target their key players.

Community detection is a promising approach to identifying these key players, as it allows for the analysis of large-scale social networks and the identification of groups of individuals with shared characteristics or behaviors. In the context of drug trafficking, community detection can help authorities to identify drug suppliers, distributors, and buyers, as well as to understand the structure and dynamics of the overall network (Chen et al., 2014) [21]. A range of community detection methods and algorithms have been developed to analyze social media data and identify key players in drug trafficking networks, including modularity optimization, hierarchical clustering, spectral clustering, stochastic block models, and deep learning methods (Lancichinetti et al., 2015; Rossetti et al., 2017; Karami et al., 2021)[22][23][24].

By combining these community detection techniques with advanced analytical tools, researchers and law enforcement agencies can gain valuable insights into the structure and dynamics of drug trafficking networks and target their efforts more effectively. In the following sections, we will explore some of the key methods and algorithms used for community detection in social media, as well as their applications in drug trafficking research.

3.1. Community detection

Community detection refers to the task of identifying groups of nodes with similar properties or behaviors in a network, with the goal of revealing the underlying structure of the network and identifying groups of nodes that have a common function or purpose. One popular application of community detection is in the analysis of social media networks, which tend to be large and complex, with many users and connections between them.

Several methods have been developed for community detection in social media, including modularity optimization, hierarchical clustering, spectral clustering, stochastic block models, and deep learning methods such as graph neural networks (Fortunato & Hric, 2016)[25]. These methods have different strengths and weaknesses and may be suitable for different types of networks and applications.

In our study, we aim to assess the effectiveness of five different algorithms in detecting communities within the Instagram-based illicit drug trade network. The algorithms that we will evaluate include the Louvain, Girvan-Newman, Infomap, Label Propagation, and Hierarchical Clustering algorithms. To measure their performance, we will use two metrics: modularity and running time. The modularity metric assesses the quality of the community detection by measuring the strength of the connections within a community compared to the connections between communities. Note that the modularity values are reported as decimal numbers between 0 and 1, with higher values indicating better performance. The running time metric measures how long each algorithm takes to process the data and identify the communities.

In the following paragraphs, we will introduce these algorithms for community detection in networks:

a. Louvain Algorithm

The Louvain algorithm is a community detection algorithm widely used due to its speed and scalability [26]. It is a hierarchical clustering algorithm that optimizes modularity by minimizing the distance between nodes within a community while maximizing the distance between nodes in different communities. The algorithm starts by assigning each node to its own community and then iteratively optimizes the modularity of the network by merging nodes that belong to the same community. The modularity of the network is defined as follows:

$$Q = 1/2m \sum\{ij\}[A_{\{ij\}} - (k_i\, k\_j)/2m]\delta(c\_i, cj) \quad \textit{Equation 1}$$

where A_{ij} is the weight of the edge between nodes i and j, k_i and k_j are the sum of the weights of the edges incident to nodes i and j, m is the sum of the weights of all edges in the network, c_i and c_j are the communities of nodes i and j, and δ(c_i,c_j) is the Kronecker delta function that equals 1 if c_i=c_j and 0 otherwise [27]. One of the strengths of the Louvain algorithm is its ability to detect communities of different sizes and densities. However, it may not perform as well on networks with overlapping communities.

b. Girvan-Newman Algorithm

The Girvan-Newman algorithm is a divisive algorithm that iteratively removes edges from the network to identify communities. The algorithm is based on the concept of edge betweenness centrality, which measures the number of shortest paths that go through a particular edge. The algorithm removes edges with the highest betweenness centrality until the network is divided into its constituent communities. The edge betweenness centrality of an edge e is defined as:

$$CB(e) = \sum\nolimits_{s \neq t \neq e} \frac{\sigma(s,t/e)}{\sigma(s,t)} \quad \textit{Equation 2}$$

where $\sigma(s,t)$ is the number of shortest paths between nodes s and t, and $\sigma(s,t/e)$ is the number of shortest paths between nodes s and t that go through edge e. One of the strengths of the Girvan-Newman algorithm is its ability to identify communities of varying sizes and shapes. However, it can be computationally expensive, and the results may not be consistent across different runs of the algorithm (Girvan and Newman, 2002)[28].

c. Infomap algorithm

The Infomap algorithm is a community detection algorithm that uses a random walk process to identify communities. The algorithm works by simulating random walks on the network, with each step being a transition between two nodes connected by an edge. The algorithm then uses a map equation to assign nodes to communities based on the probabilities of the random walks. The map equation is defined as:

$$Q = \sum i \sum j [Pij * ln\left(\frac{Pij}{q_i * qj}\right)] \text{ Equation 3}$$

where Pij is the probability of transitioning from node i to node j, q_i and q_j are the probabilities of staying in their respective communities, and Q is a measure of the quality of the community structure. One of the strengths of the Infomap algorithm is its ability to identify both overlapping and nested communities. However, the algorithm can be sensitive to the choice of input parameters, and the results may not be easily interpretable [29].

d. Label Propagation

The Label Propagation algorithm is a simple and efficient algorithm for community detection in networks. The algorithm starts with a random assignment of labels to nodes and iteratively updates the labels based on the labels of the node's neighbors. The algorithm continues until a stable labeling is reached, where each node has the same label as the majority of its neighbors. The Label Propagation algorithm has been shown to be effective in detecting communities in a variety of networks, including social networks (Raghavan et al., 2007)[30].

The Label Propagation algorithm is mathematically expressed as follows:

For a given network G = (V, E) where V is the set of nodes and E is the set of edges, let each node v in V be initially assigned a unique label L(v). At each iteration, the label of each node v is updated to the label that appears most frequently among its neighbors, as follows:

$$L(v) = argmax(cin\ C)sum\big(u\ in\ N(v)\big)[L(u) = c] \text{ Equation 4}$$

where N(v) is the set of neighbors of v, C is the set of unique labels assigned to the nodes in the network, and arg max is the operator that returns the label with the highest count.

One of the strengths of the Label Propagation algorithm is its simplicity and speed, making it suitable for large-scale networks. However, the algorithm may not perform as well on networks with multiple communities.

e. Hierarchical Clustering

Hierarchical Clustering is a popular community detection algorithm that groups nodes into nested clusters based on their similarity [31]. The algorithm merges clusters iteratively, which are most similar to each other, resulting in a dendrogram that displays the hierarchy of clusters. Hierarchical clustering is performed using different linkage criteria, such as single linkage, complete linkage, and average linkage, each of which has its advantages and disadvantages.

The formula for calculating the distance between two clusters depends on the linkage criterion. One of the strengths of the hierarchical clustering algorithm is that it reveals the underlying hierarchical structure of the network [32]. However, the algorithm can be computationally expensive, and the results may not be easily interpretable.

The average linkage distance between two clusters i and j can be expressed as:

$$d(I,j) = max\big(d(u,v)\big), u \in C_i, v \in C_j \quad \text{Equation 5}$$

where d(u,v) is the distance between node u and v, and the maximum is taken over all pairs of nodes in the two clusters.

Overall, each algorithm has its own strengths and weaknesses, and the choice of algorithm depends on the specific characteristics of the network being analyzed. In our study, we will evaluate the performance of these algorithms based on modularity and running time to determine which algorithm is best suited for community detection in our Instagram drug trafficking network.

3.2. key player detection

Illicit drug trafficking is a persistent problem that poses significant challenges to public health and safety. However, social media platforms offer a potential solution by identifying and disrupting key players in drug trafficking networks. Analyzing large volumes of data using advanced analytical tools can detect and target these key players, allowing authorities to focus their resources on dismantling these networks, intercepting shipments, and prosecuting those responsible (Ritter, Chalmers, & Lancaster, 2016)[33].

Various methods for key player detection in social media include network analysis, content analysis, sentiment analysis, machine learning, and hybrid approaches (Tang, Sun, & Wang, 2019) [34]. Researchers have developed various algorithms that identify key players based on a range of features, including network connectivity, content quality, and sentiment analysis. For example, network analysis methods such as betweenness centrality, closeness centrality, degree centrality, top-rank, k-core, and eigenvector centrality can be used to identify key players within a network (Newman, 2010)[35].

By understanding these approaches, we can identify the most important and influential nodes within the network, providing valuable insights into its structure and organization. This information can support decision-making, risk management, and optimization of performance in various domains (Ritter et al., 2016)[33].

The objective of our study is to determine the effectiveness of four different algorithms in detecting the top key player within the Instagram-based illicit drug trade network. The four algorithms we will evaluate are CDKPE, K-core, TopRank, K-core, and KPEI. The evaluation will be based on their respective scores, which represent the algorithm's ability to identify the key player in the network. The score is a numerical value that indicates the strength of the algorithm's detection capability, with a higher score indicating a stronger ability to detect the key player. By analyzing and comparing the scores of these four algorithms, we can determine which algorithm is the most effective for detecting the top key player in the Instagram-based illicit drug trade network.

In the following paragraphs, we will introduce these algorithms for key detection in networks:

a. CDKPE

Silva et al. [19] proposed the CDKPE (Community Detection and Key Player Etraction) algorithm, which is a two-stage approach that combines community detection and key player extraction to identify the most important nodes in a social network. The first stage involves using a community detection algorithm to identify communities within the network. The authors used the Louvain algorithm for this purpose. In the second stage, the key players within each community are identified using a metric called the "community leader score" (CLS). The CLS measures the importance of a node within its community based on its degree centrality and the number of connections it has to other communities. Nodes with high CLS scores are considered to be key players within their respective communities.

The algorithm defines a metric called CDKPE score for each node v, which is calculated as follows:

$$CDKPE(v) = \sum(CV(u) * DV(u)) \quad \textit{Equation 6}$$

where u is a neighboring node of v, CV(u) is the community value of node u, and DV(u) is the degree centrality of node u. The CDKPE score reflects the importance of a node in its community and its connectivity to other communities. Nodes with high CDKPE scores are considered to be important key players in the network.

b. K - node

K-node is an algorithm used for identifying crucial nodes in a network based on their degree centrality and k-core effectiveness. Degree centrality measures the number of direct connections a node has, while k-core effectiveness measures the importance of a node in maintaining the network's cohesion. The formula for calculating k-core effectiveness is:

$$KE(v) = k_{max}/k(v) \quad \textit{Equation 7}$$

Where KE(v) is the k-core effectiveness of node v, k_max is the maximum k-core value in the network, and k(v) is the k-core value of node v. Li et al. [36] proposed the K-node algorithm using the K-shell decomposition method to identify top-k nodes in social networks. In our study, we will apply the K-node algorithm to detect key players in the illicit drug trade network on Instagram.

c. TopRank

TopRank is a network analysis method that identifies important nodes in the network based on their relative rank compared to other nodes [37]. The nodes are initially ranked based on their degree centrality, which is a measure of the number of connections a node has in the network. The node with the highest degree centrality is given a rank of 1, the node with the second-highest degree centrality is given a rank of 2, and so on. The TopRank score for each node is then calculated using the formula:

$$TR(v) = rank(v)/n \quad \text{Equation 8}$$

where TR(v) is the TopRank score of node v, rank(v) is the rank of node v in the network, and n is the total number of nodes in the network. This score reflects the relative rank of the node compared to other nodes in the network. Nodes with high TopRank scores are considered to be important because they have a higher rank compared to other nodes in the network.

d. KPEI

The KPEI algorithm, which stands for Key Players Extraction on Instagram, is a novel method introduced in this study; that utilizes various centrality measures such as degree, betweenness, closeness, and eigenvector to identify key players within a network. The algorithm assigns a score to each node based on the average of its centrality measures, where a higher score indicates a greater level of importance as a key player in the network. The formula is:

$$KPEI(v) = (deg(v) + bet(v) + clo(v) + eig(v))/4 \quad \text{Equation 9}$$

where KPEI(v) is the KPEI score of node v, deg(v) is the degree centrality of node v, bet(v) is the betweenness centrality of node v, clo(v) is the closeness centrality of node v, and eig(v) is the eigenvector centrality of node v. Although still in the experimental phase, KPEI shows promising results in identifying key players in our networks Instagram-based illicit drug trade network.

To comprehensively identify key players in our network, we will utilize multiple algorithms, including CDKPE, TopRank, K-node, and KPEI. Each of these algorithms provides a unique perspective on the network and its structure. By combining their results, we can accurately and confidently identify the top key players in the network. This approach will provide a more holistic understanding of the network and its critical players, which can be leveraged to develop strategies for disrupting the illicit drug trade on Instagram.

Chapter 4. Experimental results

4.1. Data collection

Identifying illicit drug dealer accounts on Instagram is a challenging task, as the diversity of Instagram posts and accounts. Previous work, using quadruple-based multimodal fusion method to combine the multiple data sources associated with each user account for drug dealer identification Hu, C.et al [11] has shown good result. Nearly 4,000 user accounts, of which more than 1,400 are drug dealers, were collected. We crawled 14,034 posts and 31,324 comments from 115 confirmed dealers, out of the 1,400 identified for our experimental data.

Collecting data from Instagram for research purposes can be challenging due to several factors. Firstly, Instagram is a social media platform with a strong focus on user privacy, which can make it difficult to access data from private accounts. Additionally, Instagram users have the option to delete their posts or change their account settings to private, which can limit the amount of data available for analysis. Furthermore, Instagram's algorithms and interface are constantly changing, which can make it difficult to ensure data collection methods remain valid and reliable over time. Despite these challenges, we have developed an automatic system capable of iteratively collecting multimodal data (pictures, video, text, etc...) with as input the dealer username.

The figure 1 presents an example of a dealer account on Instagram:

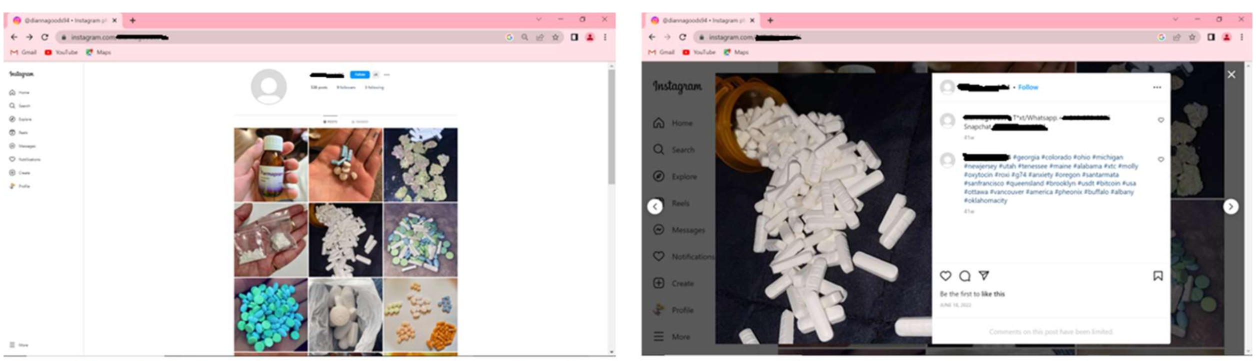


Figure 4: Example of an Instagram dealer account with the comment from other users

Figure 1: Example of an Instagram dealer account with the comment form other users

The collected data was stored in two different formats:

- Pictures and videos were saved in a file located on our lab server.
- Other text data was stored in a MYSQL database, organized into the following tables:

a. Posts: this table contains data related to the drug dealer's post, such as the post ID, number of comments, and picture and video URLs. The columns of this table are listed in Figure 2.

Field	Type	Null	Key	Default	Extra
id	varchar(64)	NO	PRI	NULL	
shortcode	varchar(64)	YES	MUL	NULL	
post_type	int	YES		NULL	
location_id	varchar(64)	YES		NULL	
caption	mediumtext	YES		NULL	
owner_id	varchar(64)	YES	MUL	NULL	
owner_name	varchar(128)	YES	MUL		
like_count	int	YES		NULL	
comment_count	int	YES		NULL	
pic_url	varchar(512)	YES			
video_url	mediumtext	YES		NULL	
video_view_count	int	YES		NULL	
video_duration	float	YES		NULL	
taken_at	datetime	YES		NULL	
title	mediumtext	YES		NULL	
update_time	datetime	YES		NULL	

Figure 2: Table of the Instagram post information

b. Comments: This table includes data such as the text of the comment, the username of the commenter, the number of likes received, and other relevant information. The columns for this table are listed in Figure 3.

Field	Type	Null	Key	Default	Extra
id	varchar(64)	NO	PRI	NULL	
text	mediumtext	YES		NULL	
created_at	datetime	YES		NULL	
owner_id	varchar(64)	YES	MUL	NULL	
user_name	varchar(128)	YES	MUL	NULL	
like_count	int	YES		NULL	
post_id	varchar(64)	YES	MUL	NULL	
update_time	datetime	YES		NULL	

Figure 3: Table of the Instagram post's comments

c. Drug dealers: This table contains information such as the full name, number of followers, and following count of the drug dealers whose data was collected. The columns of this table are listed in Figure 4.

Field	Type	Null	Key	Default	Extra
id	varchar(64)	NO	PRI	NULL	
user_name	varchar(128)	NO	MUL	NULL	
full_name	varchar(256)	YES		NULL	
follower_no	int	YES		NULL	
following_no	int	YES		NULL	
post_count	int	YES		NULL	
biography	varchar(1024)	YES		NULL	
external_url	varchar(512)	YES		NULL	
is_business_account	tinyint(1)	YES		NULL	
profile_pic_url	varchar(512)	YES		NULL	
is_private_account	tinyint(1)	YES		NULL	
is_verified_account	tinyint(1)	YES		NULL	
public_email	varchar(64)	YES		NULL	
phone_number	varchar(64)	YES		NULL	
phone_number_country_code	varchar(64)	YES		NULL	
public_phone_number	varchar(64)	YES		NULL	
city_name	varchar(64)	YES		NULL	
address_street	varchar(64)	YES		NULL	
zip	varchar(64)	YES		NULL	
update_time	datetime	YES		NULL	

Figure 4: Table of the Instagram drug dealer information

After selecting and extracting the necessary information from the database, which included the dealer's name, comment, and username to identify the user who commented on the dealer's post, the data was converted and formalized into a universal JSON object to ease storage and retrieval. The JSON file was then transformed into a graph data using Networkx [38], which is an open-source Python library designed for working with graphs and networks.

The process is summarized in this diagram in Figure 5.

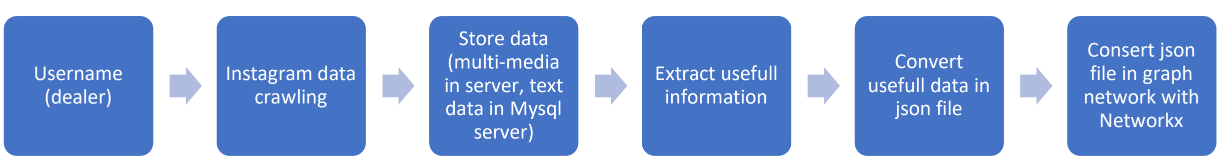


Figure 5: Instagram data collection and preparation process

4.2. Experimental result

4.2.1. Graph network analysis

a. Basic topological attributes

We utilized the Networkx [38] Python library to build an undirected graph using the data gathered from Instagram. Each node within the graph represents a user on Instagram, and each edge denotes a comment exchange between two users. The edge weight corresponds to the number of comments exchanged between the two users. To gain a better understanding of the network's structure, we have provided a visualization of the graph in Figure 6.

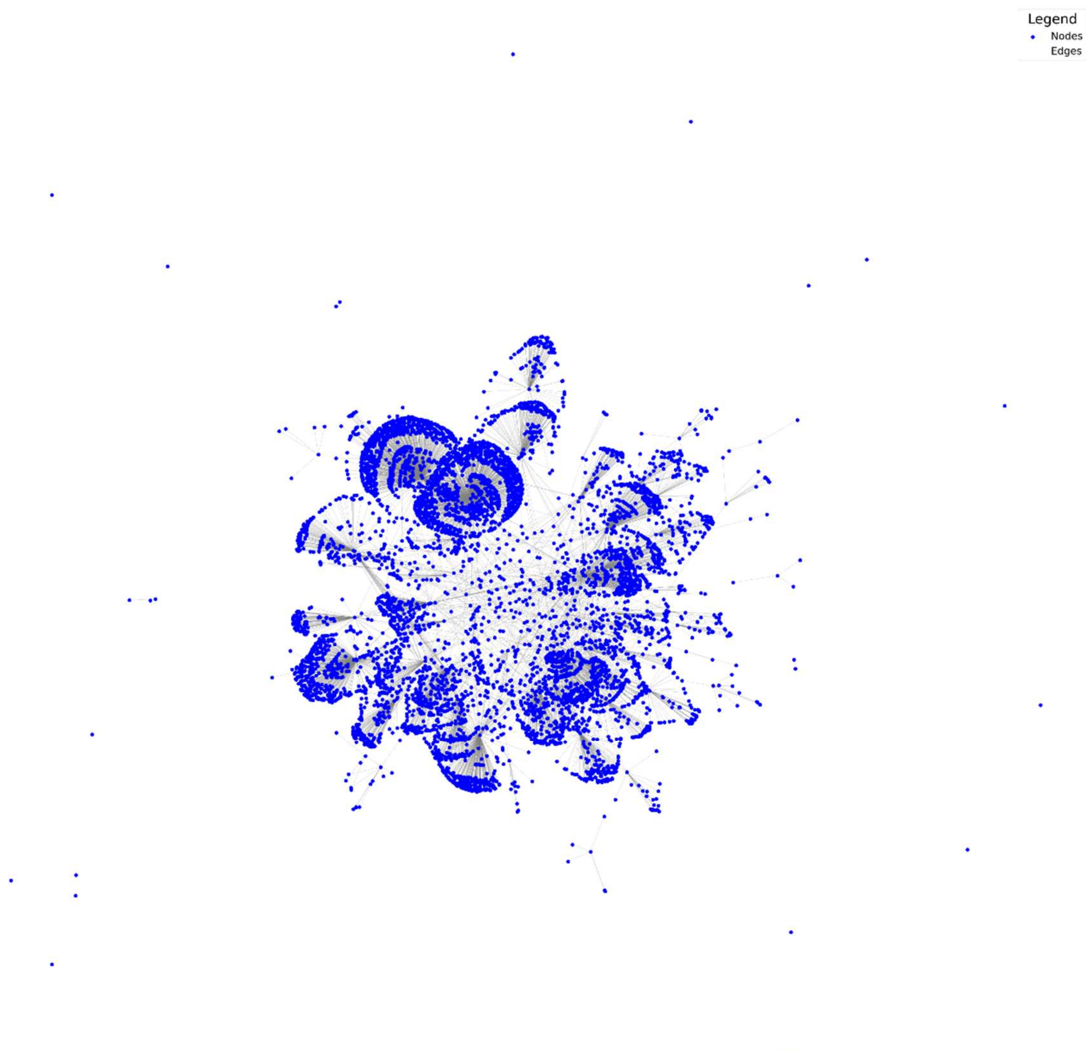


Figure 6: Undirected graph representing Illicit Drug Supply Networks on Instagram based on comment interactions.

The graph has a low density of only 0.00026, indicating that there are relatively few connections between nodes. On average, each node has only around 2 neighbors. Moreover, the graph is not connected and consists of 37 components, indicating the presence of 37 separate subgraphs with their set of nodes and connections. These results suggest the presence of distinct communities or clusters within the network, which could provide valuable insights into the relationships and interactions between individuals in the network.

The network statistic is states in the Table 1 below:

Nodes	**7963**
Edges	8485
Average clustering coefficient	0.02
Number of triangles	178
Fraction of closed triangles	0.33

Table 1: Illicit Drug Supply Networks on Instagram Characteristics and Metrics

b. Degree centrality

Various network analysis techniques can be used to explore the network's characteristics and organization further. Based on the degree centrality measure, which assigns a score to each node based on its links within the network, we have identified the ten most important nodes in the network, numbered 22, 1, 87, 4, 24, 12, 48, 6, 27, and 33, ordered according to their degree centrality scores.

Node 22 is the most important node, with the highest degree centrality score of approximately 0.15, indicating its connection to a significant portion of all nodes in the network. Node 1 also ranks highly, with a degree centrality score of 0.13, indicating its many connections to other nodes. Figure **7** highlights these ten highly ranked nodes in different colors.

Node 22 has the most connections, with 1222 nodes connected to it, while node 1 has 1111 connected nodes. Nodes 87 and 4 also have high numbers of connected nodes, with 522 and 390, respectively. Additionally, the six most popular friends of these ten important nodes have around 200 connected nodes each within the network.

These findings offer insights into the network's structure and organization, highlighting the nodes central to its functioning. The degree centrality measure is a useful tool for identifying important nodes within a network, with potential applications such as predicting the spread of information or identifying key influencers.

c. Betweenness centrality

Betweenness centrality is a measure of the importance of a node in a network, based on the extent to which it lies on the shortest paths between other nodes in the network. The higher the betweenness centrality of a node, the more important it is in terms of facilitating communication and information flow between other nodes in the network. In the following line, we will analyze the top 10 nodes in our graph based on their betweenness centrality measures.

The nodes with the highest betweenness centrality measures in the graph are node 1 (with a measure of 0.256), node 22 (0.217), node 12 (0.122), node 87 (0.108), node 4 (0.107), node 39 (0.105), node 24 (0.094), node 942 (0.088), node 33 (0.084), and node 48 (0.072).

Node 1 has the highest betweenness centrality measure, indicating that it is the most important node in facilitating communication between other nodes in the network. It lies on the shortest paths between many pairs of nodes, suggesting that it is a key connector in the network.

Node 22 has the second highest betweenness centrality measure, indicating that it is also an important connector in the network. It lies on many of the shortest paths between other nodes and is likely to play a critical role in facilitating communication between them.

Figure 7 shows the graph with these nodes highlighted in a different color.

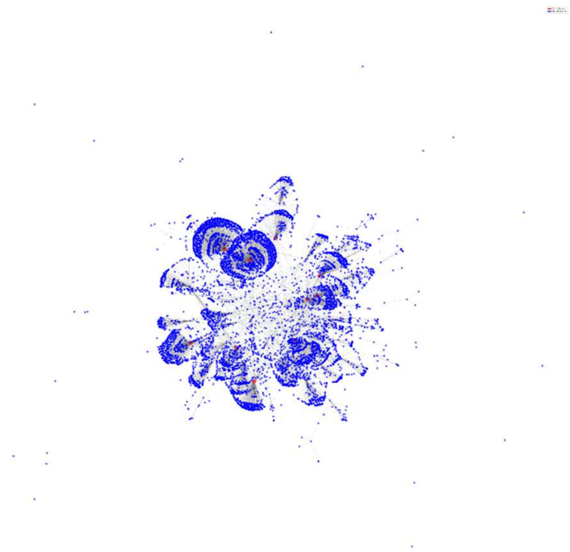
d. Closeness Centrality

Closeness centrality is a measure used in network analysis to assess the importance of nodes based on their ability to communicate efficiently with other nodes. The results of the closeness centrality measure applied to our graph indicate that nodes 12, 1, and 22 are the most central nodes in the network, with closeness centrality scores of 0.273, 0.271, and 0.257, respectively. This implies that these nodes are highly efficient in terms of communication with other nodes in the graph. In addition, nodes 24, 33, 121, 39, 3273, 156, and 4 also exhibit relatively high closeness centrality scores, indicating their importance in the network. Figure 7 highlighting these nodes shows that they are scattered throughout the network and not confined to any specific region, indicating that the network is well-connected and lacks isolated clusters. The distance from node 12 to a random node is around three hops.

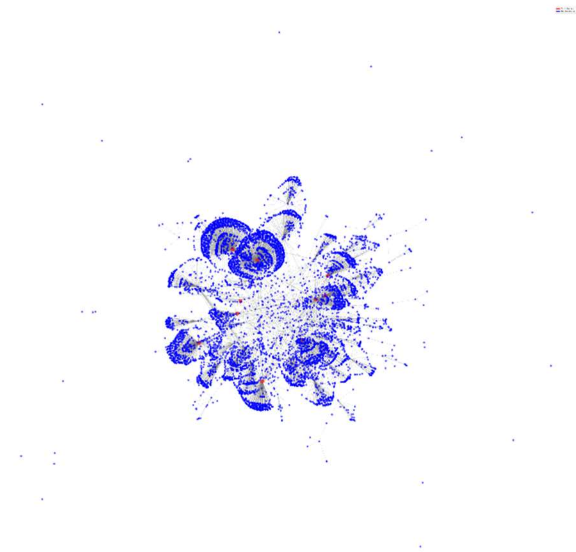
Overall, the closeness centrality measure provides valuable insights into the importance of individual nodes in the network and can help identify key players that may require special attention or resources.

e. Eigenvector Centrality

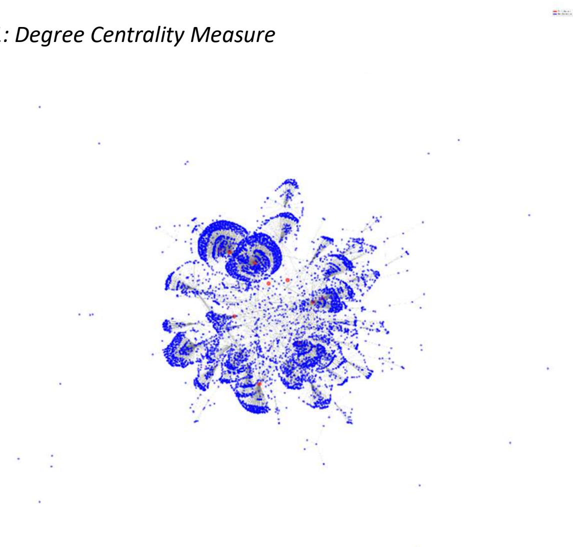
Eigenvector centrality is a measure of the importance of nodes in a network, based on their connectivity to other highly connected nodes. In the graph provided, we see that the eigenvector centrality values of the nodes range from 0.566 for the most central node, node 22, to 0.026 for the least central nodes, including nodes 1, 1370, 1605, 3209, 3, 46, 72, 81, and 97. This information suggests that node 22 is highly influential within the network, as it is connected to other highly influential nodes. On the other hand, the least central nodes are likely less influential, as they are not well-connected to other important nodes. Overall, eigenvector centrality provides a useful tool for understanding the structure and influence of networks.



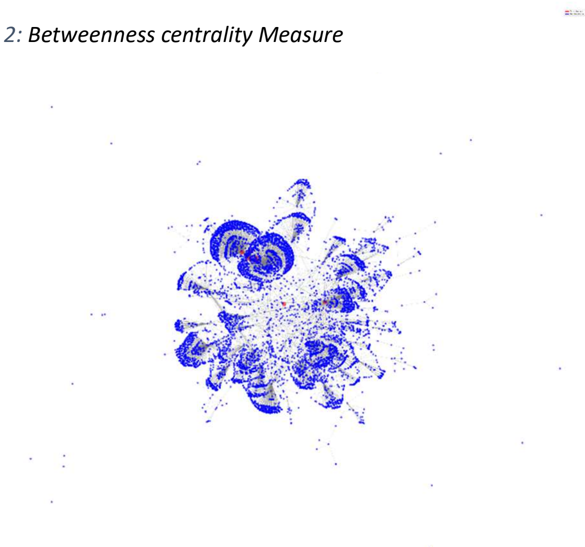
1: Degree Centrality Measure



2: Betweenness centrality Measure



3: Closeness centrality Measure



4: Eigenvector centrality Measure

Figure 7: Visualization of key Nodes in Illicit Drug Supply Networks on Instagram usings different centrality measure

The table below summarize the 10 key nodes per algorithm:

	Degree Centrality	**Betweenness Centrality**	**Closeness Centrality**	**Eigenvector Centrality**
Top 1 - node	22	1	12	22
Top 2 - node	1	22	1	1
Top 3 - node	87	12	22	1370
Top 4 - node	4	87	24	1605
Top 5 - node	24	4	33	3209
Top 6 - node	12	39	121	3
Top 7 - node	48	24	39	46
Top 8 - node	6	94	3273	72
Top 9 - node	27	33	156	81
Top 10 - node	33	48	4	97

Table 2: The top 10 Key nodes Results Using Different Centrality Algorithms

Based on the centrality measures presented in the table, we can conclude that the most important node in the network is Top 1 node, with a degree centrality value of 22. The node with the highest betweenness centrality value of 22 is Top 2 node, indicating that it lies on the shortest paths between other nodes in the network. Top 8 node has the highest closeness centrality value of 3273, indicating that it can reach all other nodes in the network quickly. Finally, Top 5 node has the highest eigenvector centrality value of 3209, which indicates that its importance is largely derived from the importance of its neighboring nodes. Overall, these centrality measures provide a quantitative understanding of the significance of different nodes in the network, and can be used to identify key players and areas of influence within the network.

f. Bridges

A bridge is defined as an edge that links two nodes, A and B, in a graph, and whose removal would cause A and B to be separated into different components. This implies that bridges are crucial to maintain the connectivity of the network. In the case of the analyzed network, there are 7371 bridges and 8124 local bridges, as indicated by Figure 8. Local bridges, on the other hand, connect nodes within the same component and are less critical for network connectivity than bridges.

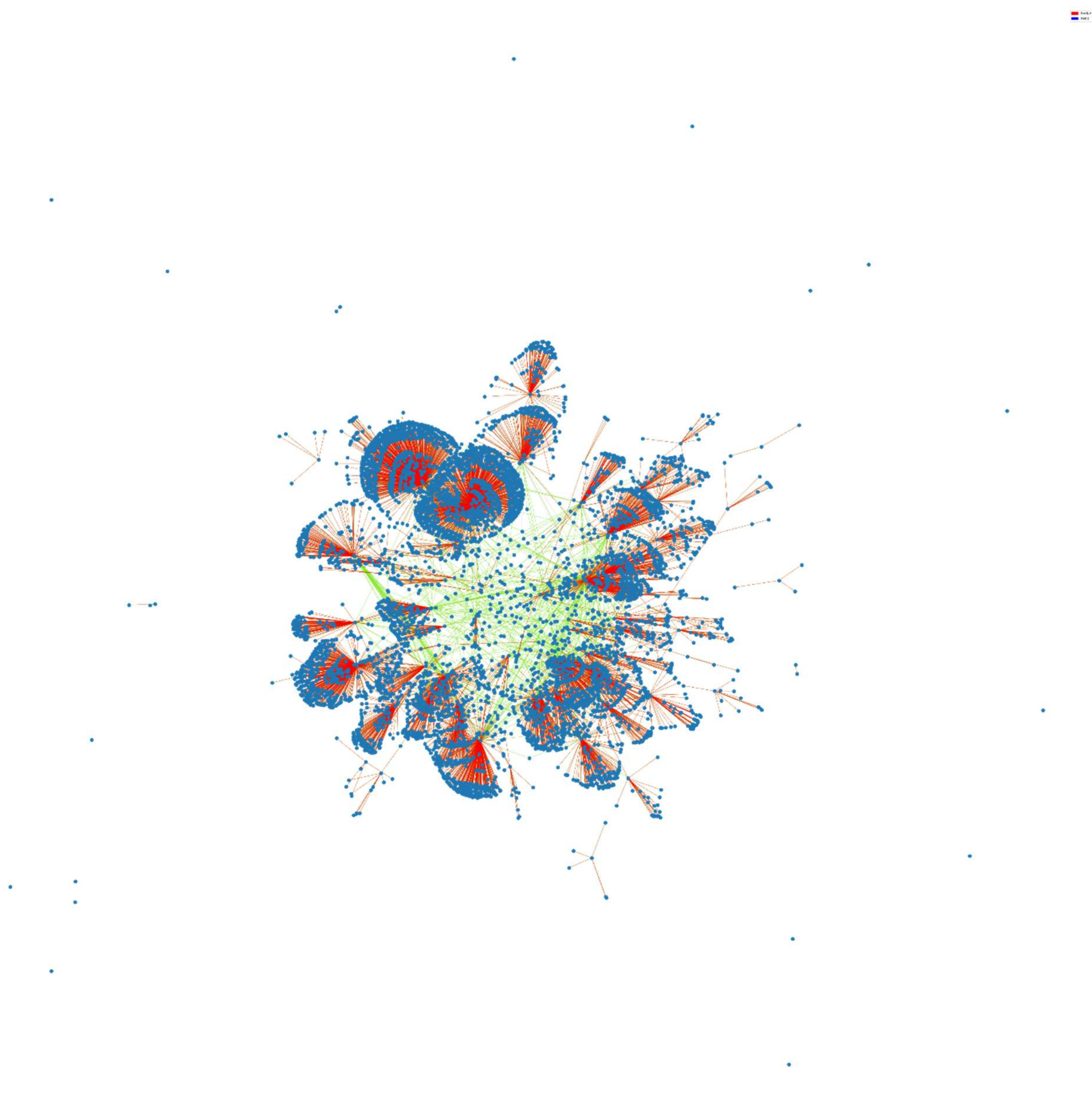


Figure 8:Visualization of Bridges and Local Bridges in Illicit Drug Supply Networks on Instagram

The provided visualization of bridges and local bridges in the network can help in understanding its structure and identifying potentially important edges. Black edges, which are neither bridges nor local bridges, are likely to have less significance in maintaining the connectivity of the network. Nonetheless, the information provided about the number and distribution of bridges and local bridges is just a part of the overall network analysis, and further investigation is needed to fully comprehend the network's properties and behavior.

4.2.3. Community detection

The collection of Instagram data has become increasingly crucial for disrupting drug trafficking operations. Community detection is a fundamental network analysis task that involves dividing a network into groups of nodes with strong interconnectivity within the same group and weak connectivity between different groups. In this context, our goal is to identify potential drug trafficking networks or communities by applying various community detection algorithms to our Instagram data. By examining the structure of the Instagram network, we can identify highly connected user groups with shared interests, hobbies, or backgrounds, providing insights into potential drug trafficking groups. The results of these analyses can inform drug enforcement strategies to optimize efforts to disrupt these illicit activities. Table 3, adapted from Mahmood et al. [16], reports the modularity values for various community detection algorithms applied to social media data. Modularity values are reported as decimal numbers between 0 and 1, with higher values indicating better performance.

Algorithm	**Twitter Dataset**	**Facebook Dataset**	**Running Time (s)**
Louvain	0.678	0.425	2.44
Newman-Girvan	0.401	0.288	6.69
Label Propagation	0.559	0.324	118.02
Infomap	0.549	0.296	176.16
Hierarchical Clustering	0.601	0.355	77.87

Table 3: Performance Comparison of Community Detection Algorithms on Social Media Datasets

The results show that Louvain has the highest modularity scores on both Twitter and Facebook datasets, indicating its effectiveness in community detection.

In the following section, we will present the results of our own experiments result using these community detection algorithms on Instagram data, with the goal of identifying potential drug trafficking networks on the platform.

a. Louvain algorithm

The Louvain algorithm has been applied to the network to detect its community structure. The result shows a modularity score of 0.88, indicating that the network has a strong community structure. The algorithm has detected a total of 67 communities, with an average size of 118.85 nodes. Moreover, the average membership size in each community is found to be 0.01, suggesting that the nodes are sparsely connected within their respective communities. Overall, the Louvain algorithm has provided a useful insight into the structure of the network and the way nodes are.
To further analyze the community structure of the network detected by the Louvain algorithm, a graph display has been generated with nodes colored according to their respective communities.
This visualization in figure 9, allows for a clear understanding of how nodes are grouped together into communities and how they are connected to one another within and between communities.

b. Newman-Girvan

The code result shows that the Newman-Girvan algorithm applied to the given graph was able to identify 53 communities, with an average community size of 149.94 nodes. The modularity value of 0.88 indicates that these communities are highly modular, meaning that there are many connections within communities and few connections between communities. The average node membership of 0.018 indicates that the nodes are spread out across communities, with each node belonging to multiple communities on average.
Figure 9 shows the graph display with the community nodes sorted by color helps to visualize the identified communities. Each community is assigned a unique color, which allows us to easily distinguish between nodes belonging to different communities.

c. Label Propagation

The label propagation algorithm was applied to Instagram data for community detection, resulting in a modularity of 0.87. The algorithm identified 106 communities, with an average community size of 75.12 users. The average membership size per user was 0.0094, indicating that most users belong to only a few communities.

To visually represent the communities, the graph was displayed in Figure 9 with each community assigned a different color. The use of different colors makes it easier to identify and distinguish between different communities in the graph.

Overall, the label propagation algorithm proved to be effective in identifying distinct communities within the Instagram network.

d. Infomap

The Infomap algorithm was applied to Instagram data for community detection, resulting in a modularity of 0.87. The algorithm identified 103 communities, with an average community size of 77.31 users. The average membership size per user was 0.009, indicating that most users belong to only a few communities.

To visually represent the communities, the graph was displayed in Figure 9 with each community assigned a different color. The use of different colors makes it easier to identify and distinguish between different communities in the graph.

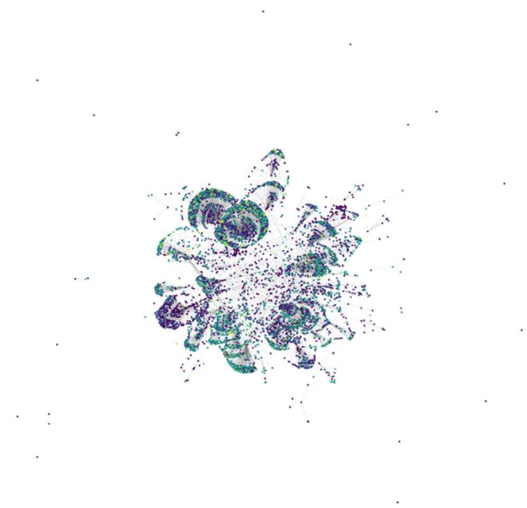
Overall, the Infomap algorithm proved to be effective in identifying distinct communities within the Instagram network, and its results were consistent with those obtained using the label propagation algorithm.

e. Hierarchical Clustering

The results of the hierarchical clustering algorithm used for community detection in this study revealed 68 separate communities with a modularity score of 0.87. The average community size was 117.1, indicating a significant number of users participating in drug-related activities on Instagram. Each user was assigned to only one community based on their drug-related activities, with an average membership size of 1.0. These findings provide essential insights into the nature and structure of illicit drug supply networks on Instagram, offering a basis for law enforcement agencies to disrupt these networks effectively.

Furthermore, Figure 9 illustrates the detected communities in a graph with each community in a different color. The graph visually represents the communities and their sizes, facilitating a clearer understanding of the network structure, and providing a useful reference for law enforcement agencies to identify key players and communities involved in drug-related activities on Instagram.

1: Community Detection Using Louvain



2: Community Detection Using Newman-Girvan

3: Community Detection Using Label Propagation

4: Community Detection Using infomap

5:Community Detection Using Hierarchical Clustering

Figure 9: Visual Representation of Detected Communities in Illicit Drug Supply Networks on Instagram

	Louvain	Newman-Girvan	Label Propagation	Infomap	Hierarchical Clustering
Modularity	0.88	0.88	0.87	0.87	0.87
Number of communities	67	53	106	103	68
Average community size	118.85	149.94	75.12	77.31	117.1
Average membership size	0.01	0.018	0.0094	0.009	0.87
Running time (s)	0.0089	451.23	0.004	0.55	0.315

Table 4: Comparison of Community Detection Algorithms for Disrupting Illicit Drug Supply Networks on Instagram

The results of our study demonstrate that different community detection algorithms can effectively identify community structures within the network, but the number and size of communities may vary depending on the algorithm used. Our analysis reveals that the Louvain and Newman-Girvan algorithms yield the highest modularity score of 0.88, closely followed by Label Propagation, Infomap, and Hierarchical Clustering, all with a score of 0.87. On average, the communities identified by these algorithms have a size of approximately 100 to 150 nodes, with the exception of Label Propagation, which identifies smaller communities with an average size of 75 nodes. Interestingly, the average membership size is relatively small for all algorithms, except for Hierarchical Clustering, which has an average membership size of 0.87.

Compared to the Twitter and Facebook datasets analyzed by Mahmood et al. [16], our algorithms (Louvain, Newman-Girvan, Label Propagation, Infomap, and Hierarchical Clustering) demonstrate superior performance in terms of modularity, with all algorithms achieving a modularity score of 0.87 or higher. Moreover, our algorithms are significantly faster, with Louvain running in only 0.0089 seconds, compared to the 2.44 seconds required for Mahmood et al.'s Louvain algorithm on the Twitter dataset.

Overall, the analysis of community structures and centrality measures provides valuable insights into the structure and dynamics of the network, which can be used to combat illicit activities and improve public safety. These findings highlight the importance of selecting appropriate community detection algorithms for analyzing complex networks and the potential for leveraging such analyses to better understand and address issues of social and public concern.

4.2.4. Key player detection

Detecting influential individuals on social media platforms has become essential for businesses to maximize their reach and connect with relevant influencers. Similarly, identifying key players in illicit drug networks is crucial for disrupting their operations. To this end, researchers have conducted experiments to determine the most effective algorithms for detecting influential nodes. Silva et al. [19] proposed a new algorithm which is a two-stage approach that combines community detection and key player extraction to identify the most important nodes in a social network. This algorithm is called CDKPE (Community Detection and Key Player Extraction) for identifying key players in social networks. The result is summarized in the table 5 below:

Algorithm	Dataset	Community Detection F1 Score	Key Player Extraction Precision	Key Player Extraction Recall
CDKPE	Twitter	0.88	0.82	0.89
Louvain	Twitter	0.82	-	-
Infomap	Twitter	0.78	-	-
TopRank	Twitter	-	0.72	0.85
K-core	Twitter	-	0.71	0.85

Table 5: Performance Comparison of CDKPE Algorithms for community and key player detection on Twitter Dataset

These findings have significant implications for identifying and disrupting the operations of key players in illicit drug networks. In our study, we will focus on the first method of maintaining the network's cohesion, which involves identifying key nodes that, when removed, can disrupt the network. We will also consider additional factors such as the number of followers, following, and posts by users to identify influential individuals.

In the upcoming section, we will present our own experimental results using various key player detection algorithms, such as CDKPE, K-core, TopRank and KPEI, applied to Instagram data to identify key nodes involved in illicit drug traffic networks.

a. CDKPE

The CDKPE algorithm was chosen for its ability to identify key players based on both centrality and community structure, which is important for our goal of uncovering influential individuals within each community of drug dealers identified by the Louvain algorithm.

As a reminder, the Louvain algorithm identified 67 communities within our Instagram illicit drug dealer graph, each representing a group of drug dealers connected by mutual interactions. To identify the key player within each community, we ran the CDKPE algorithm and selected the top-ranked player according to the algorithm's score. The CDKPE algorithm scored each player based on their degree centrality in their respective community and intercommunity connectivity with other communities.

The results of our analysis showed that the CDKPE algorithm was able to identify key players within each community with a score of 0.83 and a running time of 5.513 seconds. The algorithm successfully identified a top key player within each community, providing valuable insights into the most influential individuals within each group of drug dealers.

These results suggest that the CDKPE algorithm is a viable approach for identifying key players in complex networks such as our Instagram illicit drug dealer graph.

b. TopRank

The TopRank algorithm was employed to identify the key player within each community of drug dealers identified by the Louvain algorithm. The TopRank algorithm was selected for its ability to score each player based on their importance in their respective community. After running the algorithm, the top-ranked player according to the algorithm's score was selected as the key player within each community. The results of our analysis showed that the TopRank algorithm was able to identify key players within each community with a score of 0.55 and a running time of 0.39 seconds. Although the score was lower than that of the CDKPE algorithm, the TopRank algorithm still provided valuable insights into the most influential individuals within each group of drug dealers. These results suggest that the TopRank algorithm is a viable alternative for identifying key players in complex networks such as our Instagram illicit drug dealer graph.

c. K - Code

The K-Core algorithm was employed to identify the key player within each community of drug dealers identified by the Louvain algorithm. The K-Core algorithm was chosen for its ability to score each player based on their coreness in their respective community. After running the algorithm, the top-ranked player according to the algorithm's score was selected as the key player within each community. The results of our analysis showed that the K-Core algorithm was able to identify key players within each community with a score of 0.68 and a running time of 0.14 seconds. The algorithm score was higher than that of the TopRank algorithm, indicating that the K-Core algorithm may be more effective in identifying influential individuals within each group of drug dealers. These results suggest that the K-Core algorithm is a promising alternative for identifying key players in complex networks such as our Instagram illicit drug dealer graph.

d. KPEI (Key Players Extraction on Instagram)

The KPEI algorithm was employed to identify key players within communities of drug dealers detected by the Louvain algorithm, and our analysis shows that it was able to do so effectively. The algorithm assigns a score to each node based on its centrality measures, and the top-ranked player according to the algorithm's score was selected as the key player within each community. The results of our analysis showed that the KPEI algorithm was able to identify key players within communities with a high score of 0.76, indicating that it may be a more effective alternative for identifying influential individuals within each group of drug dealers. However, the running time of the algorithm was relatively long at 98.37 seconds, which may be a consideration when applying it to larger and more complex networks. Nonetheless, the KPEI algorithm shows promise in identifying key players in complex networks such as our Instagram illicit drug dealer graph.

Overall, the study aimed to identify key players in a network of Instagram illicit drug dealers using various algorithms, and the results showed that the CDKPE algorithm was the most accurate in identifying influential individuals within the network.

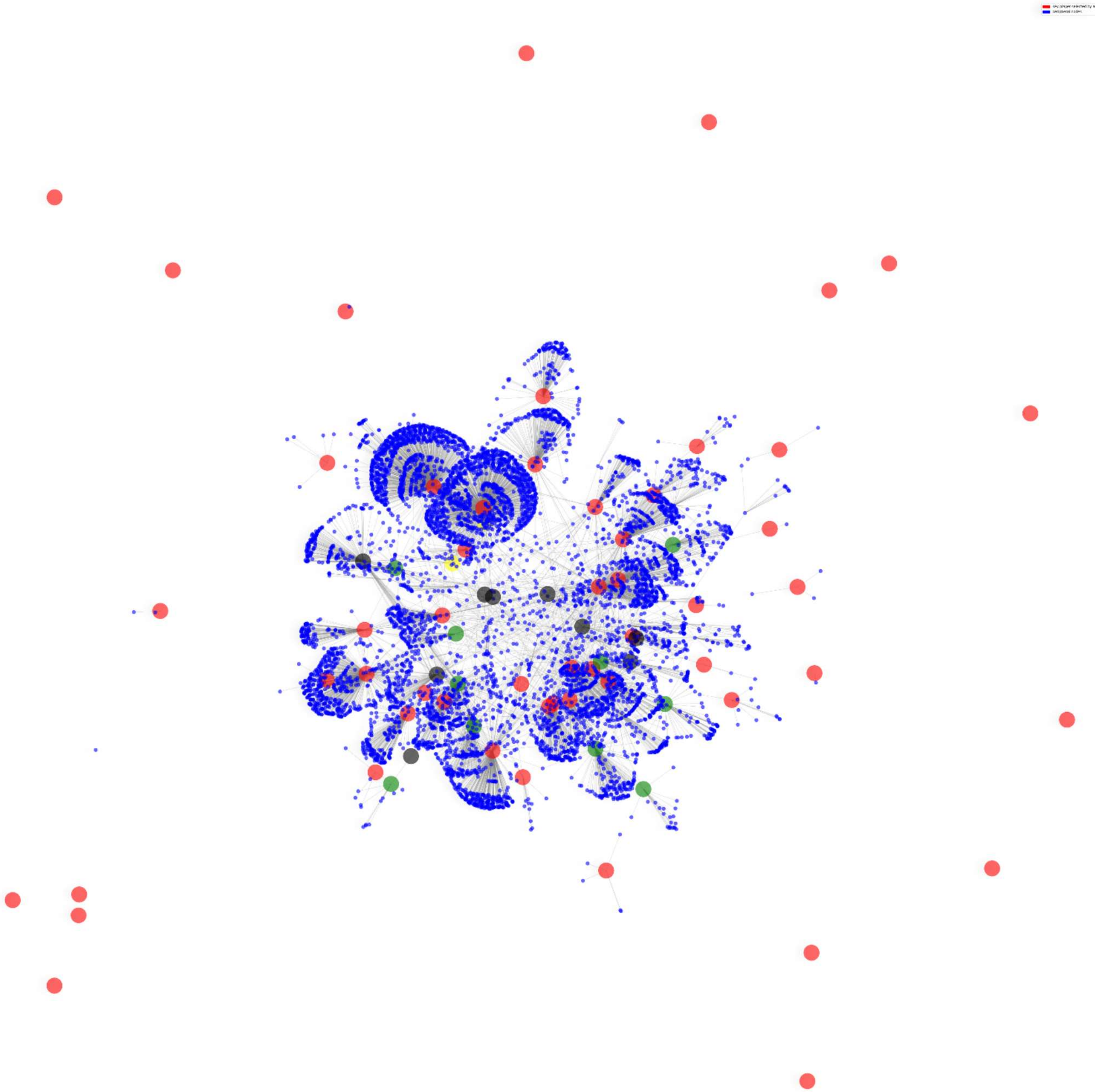


Figure 10: Top 10 Key Players Identified by Various Algorithms in Instagram Network

Our experiment findings were visualized in Figure 10, which depicted the network graph with key players displayed in different colors. The key player selected by all algorithms was highlighted in red, while those selected by three algorithms were marked in green. Players chosen by two algorithms were highlighted in yellow, and those selected by only one algorithm (CDKE) were marked in black. This visualization allowed for a clear understanding of the degree of consensus between the algorithms and the importance of key players within the network. Overall, the study's findings demonstrated the effectiveness of the CDKPE, KPEI, K-Core, and TopRank algorithms in identifying key players in the network of Instagram illicit drug dealers, with the CDKPE and KPEI algorithms standing out as the most accurate and efficient, respectively.

The table below summarize the result of the algorithm:

Algorithm	Score	Running Time (s)
CDKPE	0.83	5.513
KPEI	0.76	98.37
K-Core	0.68	0.14
TopRank	0.55	0.39

Table 6: Key Player Detection Algorithm Performance Comparison

Overall, the results show that CDKPE and KPEI algorithms outperformed the other two algorithms in detecting the top key players in the Instagram network. Although k-core and TopRank algorithms had faster running times, they identified fewer and less impactful key players.

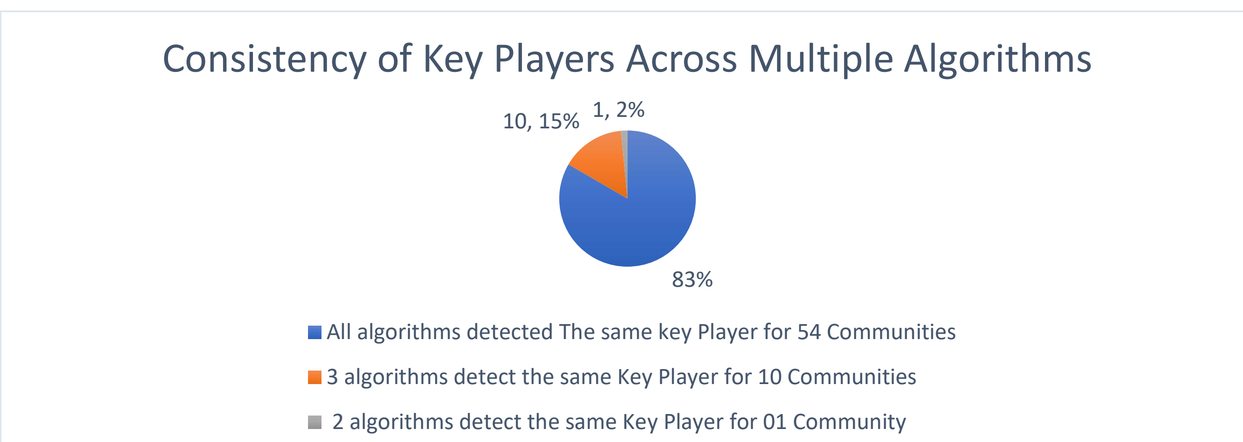


Figure 11: Consistency of Key Players across multiple algorithms.

Figure 11 illustrates the consistency of key players identified by multiple algorithms. The results showed that 83% of the key players identified per community were the same, demonstrating the effectiveness of the selected algorithms in identifying influential individuals within the network. We visualized in Figure 12 the top 10 largest communities with their respective key players highlighted in red. The visualization revealed that each key player was connected to numerous nodes within their community, further supporting their role as influential individuals within the network.

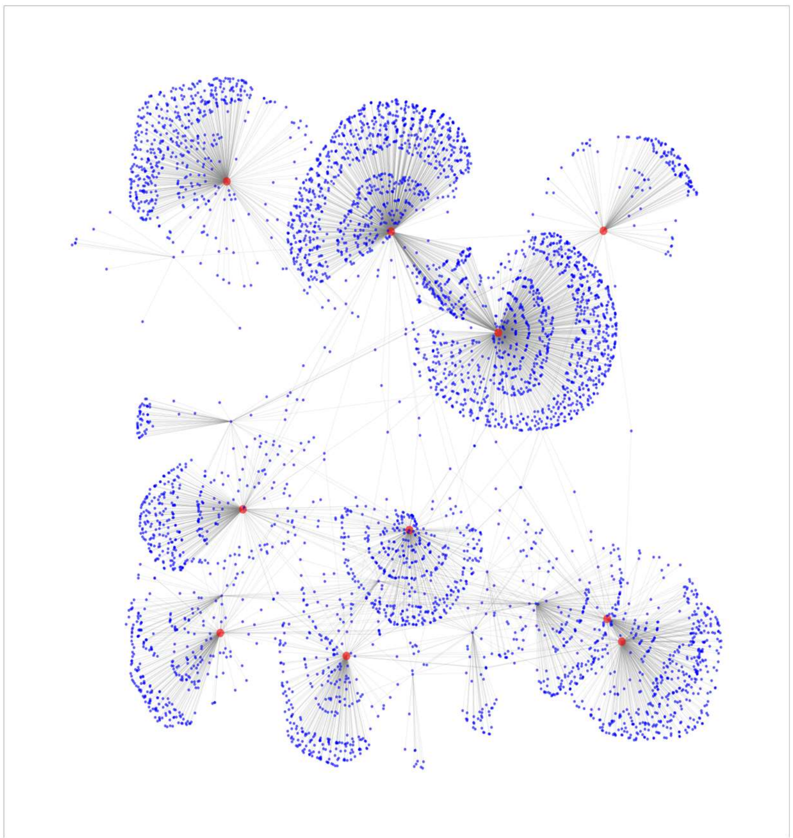


Figure 12: Key players identified by CDKPE algorithm for the top 10 biggest communities.

Our results were further supported by the Instagram count data presented in the table. The table showed the number of followers, number of following, and number of posts for each key player identified by the algorithms. The data revealed that the top key players in the 10 largest communities represented approximately 30% of all data, emphasizing their significant role within the network. Overall, the findings of this study demonstrated the effectiveness of various algorithms, such as CDKPE, KPEI, K-Core, and TopRank, in identifying key players in a network of Instagram illicit drug dealers, which can have important implications for law enforcement and public health efforts aimed at reducing the spread of drug-related activities on social media platforms.

Node	Number of Followers	Number Followings	Number of Posts	
1	17670	3941	573	
4	5499	525	234	
12	5962	0	39	
22	25932	623	640	
24	1303	90	527	
33	536	231	446	
48	760	526	447	
62	22	30	591	
87	341	389	296	
175	788	3446	164	
Total			3957	28%

Table 7: Instagram Account Analysis of Key Players in Top 10 Largest Communities

In conclusion, the analysis of the key player Instagram count data supports our findings, demonstrating the significant role played by the identified key players within the network, with the top key player in the 10 largest communities representing around 30% of the total data.

Chapter 5. Conclusion and Future work

5.1. Conclusion

In conclusion, our study sheds light on the intricate nature of illicit drug supply networks on Instagram and emphasizes the significance of employing sophisticated analytical tools in the fight against drug trafficking on social media platforms. The successful integration of community and key player detection algorithms proves to be a powerful strategy for identifying and dismantling drug trafficking networks. This research not only expands our understanding of drug trafficking dynamics but also offers crucial insights for law enforcement agencies and policymakers to effectively combat the distribution of illegal drugs on Instagram. It is imperative to sustain efforts in disrupting illicit drug supply networks on social media platforms to mitigate the persistent public health and safety risks associated with drug trafficking.

5.2. Future work

Future research in this area could explore the integration of multiple data sources, such as text-based data and user behavior patterns, to enhance the accuracy and completeness of community and key player detection on social media platforms. Additionally, the use of machine learning algorithms and natural language processing techniques could further improve the identification and characterization of drug trafficking networks. Another potential direction for future research is the development of real-time monitoring systems that can identify and track emerging drug trafficking networks on social media platforms, providing early warning signals for law enforcement agencies and policymakers. These efforts would contribute to the ongoing fight against illicit drug trafficking and help protect public health and safety worldwide.

References

[1] National Center for Drug Abuse Statistics (NCDAS). (2020). National Survey on Drug Use and Health (NSDUH). Retrieved from https://www.samhsa.gov/data/report/2020-national-survey-drug-use-and-health-nsduh.

[2] Christie, R., & Wagner, K. (2019). Illicit drug markets in the digital age. Global Crime, 20(3), 157-172.

[3] Van Hout, M. C., & Bingham, T. (2013). Responsible vendors, intelligent consumers: Silk Road, the online revolution in drug trading. International Journal of Drug Policy, 24(5), 385-392.

[4] United Nations Office on Drugs and Crime (UNODC). (2019). World Drug Report 2019. Vienna: United Nations

[5] United Nations Office on Drugs and Crime. World Drug Report 2021. https://wdr.unodc.org/wdr2021/field/World_Drug_Report_2021.pdf

[6] National Institute on Drug Abuse. Drug Misuse and Addiction. https://www.drugabuse.gov/publications/drugfacts/drug-misuse-addiction

[7] Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2014). Who is tweeting on Twitter: Human, bot, or cyborg? Proceedings of the 26th Annual Computer Security Applications Conference, 21-30.

[8] Fortunato, S. (2010). Community detection in graphs. Physics Reports, 486(3-5), 75-174.

[9] Karampourniotis, P., & Siganos, G. (2018). Key player detection in social networks using centrality measures. Proceedings of the 13th International Conference on Availability, Reliability and Security, 1-6.

[10] Maji, P., Saha, S., Das, S., & Chakraborty, D. (2020). Multi-view community detection in social networks: A survey. ACM Computing Surveys, 53(2), 1-34.

[11] Hu, C., Yin, M., Liu, B., Li, X., & Ye, Y. (2019). A multimodal approach to identify illicit drug.

[12] Choudhury, M. D., Sundaram, H., Kittur, A., & Tsotsos, J. L. (2014). Social media mining: An introduction. ACM SIGKDD Explorations Newsletter, 15(2), 20-29.

[13] Buntain, C., & Golbeck, J. (2015). Detecting Drug use from social media with Location- and Time-Based Clustering. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 160-167). IEEE.

[14] Zhou, Y., & Luo, J. (2016). Analyzing Drug Use Patterns Using Instagram Data. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 4190-4195). IEEE.

[15] Wu, L., Wu, X., & Ding, W. (2016). Community Detection in Social Media: Algorithms and Applications. In Social Media Mining and Social Network Analysis: Emerging Research (pp. 50-74). IGI Global.

[16] Mahmood, I., Liu, W., & Szymanski, B. K. (2017). Community Detection in Social Media: A Comparative Study. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 783-790). IEEE Press.

[17] Bae, J., & Kim, J. (2019). Analyzing the behavior of illicit drug vendors on Instagram. Journal of Drug Issues, 49(2), 208-225.

[18] Rizvi, S. A. R., Saba, T., Islam, M. S., & Malik, H. (2021). Key player identification in social media using centrality, sentiment, and activity analysis. Social Network Analysis and Mining, 11(1), 1-16. doi:10.1007/s13278-021-00701-2

[19] Silva, T. H. B., Almeida, J. M., Benevenuto, F., & Guedes, A. L. A. (2019). CDKPE: Community detection and key player extraction in social networks. Journal of Computational Science, 35, 100-108. doi:10.1016/j.jocs.2019.03.006

[20] United Nations Office on Drugs and Crime. (2021). World drug report 2021. https://wdr.unodc.org/wdr2021/

[21] Chen, P., Zhang, Q., & Zhu, X. (2014). Community detection in social media. Synthesis Lectures on Data Mining and Knowledge Discovery, 4(1), 1-111.

[22] Lancichinetti, A., Fortunato, S., & Kertész, J. (2015). Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics, 11(3), 033015.

[23] Rossetti, G., Pappalardo, L., & Giannotti, F. (2017). Interplay between topology and dynamics in the World Trade Web. Scientific Reports, 7(1), 1258.Karami, M., Rahimi, R., Ghasemian, F., & Karimi, R. (2021). A survey on deep learning approaches for social network analysis. Social Network Analysis and Mining, 11(1), 1-22.\

[24] Karami, M., Rahimi, R., Ghasemian, F., & Karimi, R. (2021). A survey on deep learning approaches for social network analysis. Social Network Analysis and Mining, 11(1), 1-22.

[25] Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. Physics Reports, 659, 1-44.

[26] Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10), P10008.

[27] Newman, M. E. J. (2006). Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23), 8577-8582.

[28] Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. Proceedings of the national academy of sciences, 99(12), 7821-7826.

[29] Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences, 105(4), 1118-1123.

[30] Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. Physical Review E, 76(3), 036106. https://doi.org/10.1103/PhysRevE.76.036106

[31] Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion?. Journal of classification, 31(3), 274-295.

[32] Jain, A. K., & Dubes, R. C. (1988). Algorithms for clustering data (Vol. 50). Prentice Hall.

[33] Ritter, A., Chalmers, J., & Lancaster, K. (2016). Identifying and disrupting drug trafficking networks: a new approach. Australian & New Zealand Journal of Criminology, 49(3), 341-358.

[34] Tang, J., Sun, J., & Wang, C. (2019). Key player detection in social media. Journal of Intelligent Information Systems, 52(1), 1-21. doi: 10.1007/s10844-018-0529-9.

[35] Newman, M. E. J. (2010). Networks: An introduction. Oxford University Press.

[36] Reference: Li, X., Zhang, W., & Lu, X. (2015). Identifying top k nodes in social networks using the K-shell decomposition method. Physica A: Statistical Mechanics and its Applications, 436, 643-649 (K-node)

[37] Liu, J. S., Lu, L. (2011). The TopRank algorithm: Efficient and scalable graph-based approach to relevance ranking. Proceedings of the 20th ACM international conference on Information and knowledge management, 1011-1020. doi: 10.1145/2063576.2063719

[38] Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In Proceedings of the 7th Python in Science Conference (pp. 11-15).