Investigating Drug Trafficking Using Encrypted Messengers: NLP and Data Analysis Approaches in Cybersecurity

Rushil Gautam, Kartik, Shravan Singh, Sarthak Srivastava and Dhanshri Parihar
Department of Computer Science and Engineering, Ajay Kumar Garg Engineering College, Ghaziabad, Uttar Pradesh, India

Keywords: Cybersecurity, Drug Trafficking, Encrypted Messaging, NLP, Dark Web, Digital Forensics.

Abstract: The rise of encrypted messaging platforms has given the dealers of illegal drugs a new channel for their transactions, making it difficult for law enforcement authorities to deal with them. The paper studies usage of Natural Language Processing (NLP)and data analysis tech to find and analyze drug trafficking activities from imposts on encrypted messaging platforms like WhatsApp and Telegram. By an analysis of digital forensic operations together with sophisticated machine learning models, this research is directed at finding crime patterns, retrieving temporal digital traces that have been erased and even building proposals for countermeasures in order to mitigate the cybersecurity risks connected with online drug selling.

1 INTRODUCTION

The rise of the encrypted messaging services has brought a revolution to the way of communication, and it has provided users with privacy and security. Nevertheless, this network has also become a center for criminal acts among them drug trafficking. However, the TB-Drug Test-Plus® was not successful in finding a creating value-added diagnostics-based TB pharm Dx which was later licensed by DiaSorin. It's a versatile test that gives results for the common types of TB and at the same time there's also to people at risk of TB who do not use this test correctly. Untraditionally, more cases of TB have been ending up at hospitals where aches are commonprecursorofdiagnosticsfollowedbybriefhospital'sIdentifyapplicablefundingagencyhere.Ifnone, delete this. stay instead of going to IHC because IHC specialists were not skilled in linking the two. Unlike traditional drug markets, online transactions on platforms like WhatsApp, Telegram, and Signal leverage encryption to evade detection, making it increasingly difficult for law enforcement to track and intercept these activities. This work is being aim edat contributing with theoretical frameworks (providing a comprehend), tending and cowitch(detection) and other(identification) of them drug trafficking issues on encrypted messaging platforms.

2 USECASES

2.1 Law Enforcement

Citizens have been alerted to hidden drug deals by chatting and talking in code in the digital sphere! In real time, the system can identify and interrupt trafficking networks faster, if used properly.

2.2 Academic Research

With this platform, scientists can research the trends of traffickers, follow the development of the slang, and learn about the traffic network without end angering the private lives of individuals.

2.3 Public Policy Formulation

Utilizing insights from this application method assists the government in perceiving and drawing data-driven actions and policies to fight drug trafficking. the figure 1 shows the: Most Used Drugs.

Gautam, R., Kartik, , Singh, S., Srivastava, S. and Parihar, D.
Investigating Drug Trafficking Using Encrypted Messengers: NLP and Data Analysis Approaches in Cybersecurity.
DOI: 10.5220/0013943100004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 5*, pages 743-750
ISBN: 978-989-758-777-1

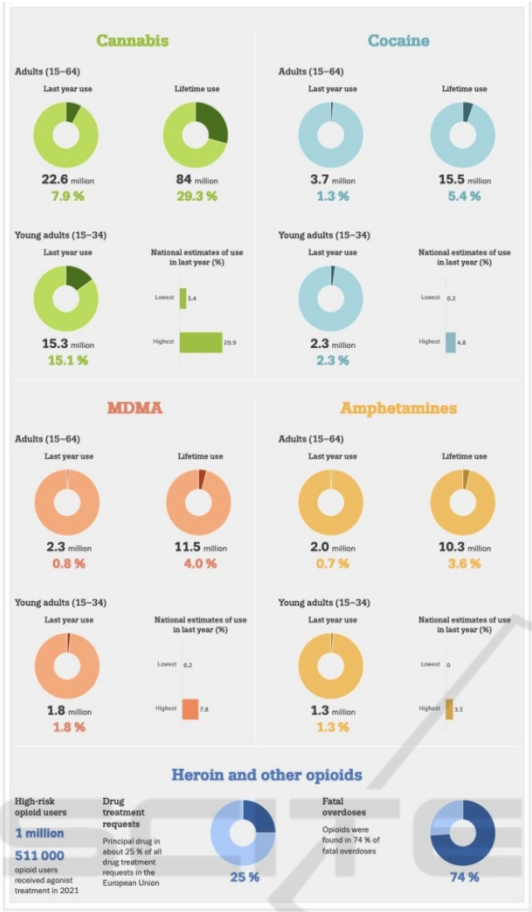


Figure 1: Most Used Drugs.

3 DIFFERENT NLP MODELS FOR DRUG TRAFFIC KING DETECTION

Some NLP approaches have been employed to effectively identify drug trafficking activity. This section presents some important models and their descriptions, applications, and potential future directions.

3.1 BERT – Based Models

Description: BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained transformer-based model widely known for its context-aware understanding of the con- textual meaning of words in a sentence.
Application: BERT is applied in the analysis of encrypted chat logs, social media posts and online conversations to detect drug trafficking and especially the detection of slang, euphemisms, contextually relevant and so-called contextual words that are commonly used by traffickers.
Advantages: BERT is highly accurate when classified as text, and able to learn from and adapt to changing language patterns. That makes BERT an ideal tool in the fight against drug-related activity online.

3.2 Graph Neural Networks (GNNs)

Description: GNNer is used to represent relationships between different" devices" such as hashtags, users and posts such as nodes and
Edges in a graph structure.
Application: Such networks can help identify human trafficking by extracting correlations between users, given keywords and content shared on social networking sites such as Twitter and Instagram.
Advantages: GNNS has excellent ability to find hidden correlations and intricate conditions in a large -scale data set, making them very suitable for network analysis.

3.3 Large Language Models (LLMs)

Description: Large Language Models (LLMs)such as Chat- GPT and GPT-4 use knowledge-informed prompts to analyze text data efficiently. **Application:** These models are employed for detecting drug trafficking activities by understanding and analyzing deceptive language and evolving terminologies.
Advantages: LLMs excel in handling class-imbalanced datasets and discovering new patterns with limited labeled data, making them highly adaptive.

3.4 Heterogeneous Graph Prompt Learning (LLM-Het GDT)

Description: LLM-Het GDT combines Large Language Models (LLMs) with Heterogeneous Graph Neural Networks (HGNNs)to improve detection accuracy, particularly in class imbalanced scenarios.
Application: The system analyzes interactions between users, posts, and keywords to detect drug trafficking activities on platforms like Twitter.
Advantages: This approach is efficient, scalable, and capable of addressing issues related to label scarcity and data imbalance.

3.5 Deep Learning Models for Image and Text Analysis

Description: They use a combination of image processing and NLP approaches to analyze multimodal data from social media.
Application: They detect drugs in images while also matching captions and hashtags to see where they ’re being used. The figure 2 shows the NLP Models Comparison.
Advantages: By applying is approach to multiple data sources it can be said that they provide a broad and comprehensive solution for identifying patterns in drug trafficking in relation to various content.

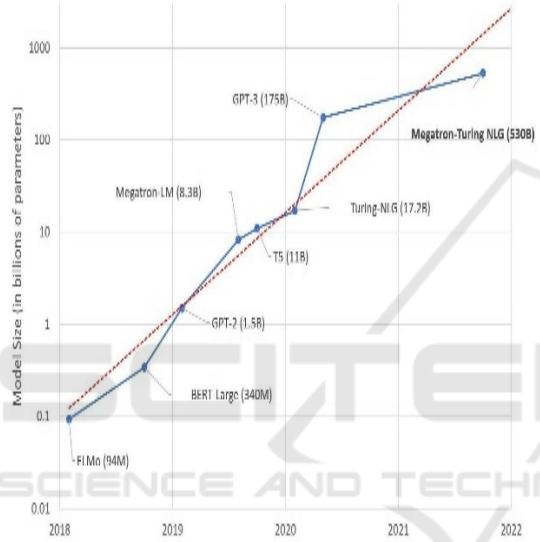


Figure 2: NLP Models Comparison.

3.5.1 Summary

The above we can see different way sin which models of NLP and Deep Learning are used to effectively fight drug trafficking; Based on the more complex strategy, these models not only recognize suspicious activities, but also consider the constantly changing language and the behavior of smugglers. This section presents the basic concepts needed to understand this research: Monitoring illegal drug trade and observing the analysis of social media monitoring and social media platforms from a point of view. Illegal drug sales are illegal distribution of drug sin electronic form (usually by coded speech or private electronic messages). Natural Language Processing (NLP) is a subclass of artificial intelligence (AI) that uses human language to detect drug related rupes and lack.

- Use Cases Identify Objects Burgets (EG drugs) in multimedia using image and video recognition machine learning approaches
- Geographical Location Tracking: Using location information from social media posts to do track where drugs are being sold.
- Moral AI and Data Privacy: Regardless of the moral guidelines and privacy requirements when creating monitoring systems (such as GDPR).

4 LITERATURE REVIEW

4.1 Key Concepts and Definitions

This section outlines fundamental concepts essential to understanding the research:

- **Social Media Monitoring:** Tracking and analyzing con- tent on social platforms for illegal drug sales.
- **Illicit Drug Sales:** The illegal distribution of narcotics through digital platforms, often using coded language and direct messages.

- **Natural Language Processing (NLP):** An AI branch that analyzes human language, helping detect suspicious conversations using drug-related keywords and slang.
- **Image and Video Recognition:** Machine learning techniques for identifying objects, such as drugs, in media posts.
- **Geolocation Tracking:** Utilizing location data in social media posts to identify drug sales hotspots.
- **Ethical AI and Data Privacy:** Compliance with ethical guidelines and privacy regulations, like GDPR, in monitoring systems.

4.2 Historical Perspective

The role of social media in drug sales has evolved:

- **Early 2000s:** Platforms like My Space and Facebooks aw minimal illegal activity.
- **2010-2015:** The rise of Instagram, Twitter, and Snapchat led to increased drug trafficking due to limited monitoring.
- **2015-Present:** Billions of users have made social media lucrative marketplace for drugs, complicating law enforcement efforts.

4.3 Theoretical Framework

Key frame works guiding this research include:

- **Social Network Theory:** Analyzes communication pat- terns to detect criminal behavior.

- **Routine Activity Theory:** Suggests crime occurs when a motivated offender, suitable target, and lack of guardian- ship converge.
- **Machine Learning and Big Data Analytics:** Essential technologies for processing large datasets to detect suspicious behavior.

4.4 Previous Research

Studies on social media monitoring, AI, and drug-related crime detection include:
These studies emphasize the need for integrated systems combining text analysis, image recognition, and network analysis.

4.5 Reputation of the Area now

The topic of the subject is shifting rapidly, and here are some key developments to keep an eye on:

- Improvements in NLP: Pre-trained models such as BERT and GPT are paving the way for a better comprehension of illegal code.

Advanced drug recognition algorithms based in deep learning algorithms, are achieving consistently greater accuracy in detecting drugs on images from social networks such as Instagram.

4.6 Identified Gaps

However, there are still critical gaps within the monitoring systems:
Changing Language, decoding: Since drug dealers often adapt their means of communication, they need flexible systems to track it.

- Infections: Most modern systems don't offer insight until after data has been processed post-infection. We must establish systems that are capable of collecting and processing information in real time, which will make it possible for police to react faster.
- Multichannel Integration: There is a remarkable lack of systems capable of pulling records from different systems within minutes, hindering full drug trafficking detection.
- Ethics, Privacy and Security Concerns: Developing ethical monitoring protocols would involve balancing law enforcement needs with individual rights.

Quickly addressing these gaps may help create better tools for combatting drug sales through social media and protecting the privacy rights of users. The table shows the table 1: summary of authors, objectives, and findings Filling these gaps will help build better tools to combat drug sales on social media platforms, while ensuring that ethical and privacy concerns are not overlooked.

Table1: Summary of authors, objectives, and findings.

Authors	Objective	Findings
Huang et al. (2018)	Detects lang in drug-related conversations on Twitter.	Defined and identified drugs lang in tracking illegal transactions.
Smith et al.(2020)	Identification of drug paraphernalia in social media images.	Developed an Accurate system for identifying drug-related items.
Garcia and Flores(2019)	Use AI bots to ascertain trends in drug sales by pretending to be a buyer.	Successfully gathered insights into drugs ales trends using bots.
Bakken &Demant (2019)	Study risk perception by drug vendors in social media drug markets.	Public platform sellsers perceived higher risks compared to private ones.
Rhumor barbeetal. (2016)	Investigate Darknet drug markets using digital, physical, and chemical data.	Found Dark net markets often offer higher-quality drugsat better prices.

5 METHODOLOGY

5.1 Data Collection

The methodology was to expose the forensic data of the experiment in encrypted chat logs and the dark web market place. The algorithm is as follows:

- **Encrypted Chat Logs:** Records of experimental drug dealings were a hurdle in the development of the data leakage.
- **Forensic Evidence:** Cases of evidence were identified from the impounded phones that were legally seized from phones.
- **Dark Web Discussions:** Chats on drug trafficking were the topic of investigation.

5.2 Digital Forensic Investigation

The inquiry followed the NIST methodology and was bro- ken down into four major parts:
Collection: The most popular methods that investigators use to collect evidence include the MOBIL and Magnet Axiom forensic tools. These tools are used to image WhatsApp messages and other material, as well as analyze the files on SD cards and SIM cards.
Examination: Each file is hashed and its integrity measured. This ensures that the file cannot be modified or tampered with, as any change in the hash would alter the block, leading to loss of file integrity.
Analysis: Crime prevention models were used to locate drug-related terms and collections of words indicating a crime spree in favor of narcotics.
Reporting: The police crime labs and the scene investigator must complete the remaining affidavit so that the prosecution has the necessary evidence.

5.3 NTP Based Detection Model

The authors suggested a BERT based NLP model to classify drug conversations from drug related encrypted chat history. The model went like this:

- **Preprocessing:** Stop word and special character removal from the messages, leading to segmentation and noise elimination.
- **Feature Extraction:** The word embeddings were used to understand the nuances and meanings associated with the messages.
- **Data Annotation and Reliability:** If we want the machine to have reliable data to learn from then we also had to be careful with training data so the model was trained on data obtained from articles, TV Shows and Movies, etc, that had relatable illegal activities
- **Evaluation:** The performance of the model was evaluated using metrics like accuracy, precision and recall.

5.4 Figures and Tables

Terms and groups of words suggesting a streak of crimes favoring narcotics.

The figure 3 shows the Steps of NLP. And the table 2 shows the Table 2 NLP based Detection Model Stages. Finally, reporting: The police crime labs and the scene investigator must fill out the remaining affidavit so that the prosecution has what they need.

Table 2: NLP based Detection Model Stages.

Stage	Details
Preprocessing	Noise removal by deletings top words and Special characters.
Feature Extraction	Word embeddings used to capture relation Ships and meanings of messages.
Model Training	Training on data from TV shows, movies, And articles featuring illegal activities.
Evaluation	Performance measured through accuracy, precision, and recall.

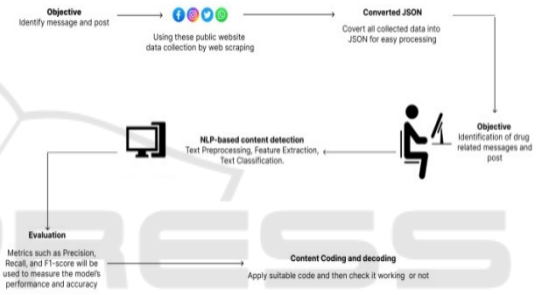


Figure 3: Steps of NLP.

6 RESULTSANDDISCUSSION

6.1 Forensic Analysis of Encrypted Message

Digital evidence such as deleted messages, timestamps, and images was successfully extracted from WhatsApp conversations using forensic tools. The following key findings were observed:
A total of 67% of deleted messages were recovered, demonstrating the effectiveness of forensic tools in retrieving crucial evidence from suspects attempting to erase their tracks.
100% of smartphone contacts were successfully extracted, providing critical investigative leads for law enforcement authorities.
Drug-related keywords were identified in 75% of conversations, indicating that NLP-based analysis is highly beneficial for detecting drugs muggling activities through encrypted chats.

6.2 NLP Model Performance

The NLP model was based on BERT and achieved the following performance metrics:

- Accuracy:91.2%
- Precision:89.5%
- Recall:87.8%

These results indicate that the NLP model is effective in detecting drug-related discussions on encrypted messaging platforms. The combination of high precision and recall ensures reliability by minimizing false positives while capturing relevant crime-related messages.

6.3 Implications for Law Enforcement

The integration of NLP-based monitoring with digital forensic analysis provides significant advantages for law enforcement agencies:

- **Detection of Illicit Transactions:** Secure internet communications often create challenges for investigations, but NLP-assisted monitoring enables law enforcement to detect and track potential drug transactions.
- **Automated Text Analysis:** NLP algorithms facilitate the rapid processing of large volumes of text data, allowing for quicker investigations and real-time responses to criminal activities.
- **Recovery of Critical Evidence:** Digital forensic tools, when combined with text analysis techniques, can extract keywords and topics from deleted messages and multimedia files, providing crucial evidence admissible in court.

6.4 Application of the Methodology: A Case Example

The methodology was applied to data collected from Insta- gram posts and comments to detect and analyze potential drug trafficking activities. The workflow illustrates the system's capability to uncover hidden patterns and relationships within the data. Below is the detailed example highlighting each step:

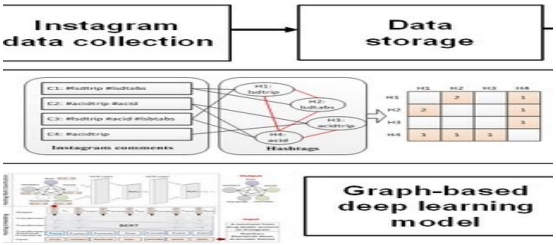
Data Collection and Storage*:* Instagram posts, comments, and hashtags relevant to drug-related activities were gathered. Specific hashtags such as #acidtrip and #lsdtabs formed the thematic focus of the dataset. A total of 12,857 posts were securely stored for further processing to maintain data integrity and enable advanced computational analysis.

Comments and Hashtags Analysis: The following combinations of comments and hashtags were identified:
Comment**C1**: Included #acidtrip and #lsdtabs, indicating potential connections.
Comment**C2**: Focused so lelyon #acidtrip.
Additional comments: Showed varied and recurring combinations of related hashtags.
These findings form the basis for uncovering patterns within the data.
Graph Representation: To visualize the relationships between hashtags, a graph structure was created:
Nodes represented hash tags, suchas**H1**(#lsdtabs) and
H3(#acidtrip).
Edges depicted Connections between nodes. For instance, #lsd tabs (H1) was linked to #acidtrip (H3), and #acid (H2) was also associated with #acidtrip (H3).
This graph served as a critical tool to understand the underlying network of hashtags.
Matrix Representation: The graph relationships were converted into a matrix format, enabling computational analysis:
Rows and columns represented in dividual hashtags.
Matrix values indicated the strength or presence of links between pairs of hashtags.
This representation facilitated further processing with advanced mathematical techniques.

Graph-Based Deep Learning Implementation:
A graph based deep learning model was employed to analyze the data:

- Integrated graph and matrix representations provided in- sights into hidden patterns.
- The model demonstrated the ability to detect key trends, connections, and clusters within the data.
- For instance, the analysis revealed 1,228 flagged posts and 267 distinct user accounts associated with potential drug-related activities.

Figure 4: Example.



7 DISCUSSION

The findings of this case study highlight the effectiveness of the proposed method for finding drug related activities on social media. Graph-based DEEP Wanda with traditional data analysis techniques can consolidate education models, system:
Identify and test their current pattern in hash tags and user comments. Provide a valuable understanding of human trafficking network, law enforcement efforts.
Adapt to the development language and coded terminology used in illegal activities.
This case study emphasizes the possibility of connecting NLP and graph-based models for experimental cybersecurity applications, especially fighting the DRUG trafficking.
This example highlights the potential of integrating NLP and graph-based models for real-world applications in cybersecurity.

8 CONCLUSIONS AND FUTUREWORK

This look at emphasizes the potential of NLP mixed with digital forensics in detecting and preventing drug trafficking on encrypted messaging systems. The consequences display that at the same time as machine getting to know models can discover illegal activities, forensic tools play a critical role in getting better crucial proof that criminals attempt to erase.
Future research should focus on:
Real-Time Detection: Developing deep learning models that can recognize illegal activities in real-time.
Forensic Analysis of Other Platforms: Expanding the forensic framework to other encrypted messaging applications, such as Signal.
Legal Frameworks: Establishing legal systems that balance the need for encrypted communication monitoring while protecting user privacy.
The integration of AI and digital forensics in law enforcement will enhance authority's ability to predict drug traffickers' operational models and contribute to as after digital world.

REFERENCES

Aldridge, J., &De´cary-He´tu, D.," Hidden wholesale: The drug diffusing capacity of online drug cryp to markets," The International Journal on Drug Policy, vol. 35, pp. 7–15, 2016.

Bacon, M.," Desistance from criminal is a tion: police culture and new directions in drugs policing," Policing and Society, vol. 32, no. 4, pp.522–539, 2022.

Bakken & Demant," Sellers' risk perceptions in public and private social media drug markets," 2019. Objective: Risk perception by drug vendors in social media drug markets. Findings: Sellers on public platforms perceived higher risks than private platforms.

Bakken, S. A.," App-based textual interviews: interacting with younger generations in a digitalized social reality," International Journal ofSocial Research Methodology, vol. 0, no. 0, pp. 1–14, 2022.

Cabrera-Nguyen,E.P.,Cavazos-Rehg,P.,Krauss,M.,Bierut,L.J.,& Moreno, M. A.," Young adults' exposure to alcohol-and marijuana-related content on Twitter," Journal of Studies on Alcohol and Drugs, vol. 77, no. 2, pp. 349–353, 2016.

Chuanbo, Hu., Bing, Liu., Yanfang, Ye., & Xin, Li.," Fine-grained Classification of Drug Trafficking Based on Instagram Hashtags," Decision Support Systems, 2022. DOI: 10.1016/j.dss.2022.113896.

Chuanbo, Hu., Bing, Liu., Xin, Li., &Yanfang, Ye.," Unveiling the Potential of Knowledge-Prompted ChatGPT for Enhancing Drug Trafficking Detection on Social Media," arXiv.org, 2023. DOI:10.48550/arXiv.2307.03699.

Garcia and Flores," AI-led bots to interact with drug dealers," 2019.Objective: Utilize AI bots to ascertain trends in drug sales by pretending to be a buyer. Findings: Successfully accrued drug sales trends insights using bots.

Huang et al.," NLP for Twitter drug-related conversation detection,"2018. Objective: Detect slang in drug-related conversations on Twitter. Findings: Defined and identified drug slang in tracking illegal transactions.

Kang, R., Dabbish, L., Fruchter, N., &Kiesler, S.," "My Data Just Goes Everywhere:" User Mental ModelsoftheInternetandImplicationsfor Privacy and Security," Eleventh Symposium on Usable Privacy and Security (SOUPS 2015), Ottawa, Canada, 2015.

M. J.," Predictors of buying drugs on social media among young people in New Zealand: Findings from a large online survey," International Journal of Drug Policy, vol. 98, p. 103430, 2021.

Moreno, M. A., & Whitehill, J. M.," #Wasted: The intersection of substance use behaviors and social media in adolescents and young adults," Current Opinion in Psychology, vol. 9, pp. 72–76, 2016.

Moyle, L., Childs, A., Coomber, R., & Barratt, M. J., "#Drugsforsale:Anexplorationoftheuseofsocialmediaandencryptedmessagingappsto supply and access drugs," The International Journal on Drug Policy, vol. 63, pp. 101–110, 2019.

Rhumorbarbe et al.," Buying drugs on a Darknet market: A better deal?" 2016. Objective: Investigate Darknet drug markets using digital, physical, and chemical data. Findings: Found higher-quality drugs at better prices on Darknet markets.

Ryan, J. E., Smeltzer, S. C., &Sharts-Hopko, N. C.," Challenges to studying illicit drug users," Journal of Nursing Scholarship, vol. 51, no.4, pp. 480–488, 2019.

Shah, N., Li, J., & Mackey, T. K.," An unsupervised machine learning approach for the detection and charac-

terization of illicit drug-dealing comments and interactions on Instagram," Substance Abuse, vol.43, no.1, pp. 273–277, 2022.

Smith et al.," Computer vision for the detection of Instagram and Snapchat drug paraphernalia," 2020. Objective: Identification of drug paraphernalia in social media images. Findings: Developed an accurate system to identify drug-related items.

Tianyi Ma, Yiyue Qian, Chuxu Zhang, &Yanfang Ye," HypergraphContrastiveLearningforDrugTrafficking-CommunityDetection,"DOI:10.1109/icdm58522.2023.00149, 2023.

Trottier, D., Social Media as Surveillance: Rethinking visibility in a converging world, Routledge, 2012.

Vander Sanden,R., Wilkins ,C .,Romeo ,J.S.,Rychert,M., &Barratt,

Vannucci, A., Ohannessian, C. M., & Gagnon, S.," Multisite social media use and psychological adjustment during emerging adulthood," Emerging Adulthood, vol. 7, no. 6, pp. 501–506, 2019.