



Social drug dealing: how peer-to-peer fintech platforms have transformed illicit drug markets

Jörn Boehnke¹ · Pantelis Loupos¹ · Ying Gu²

Received: 1 September 2022 / Accepted: 14 September 2023 / Published online: 27 October 2023
© The Author(s) 2023

Abstract

Digital platforms have revolutionized the way illegal drug trafficking is taking place. Modern drug dealers use social network platforms, such as Instagram and TikTok, as direct-to-consumer marketing tools. But apart from the marketing side, drug dealers also use fintech payment apps to engage in financial transactions with their clients. In this work, we leverage a large dataset from Venmo to investigate the digital money trail of drug dealers and the social networks they create. Using text and social network analytics, we identify two types of illicit users: mixed-activity participants and heavy drug traffickers and build a random forest classifier that accurately predicts both types of illicit nodes. We then investigate the social network structure of drug dealers on Venmo and find that heavy drug traffickers share similar network characteristics with previous literature findings on drug trafficking networks. However, mixed-activity participants exhibit different patterns of network structure characteristics, including a higher clustering coefficient, suggesting that they may be accessing multiple networks and bridging those networks through their illicit activities. Our findings highlight the importance of distinguishing between these two types of illicit users and provide law enforcement agencies with valuable insights that can aid in combating illegal drug transactions in digital payment apps.

Keywords Illicit markets · Peer-to-peer platforms · Text analytics · Social network analysis · Machine learning · Venmo

✉ Jörn Boehnke
jb@ucdavis.edu

✉ Pantelis Loupos
ploupos@ucdavis.edu

Ying Gu
yg2689@uw.edu

¹ Graduate School of Management, University of California Davis, 1 Shields Ave., Davis, CA 95616, USA

² Foster School of Business, University of Washington, 4273 E Stevens Way, Seattle, WA 98195, USA

1 Introduction

Digital platforms have revolutionized all kinds of businesses including illegal drug trafficking. Modern drug dealers have adopted social network platforms like Instagram and TikTok as direct-to-consumer marketing tools. They utilize these platforms to upload short-lived videos and images showcasing their illicit merchandise accompanied by code words and emoji to entice potential buyers (Hu et al., 2021). In addition to their marketing efforts, drug dealers also employ fintech payment apps to conduct financial transactions with their customers. After striking a deal with a prospective buyer, the drug dealer requests payment via mobile payment apps like Venmo and CashApp. A recent anonymous survey¹ showed that nearly one-third of millennials have used Venmo to pay for drugs. While previous works have focused on identifying the marketing practices used to promote illegal drugs on social media (Hu et al., 2023; Li et al., 2019; Mackey et al., 2017, 2018), research has yet to investigate drug dealers' financial activities on peer-to-peer (P2P) payment platforms.

Plenty of anecdotal evidence suggests that drug dealers have increasingly turned to mobile payment apps as a means of conducting transactions. According to a recent report² by the United States Government Accountability Office to the congressional committees, this shift towards digital platforms has been attributed to their ease of use and convenience. Moreover, the use of mobile payment apps by drug dealers has been further accelerated by the COVID-19 pandemic, as social distancing measures and lockdowns made traditional forms of transaction more difficult. So, how do drug dealers use fintech payment apps to facilitate their illicit activities, and what are the implications for law enforcement and regulatory authorities? Moreover, how can data analytics and machine learning be used to detect and prevent the use of fintech payment apps by drug dealers?

Previous research on drug-related networks has shown that these networks often exhibit complex structures, with a small number of key players exerting a disproportionate influence on the overall network (Desroches, 2005; Malm & Bichler, 2011; Morselli, 2010; Natarajan, 2006; Zaitch, 2002). These key players, or “hubs”, act as intermediaries between different parts of the network and play a critical role in facilitating drug transactions. Additionally, research has highlighted that these networks exhibit low density and low cohesion (Framis & Morselli, 2014; Framis & Regadera, 2017; Klein & Crawford, 1967; Klein & Maxson, 2010; McGloin, 2005; Natarajan, 2000; Papachristos, 2006). That is, drug dealers tend to operate in relatively isolated clusters rather than in highly connected, tightly knit groups. All this prior work, however, focused on traditional offline drug-related networks. Therefore, it is important to explore whether the disruption of this illicit market by fintech payment apps led to the emergence of new social structures within drug-related networks.

To address these questions, we leverage a large dataset from Venmo, the most popular mobile payment app in the U.S., to investigate the digital money trail of drug dealers and the social networks they create. What distinguishes Venmo from other mobile payment apps is its social aspect. Venmo transactions are by default posted on a public newsfeed and are accompanied by a comment describing the purpose of the transaction. This distinct feature allows us to utilize text analytics to classify drug-related transactions. One might expect drug dealers to conceal their transactions from the public newsfeed, but in practice, this is not often the case. On the contrary, drug dealers willingly post transactions on the public newsfeed as it bolsters their (illicit) reputation and attracts new customers.

¹ <https://lendedu.com/blog/nearly-third-millennials-used-venmo-pay-drugs/>

² <https://www.gao.gov/assets/gao-22-105101.pdf>

As we lack labeled data to distinguish between licit and illicit Venmo users, we employ a combination of text analysis and heuristic rules to classify which users are engaging in drug-related transactions. In the first stage, we utilize text analytics to classify Venmo transactions as either illicit or not. This stage, however, poses a significant challenge due to the prevalence of sarcastic comments in Venmo transactions (Potamias et al., 2020). To address this issue, we develop a set of heuristic rules that enable us to create a list of potential illicit users. We then manually inspect and annotate these users to create two distinct labels. The first label is assigned to users who engage in drug-related transactions, but also have legitimate transactions in their Venmo profile, hereafter called “mixed-activity participants.” The second label is assigned to users who heavily engage in drug-related transactions, hereafter called “heavy drug traffickers.” Distinguishing between these two types of illicit users is important because it provides a more nuanced understanding of the social network structure and behavior of drug traffickers on Venmo.

After labeling the two types of illicit users on Venmo, we proceed to create user profile metrics that describe their behavior on the platform. To better understand a user’s patterns, we construct transactional and text analytics variables. In addition, we create several network variables that have been identified in prior criminology research to describe the network characteristics of these users. We then use these features to train a random forest classifier to gain deeper insights into illegal drug dealing activity on Venmo. Our model identifies both types of illicit users with high accuracy.

Subsequently, we explore the feature importance of transactional, text, and network metrics in identifying illicit users. We find that the number of incoming transactions is the most informative feature for identifying drug trafficking users in terms of transactional metrics. For text metrics, we find that the number of illicit words and emoji are the most informative variables. Finally, with respect to network metrics, our analysis reveals that degree centrality, eigenvector centrality, page rank, and local clustering coefficient are the most informative variables for identifying illicit nodes. In fact, our findings demonstrate that social network metrics are equally important predictors of drug-related activity as transactional and text analytics variables. This finding emphasizes the critical role of social network analysis in identifying illicit users on mobile payment apps, as it may be extended to fintech platforms without text-based features.

Finally, we examine the social network structure of illicit Venmo users and compare it to that of legitimate users. Consistent with prior literature, we find that heavy drug traffickers exhibit higher degree, Katz and page rank centrality measures, but have a lower local clustering coefficient compared to regular Venmo users. On the other hand, mixed-activity participants exhibit a distinct social network structure compared to both heavy drug traffickers and regular Venmo users. Specifically, mixed-activity participants have significantly higher degree centrality measures than regular Venmo users, along with higher Katz and page rank centrality measures. However, they also exhibit a much higher clustering coefficient than a normal Venmo user. These findings suggest that mixed-activity participants act as a bridge between different networks on the platform and play a pivotal role in facilitating the spread of illicit activity across these networks.

To sum up, the main contributions of this work are as follows. First, to our knowledge, we are the first to study drug-related financial networks on a social digital fintech platform. This is an important and timely topic given the rising prevalence of mobile payment apps in today’s world, and our findings open up a new and highly relevant research avenue for criminologists, cybercrime investigators, and policymakers who are interested in understanding and tackling drug-related activities in the digital world. Second, we provide a reliable and scalable methodology that combines text analytics with social network analysis to detect

illicit users. This methodology is highly effective in detecting and classifying illicit users on Venmo and can be adapted for use on other similar platforms. Last, we document the structural network differences between licit and illicit users providing law enforcement agencies with helpful network insights to help combat illegal drug trafficking on digital payment platforms. By leveraging these insights, agencies may be able to identify potential drug traffickers and disrupt their activities on other fintech platforms, even in the absence of text data.

2 Literature review

There is a significant amount of literature on the economics of crime and drug-related activities (Freeman, 1999). However, due to the complexity of acquiring accurate data on these transactions, most studies have relied on self-reported data (Hagedorn, 1994; Reuter et al., 1990). Two notable exceptions are Levitt and Venkatesh (2000) and Morselli et al. (2017), who analyzed data from Chicago's Black Knights and Quebec's Hells Angels, respectively. Levitt and Venkatesh (2000) found that economic factors alone are inadequate to explain participation in gang behavior, while Morselli et al. (2017) investigated the issue from a social network perspective and found that the transactional network of the Hells Angels did not revolve around a single person or group of people. Our work adds to this literature by exploring a much richer and more precise transactional dataset that spans the entire United States. Additionally, our study examines Venmo, which represents an entirely new transactional medium for drug dealers that has not been investigated before.

Venmo has created a financial ecosystem that attracts both licit and illicit users, which is a phenomenon that has been well-documented in the literature. For instance, Reuter and Haaga (1989) documented how legitimate actors bring complementary resources to illegal drug trafficking operations. Cook et al. (1983) found that in a traditional drug-related social network, each actor possesses unique resources that others lack and value. As a result, it is the ensemble of connections in the network that represents the opportunity structure and incentivizes everyone to participate in the illegal process. Additionally, the "social snowball effect" has been documented in numerous studies (Kleemans & Van de Bunt, 1999; Kleemans, 2007; Van de Bunt et al., 2014; Bodoh-Creed et al., 2018; Bhargava et al., 2020), where individuals who come into contact with participants in illicit operations may later become participants themselves and even serve as entry points for others who are attracted to such activities. These findings suggest that drug dealers are attracted to using digital payment platforms like Venmo as it offers new market opportunities. Moreover, it helps explain why drug dealers are willing to post their illegal transactions on Venmo's public newsfeed; essentially, drug dealers see their existing clients as a means to expand their drug trading activity into new networks. This can occur when friends of users who participate in a drug-related transaction observe these on Venmo's newsfeed and contacted the dealer directly or asked their friends for a referral.

Numerous studies have documented that drug trafficking networks share high values of brokerage (Desroches, 2005; Natarajan, 2006; Zaitch, 2002), low density, and low cohesion (Klein & Crawford, 1967; Klein & Maxson, 2010; McGloin, 2005; Papachristos, 2006). Our research contributes to this literature by examining how these characteristics manifest on digital payment platforms. Specifically, we find that drug dealers form tightly knit groups on Venmo, which is contrary to previous findings of low cohesion. It is important to note, however, that our research setting differs from prior research in one significant dimension. Previous studies, such as (Morselli & Giguere, 2006), found that if all drug traffickers are

removed from the network, the subset of non-traffickers becomes disconnected. In contrast, both lawful and illicit Venmo users remain interconnected through their lawful transactional networks, even after illicit transactions are excluded.

Last, research on the impact of social media platforms on drug-trafficking networks has revealed various insights into how these illicit networks operate. Several studies have explored the marketing strategies used by drug dealers on social media, such as Twitter (Buntain & Golbeck, 2015; Lossio-Ventura & Bian, 2018; Mackey et al., 2017, 2018; Sequeira et al., 2019; Tofighi et al., 2020), Instagram (Cherian et al., 2018; Demant et al., 2019; Hu et al., 2021, 2023; Li et al., 2019; Moyle et al., 2019; Qian et al., 2021), Facebook (Bakken, 2021; Demant et al., 2019, 2020), Reddit (Haupt et al., 2022; Wright et al., 2021), and dark net marketplaces (Pergolizzi Jr., et al., 2017; Du et al., 2019; Bracci et al., 2022). These studies reveal that social media is becoming an increasingly popular avenue for drug dealers to promote their products and services. In addition to uncovering the extent of drug-related activity on social media, these studies offer scalable machine-learning algorithms that can be utilized to detect and prevent drug trafficking. For example, Hu et al. (2021) develop a deep learning framework that uses image and text analysis to detect various forms of drug trafficking activities on Instagram. Similarly, Roy et al. (2017) proposed a method for identifying and filtering out drug-related social media posts, with the aim of reducing teens' exposure to such material. Our study contributes to this literature by examining the financial networks that drug dealers establish on digital fintech platforms, taking a step beyond existing research on the marketing aspect of drug trafficking on social media platforms.

3 Overview of data and methods

Research setting and data overview

To conduct this research, we leveraged a large dataset from Venmo. Venmo is a mobile payment application that allows users to make P2P transactions in a convenient and user-friendly way. A comprehensive overview of Venmo's descriptive statistics can be found in Loupos and Nathan (2019).

Venmo has become incredibly popular in the United States, particularly among young adults, and has transformed the way people exchange money. However, Venmo has been exploited by drug dealers as a means of facilitating illicit transactions. People who purchase drugs on Venmo often use code words or emoji to indicate the type and quantity of drugs they are buying. For example, they might use a pill or syringe emoji to represent a drug, or use slang terms like "greens", "blues", or "shrooms" to refer to marijuana, opioids, or hallucinogenic mushrooms, respectively. Similarly, the amount of the drug might be disguised using coded language, such as "pizza" for a kilogram of cocaine or "cupcakes" for a small amount of marijuana. It's also common for drug dealers to use euphemistic phrases like "pay for dinner" or "pitch in for gas" to describe drug transactions in Venmo messages.

Our dataset contains all the transactions of Venmo's public users from January 2013 up until May 2015. This corresponds to 2,291,655 users and approximately 23 million transactions. Each transaction contains the unique user IDs of the sender and receiver, a timestamp, and the message associated with the transaction written by the sender. By analyzing these written messages, we are able to identify illicit transactions. Of all the transactions available in our dataset, 0.44% contain a drug-related word or emoji, and they are made by 83,068 unique users. Figure 1 provides a representative example of drug-related interactions between

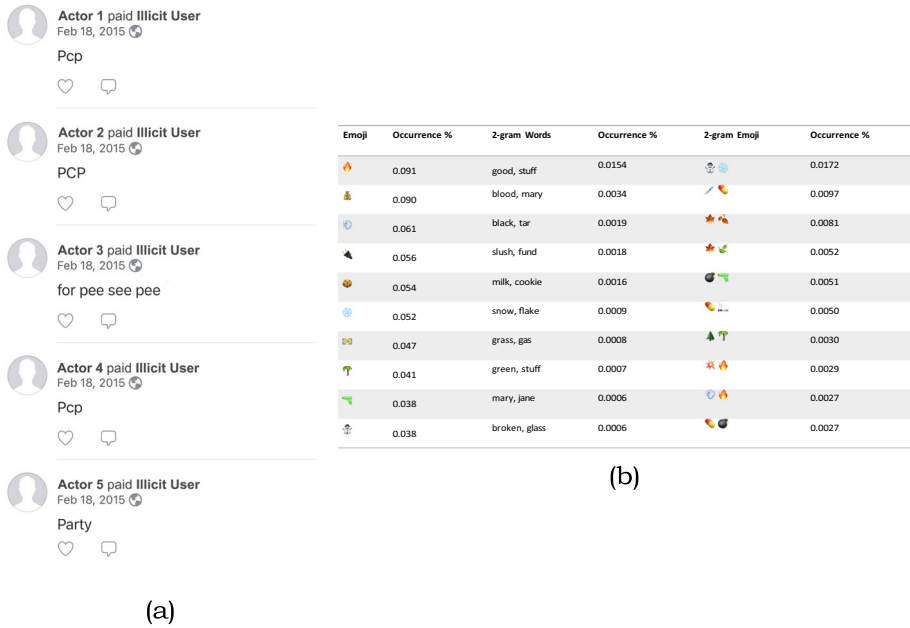


Fig. 1 **a** Shows the public Venmo newsfeed of a drug dealer. User names and pictures have been anonymized. **b** shows the most frequently used drug-related words and emoji

a drug dealer and regular Venmo users and highlights the most commonly used drug-related words and emoji.

4 Methodology

Using text analytics to identify illicit users

The first stage of our analysis is the creation of labeled data identifying drug related activities. To achieve this, we first compile a dictionary of drug-related keywords and emojis. The dictionary is created using a variety of sources, including the Drug Enforcement Agency’s list of commonly used drug-related emojis,³ common street slang terms for drugs, and manual identification of drug-related words and emojis in the Venmo dataset. Using this dictionary, we preclassify each transaction as either drug-related or non-drug-related.

We further create various variables to build a user profile. These variables include features such as total number of illicit outgoing transactions and total number of illegal emojis in the transaction descriptions sent to other users. A comprehensive description of these profile variables is available in Appendix A. We then use a heuristic rule to generate a list of potential illicit users. The rule necessitates three key criteria: a user must have at least 50 incoming transactions, at least one illegal token (word or emoji) in the description of an incoming transaction, and a minimum of 10% of all transactions should contain illegal tokens. The reasoning of this heuristic rule is grounded in the understanding that illicit nodes typically have a significant volume of transactions, hence the need for a minimum of 50 transactions.

³ dea.gov/sites/default/files/2021-12/Emoji%20Decoded.pdf

The presence of an illegal token in at least one transaction description is treated as a red flag indicative of potential illicit activity. Last, by stipulating that at least 10% of all transactions contain illegal tokens, we effectively reduce the possibility of false positives arising from sporadic instances of sarcasm or playful banter amongst Venmo users.

Upon application of our heuristic rule, we garner a list of potential illicit users. However, to further refine the accuracy of our findings and mitigate the risk of false positives due to sarcastic remarks masquerading as illicit comments, we manually inspect and annotate our list. While this process is time-consuming, it allows us to ensure the validity of our labels. Ultimately, by using a combination of heuristics and human annotation, we are able to identify and label a subset of Venmo users who engage in illicit activities using the app.

5 Social network metrics

In the second stage of our analysis, we construct various network variables based on previous criminology research. These variables are based on established metrics that measure both the influence of nodes, such as betweenness centrality and degree centrality, and the local structure of the network, such as clustering coefficient, which indicates the degree to which nodes in a network are connected to one another.

To create these variables, we model Venmo as a weighted static social network. Specifically, we consider the network to be a static graph, denoted by G , which contains V nodes and E edges. The weighted adjacency matrix W fully describes the structure of G , with each link weight corresponding to the total number of transactions between users i and j .

6 Model

Having at our disposal user profile data, including text variables, transactional variables, and network variables, as well as labeled transactions, we train a random forest classifier (RFC). The RFC is an ensemble learning technique that has proven to be both powerful and popular in the field of machine learning, particularly for high-dimensional classification and skewed problems (Breiman, 2001). The RFC combines multiple decision trees, where each tree is generated using vectors sampled independently (with replacement) from our observations, and each tree casts a unit vote for the most popular class to classify an input vector. In our study, we use the random forest classifier to evaluate the performance of three different inputs at each node: (i) text and transactional metrics, (ii) social network metrics, and (iii) the combination of all features. This approach allows us to explore which features are most informative for classification purposes, and generate insights into the underlying patterns in our data.

Because we are dealing with highly imbalanced data, we use a hybrid sampling approach to train the RFC. Specifically, we employ a combination of upsampling the minority class and downsampling the majority class to achieve a balance between the classes. This approach allows us to improve the model's ability to identify the minority class, by mitigating the effects of class imbalance.

7 Model evaluation

To evaluate our models, we split the data into two non-overlapping sets: a 70% classification set and a 30% prediction set. These sets were derived from the final sample after applying the hybrid sampling. To measure the performance of our models in the prediction set, we report the confusion matrix, balanced accuracy, and F1-score. Balanced accuracy is a performance metric that accounts for the imbalance in a dataset and provides a more accurate measure of a model's overall performance by equally weighing each class.

8 Empirical results

Identifying drug-dealers

We present our model's performance for a decision threshold of 0.7 in Table 1, and provide robustness checks using different thresholds in Appendix B. First, we observe that our random forest classifier (RFC) performs very well in predicting both types of illicit Venmo users, namely mixed-activity participants and heavy drug traffickers. Across all sets of input variables, the RFC achieves high levels of balanced accuracy and F1-score, indicating that it is able to accurately classify the data. Notably, social network metrics achieve high predictive accuracy on their own, indicating their potential as a standalone tool for detecting illicit activity on fintech platforms without relying on text data. This finding is particularly valuable as it suggests that we can generalize these results to other fintech platforms that do not require user transactions to provide description messages.

In addition to evaluating the predictive performance of our random forest model, we also conduct a feature importance analysis to determine the relative contributions of each input variable to the model's accuracy. Figure 2 shows the performance metrics for each set of input variables, namely social, text, transactional, and all three combined, for the case of Venmo's mixedactivity participants. We present the feature importance for Venmo's heavy drug traffickers in Fig. 3 of Appendix C.

Table 1 This table presents the performance metrics of our random forest model, which was trained to classify data based on various sets of input variables, including transactional, social, and text data

Metrics	TN	FN	FP	TP	Balanced accuracy	F1
<i>Results for Mixed-Activity Venmo Users</i>						
Social	3037	284	40	2953	0.95	0.95
Text	3047	10	30	3227	0.99	0.99
Transactional	3061	3	16	3234	0.99	0.99
All	3055	3	22	3234	0.99	0.99
<i>Results for Heavy Drug Trafficker Venmo Users</i>						
Social	2844	21	38	2940	0.99	0.98
Text	2876	4	6	2957	0.99	0.99
Transactional	2877	1	5	2960	0.99	0.99
All	2879	3	3	2958	0.99	0.99

A decision threshold of 0.7 was used for this analysis, meaning that only predictions with a probability of 0.7 or higher were classified as positive

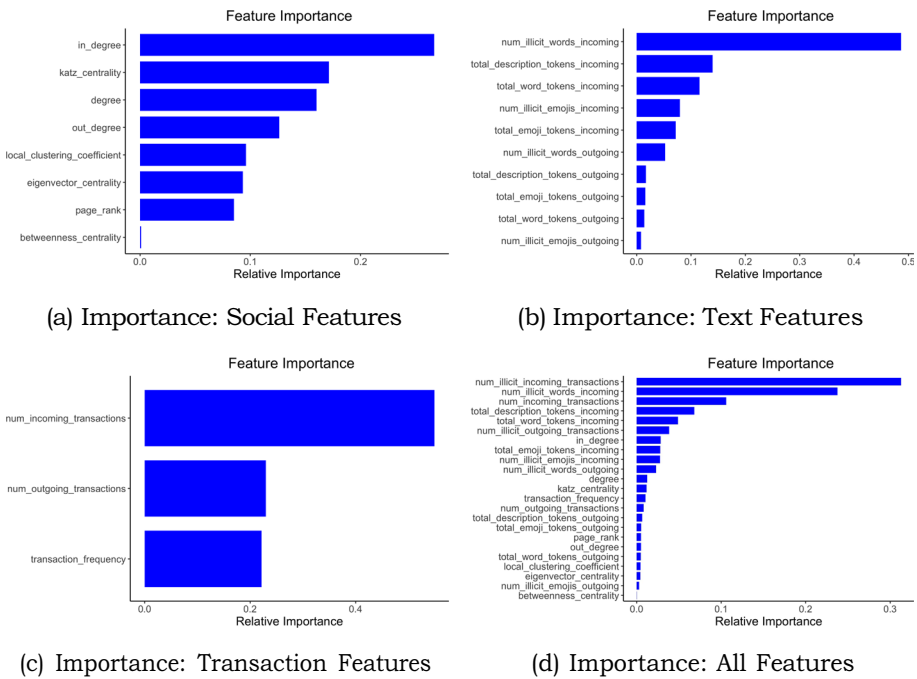


Fig. 2 We present the feature importance for the social, text, and transactional data separately, as well as for all features combined. The importance score measures the contribution of each feature to the performance of the random forest classifier. The higher the score, the more important the feature is for predicting illicit activity

With regard to transactional metrics, our investigation reveals that the number of incoming transactions is the most informative feature. This suggests that users who receive a large number of incoming transactions may be more likely to engage in illicit activities, such as drug trafficking. However, it’s important to note that this feature should be used with caution, as it may not be applicable to other fintech platforms. For example, legitimate business accounts may also receive a high volume of incoming transactions, which could result in false positives if this feature is used to identify illicit activity on other platforms.

In terms of text metrics, our analysis finds that the number of illicit words and emojis were the most informative variables, which is consistent with our expectations. These text-based features can be used as red flags for the identification of potentially illicit transactions and users. However, it’s worth noting that not all fintech platforms use description messages in their transactions, which limits the extrapolation of these findings to other fintech platforms without text-based features.

Last, with respect to network metrics, our analysis reveals that degree centrality, eigenvector centrality, page rank, and local clustering coefficient were the most informative variables for identifying mixed-activity illicit nodes within Venmo’s social network. The fact that these network metrics proved to be informative in identifying illicit users is particularly significant because these measures can be extrapolated to other fintech platforms. Therefore, these findings can help inform the development of more effective strategies for identifying and preventing illicit activities within the financial industry.

9 The network structure of drug-dealing networks

Building on these findings, we further conduct a comparative analysis of the structural social network differences between licit and the two groups of illicit users. The comparison reveals interesting findings, which are presented in Table 2. We report the Cohen effect size instead of t-statistic because the latter may become statistically significant even for small effect sizes in large samples, making it misleading to determine the practical significance of results.

The first group of illicit users, whom we refer to as mixed-activity participants, exhibit several differences in their network structure compared to regular Venmo users. Specifically, mixed-activity participants have significantly higher degree, in-degree, and out-degree than regular Venmo users. Additionally, they display higher Katz and page rank centrality measures and a higher clustering coefficient. The combination of a high clustering coefficient and high page rank and Katz centrality measures among mixed-activity participants indicate that they are accessing multiple networks, bridging the networks through their illicit activities. Therefore, it is critical to study the mixed-activity participants as they might play a pivotal role in facilitating the spread of illicit activity across different groups within the platform.

The second group of illicit users, whom we refer to as heavy drug traffickers, reveal several intriguing findings. First, heavy drug traffickers have much higher degree, in-degree, and out-degree measures than regular Venmo users. Notably, these measures are even higher than those observed for mixed-activity participants. Second, heavy drug traffickers exhibit

Table 2 This table compares the structural social network variables between licit and illicit Venmo users

Social Network variable	Mean of mixed-activity participants (SD)	Mean of licit users (SD)	Cohen's
<i>Results of comparing mixed-activity vs licit venmo users</i>			
Degree	22.13 (22.96)	8.15 (17.24)	0.81
In-degree	12.81 (13.91)	4.07 (14.06)	0.62
Out-degree	9.32 (10.31)	4.07 (6.25)	0.84
Betweenness centrality	1.98×10^{-5} (1.63×10^{-4})	2.43×10^{-6} (3.61×10^{-4})	0.05
Clustering coefficient	0.20 (0.20)	0.16 (0.26)	0.15
Katz	6.65×10^{-4} (7.79×10^{-6})	6.60×10^{-4} (6.68×10^{-6})	0.75
Page rank	8.76×10^{-7} (6.28×10^{-7})	4.36×10^{-7} (8.20×10^{-7})	0.54
Eigenvector centrality	1.06×10^{-4} (5.93×10^{-4})	4.91×10^{-5} (6.58×10^{-4})	0.09
<i>Results of comparing heavy drug trafficker vs licit venmo users</i>			
Degree	43.80 (32.43)	8.15 (17.24)	2.07
In-degree	27.48 (22.55)	4.07 (14.06)	1.66
Out-degree	16.31 (12.08)	4.07 (6.25)	1.96
Betweenness centrality	2.53×10^{-5} (1.86×10^{-4})	2.43×10^{-6} (3.61×10^{-4})	0.06
Clustering coefficient	0.14 (0.11)	0.16 (0.26)	0.08
Katz	6.72×10^{-4} (1.17×10^{-5})	6.60×10^{-4} (6.68×10^{-6})	1.80
Page rank	1.42×10^{-6} (7.86×10^{-7})	4.36×10^{-7} (8.20×10^{-7})	1.20
Eigenvector centrality	1.03×10^{-4} (2.17×10^{-4})	4.91×10^{-5} (6.58×10^{-4})	0.08

We report the average of each group and the Cohen effect size from their comparison

Standard deviations are in parentheses. Cohen (2013) considers a value of 0.2 to be a “small” effect size, 0.5 represents a “medium” effect size, and 0.8 and above a “large” effect size

significantly higher Katz and page rank centrality measures compared to both regular Venmo users and mixed-activity participants. This suggests that heavy drug traffickers have a greater ability to spread illicit activities throughout the Venmo network, as they are able to connect with a larger number of users and potentially influence their behavior. Last, in contrast to the mixed-activity participants, heavy drug traffickers have a lower clustering coefficient than regular Venmo users. This lower clustering coefficient suggests that heavy drug traffickers do not form tightly-knit social circles on the platform, but rather maintain a more diverse set of connections.

Overall, these results highlight the importance of studying the two different types of illicit users on Venmo's platform. Mixed-activity participants and heavy drug traffickers exhibit different patterns of network structure characteristics compared to regular Venmo users, suggesting that they may be playing distinct roles in the spread of illicit activities on the platform. By better understanding the network structure differences between these groups, law enforcement and fraud detection efforts can develop more targeted interventions to prevent and disrupt illicit activity on platforms.

10 Discussion

In the realm of crime analytics, digital platforms have become the new frontier for illegal drug trafficking, with legitimate businesses often acting as inadvertent catalysts for criminal innovation (Rawat et al., 2022; Storrod & Densley, 2017; Tremblay et al., 2001). While the role of social network platforms in this domain has been explored, understanding the part played by mobile payment apps in these illicit activities remains largely uncharted (Hu et al., 2023; Yang & Luo, 2017). This study bridges this gap by examining the financial footprint of drug-related transactions on Venmo, a mobile payment app that mandates comments with every transaction. The insights obtained, which align with studies highlighting the enabling role of technology in scaling illicit activities (Barone & Masciandaro, 2011; Embar-Seddon, 2002; Holt & Bossler, 2015), spotlight mobile payment apps as facilitators of drug transactions, necessitating further research and regulatory scrutiny.

Our work contributes to crime analytics by extending the understanding how digital platforms are leveraged for illegal activities. We shed light on the existence and unique behaviors of two specific illicit user groups—mixed-activity participants and heavy drug traffickers. This not only broadens the scope of online deviance by illuminating the heterogeneous nature of digital offenders, but it also serves as a crucial stepping stone in the development of refined theoretical models of illicit digital behavior. Additionally, we reveal the impact of fintech disruption on traditional drug markets, leading to unique social structures, particularly amongst mixed-activity participants (Morselli & Giguere, 2006; Van der Hulst, 2009; Von Lampe, 2015). These insights deepen our understanding of how digital transformation affects traditional crime structures.

Methodologically, this study underscores the efficacy of a holistic data approach, combining text, transactional, and social network data, for robust and accurate results. This approach, which could be applied and refined across digital platforms, opens the gateway to advanced data mining practices (Bogdanov et al., 2011; Chen et al., 2004). Given the vast availability of diverse data types and the evolution of sophisticated data fusion techniques (Lim et al., 2020; Meng et al., 2020), integrating various data sources into a comprehensive analytical framework could significantly enhance the understanding and mitigation of digital illicit activities.

In addition to its theoretical and methodological contributions, our research carries substantial implications for mobile payment industry operators and law enforcement agencies. The study stresses the importance of developing robust strategies to identify and prevent illicit activities on their platforms. Businesses could, for example, invest in sophisticated algorithms and machine learning models to detect suspicious transactions and network structures. Meanwhile, law enforcement agencies can benefit from the study's findings to identify suspicious networks related to drug transactions, even without transaction descriptions, aiding in the proactive prevention of illicit drug activities and enhancing public safety.

11 Conclusions

In conclusion, our study has shed light on the complex interplay between illicit activities and fintech digital platforms, and the potential of machine learning and social network analysis techniques in tackling these challenges.

In the initial phase of our study, we employ text analytics to detect transactions involving illicit drugs. However, identifying illicit nodes accurately through text analytics alone is extremely challenging due to the prevalence of sarcastic comments made by Venmo users. To address this challenge, we first employ a heuristic rule to flag potential illicit nodes, providing a preliminary list of suspicious users. In order to further differentiate between truly illicit users and those merely using sarcastic remarks, we undertake a manual annotation of these flagged users. This combined approach ensures the validity of our labels. Having the labeled data at our disposal, we train a supervised random forest classifier, which leverages text, transactional, and social network data to further investigate financial activity related to drug dealing. Our random forest model is able to identify and predict both mixed-activity participants and heavy drug traffickers on Venmo.

Motivated by the above findings, we further explore the structural social network differences between licit users and the two groups of illicit users on Venmo's platform. The first group of illicit users, mixed-activity participants, have a significantly higher degree, Katz and page rank centrality measures, and clustering coefficient than regular Venmo users. On the other hand, the second group, heavy drug traffickers, exhibit higher degree measures and Katz and page rank centrality measures compared to both regular Venmo users and mixed-activity participants. However, heavy drug traffickers display a lower clustering coefficient than regular Venmo users, indicating that they do not form tightly-knit social circles. These findings suggest that both types of illicit users play distinct roles in the spread of illicit activities on the platform. Therefore, gaining a better understanding of their networks' structural characteristics could aid in devising more tailored interventions to prevent and combat such activities.

In future research, it would be worthwhile to expand the analysis of drug-related networks beyond their steady-state and investigate their dynamic evolution over time. Examining the mechanisms by which these networks are created and expanded, we could gain a deeper understanding of the factors that facilitate the growth of illicit activities within fintech platforms. Additionally, it would be interesting to explore whether our findings on drug-related Venmo transactions generalize to other similar fintech settings, such as cryptocurrency tokens. By examining the prevalence of illicit activities on other fintech platforms, we could develop a better understanding of the challenges associated with identifying and preventing illegal transactions in the digital economy. This could ultimately inform the development of more effective regulatory and enforcement mechanisms to combat illicit financial activities.

Acknowledgements We thank the guest editor, Michel Fathi, and all anonymous referees for their constructive comments and suggestions that greatly enhanced the clarity of this article. We also want to extend our heartfelt appreciation to our students—Yuntong Xiao, Ayan Ghosh, Jeet Patel, and Ethan Meng—who contributed to the early stage of this research project.

Funding No funding was received for conducting this study.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval The authors declare compliance with the ethical standards. This research does not involve Human Participants and/or Animals.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Variable Description

See Table 3.

Table 3 Variable description

Variable type	Variable name	Description
Text	Number of illegal words incoming	Total number of illegal words in the transaction descriptions received from other users
	Number of Illegal emojis incoming	Total number of illegal emojis in the transaction descriptions received from other users
	Total description tokens incoming	Total words and emojis in the transaction descriptions received from other users
	Total emoji tokens incoming	Total emojis in the transaction descriptions received from other users
	Total word tokens incoming	Total words in the transaction descriptions received from other users
	Number of illegal incoming transactions	Total number of transactions including at least one illegal word or emoji received from other users
	Number of Illegal words outgoing	Total number of illegal words in the transaction descriptions sent to other users

Table 3 (continued)

Variable type	Variable name	Description
	Number of illegal emojis outgoing	Total number of illegal emojis in the transaction descriptions sent to other users
	Total description tokens outgoing	Total words and emojis in the transaction descriptions sent to other users
	Total emoji tokens outgoing	Total emojis in the transaction descriptions sent to other users
	Total word tokens outgoing	Total words in the transaction descriptions sent to other users
	Number of illegal outgoing transactions	Total number of transactions including at least one illegal word or emoji sent to other users
Transaction	Number of incoming transactions	Total number of transactions received from other users
	Number of outgoing transactions	Total number of transactions sent to other users
	Transaction frequency	The total number of transactions made by the user divided by their lifetime on the platform
Social network	Degree	Counts the number of distinct users with whom the user has conducted transactions
	InDegree	Counts the number of distinct users from whom the user has received money
	OutDegree	Counts the number of distinct users to whom the user has sent money
	Local clustering coefficient	The local clustering coefficient of a node counts how many of the possible edges between its neighbors exist
	Betweenness centrality	Counts the number of shortest paths between all pairs of nodes in the network that pass through the focal node
	Katz centrality	It is the sum of the centrality scores of all nodes that can be reached from a node, with a discount factor applied to more distant nodes
	Page rank	It captures the significance of a node by taking into account its connections, and gives more weight to links from nodes that already have high importance scores

Appendix B: Robustness checks

We present in Table 4 the robustness checks of our random forest classifier by varying the decision threshold and re-evaluating the model's performance metrics. Overall, we find that the performance of the model is relatively consistent across a range of decision thresholds.

While the exact values of the performance metrics may vary slightly depending on the threshold chosen, the general trends and patterns remain the same.

Table 4 This table presents the performance metrics of our random forest model, which was trained to classify data based on various sets of input variables, including transactional, social, and text data

Metrics	Threshold	TN	FN	FP	TP	Balanced accuracy	F1
<i>Results for mixed-activity venmo users</i>							
Social	0.5	2933	112	144	3125	0.96	0.96
Text	0.5	3033	3	44	3234	0.99	0.99
Transactional	0.5	3047	0	30	3237	0.99	0.99
All	0.5	3039	0	38	3237	0.99	0.99
Social	0.8	3062	537	15	2700	0.93	0.93
Text	0.8	3055	20	22	3217	0.99	0.99
Transactional	0.8	3064	16	13	3221	0.99	0.99
All	0.8	3065	7	12	3230	0.99	0.99
Social	0.9	3074	1047	3	2190	0.87	0.88
Text	0.9	3070	75	7	3162	0.97	0.97
Transactional	0.9	3069	38	8	3199	0.99	0.99
All	0.9	3073	43	4	3194	0.99	0.99
<i>Results for heavy trafficker venmo users</i>							
Social	0.5	2795	6	87	2955	0.98	0.98
Text	0.5	2869	3	13	2958	0.99	0.99
Transactional	0.5	2873	1	9	2960	0.99	0.99
All	0.5	2874	0	8	2961	0.99	0.99
Social	0.8	2865	59	17	2902	0.98	0.98
Text	0.8	2877	5	5	2956	0.99	0.99
Transactional	0.8	2879	5	3	2956	0.99	0.99
All	0.8	2881	4	1	2957	0.99	0.99
Social	0.9	2875	176	7	2785	0.96	0.96
Text	0.9	2879	15	3	2946	0.99	0.99
Transactional	0.9	2880	16	2	2945	0.99	0.99
All	0.9	2882	8	0	2953	0.99	0.99

We present in Table 5 a robustness check for an alternative model specification. Instead of the previously used random forest classifier, we have implemented a logistic regression model. Despite slight variations in performance metrics, the overall patterns of the results remain unchanged.

Table 5 This table presents the performance metrics of a logistic regression model, which was trained to classify data based on various sets of input variables, including transactional, social, and text data

Metrics	TN	FN	FP	TP	Balanced Accuracy	F1
<i>Results for mixed-activity venmo users</i>						
Social	2842	2114	235	1123	0.64	0.71
Text	3018	144	59	3093	0.96	0.96
Transactional	2878	2127	199	1110	0.64	0.71
All	3037	86	40	3151	0.98	0.97
<i>Results for heavy drug trafficker venmo users</i>						
Social	2706	818	176	2143	0.83	0.84
Text	2848	88	34	2873	0.97	0.97
Transactional	2722	815	160	2146	0.83	0.84
All	2855	1	27	2960	0.99	0.99

A decision threshold of 0.7 was used for this analysis, meaning that only predictions with a probability of 0.7 or higher were classified as positive

Appendix C: Feature importance for heavy drug traffickers

See Fig. 3.

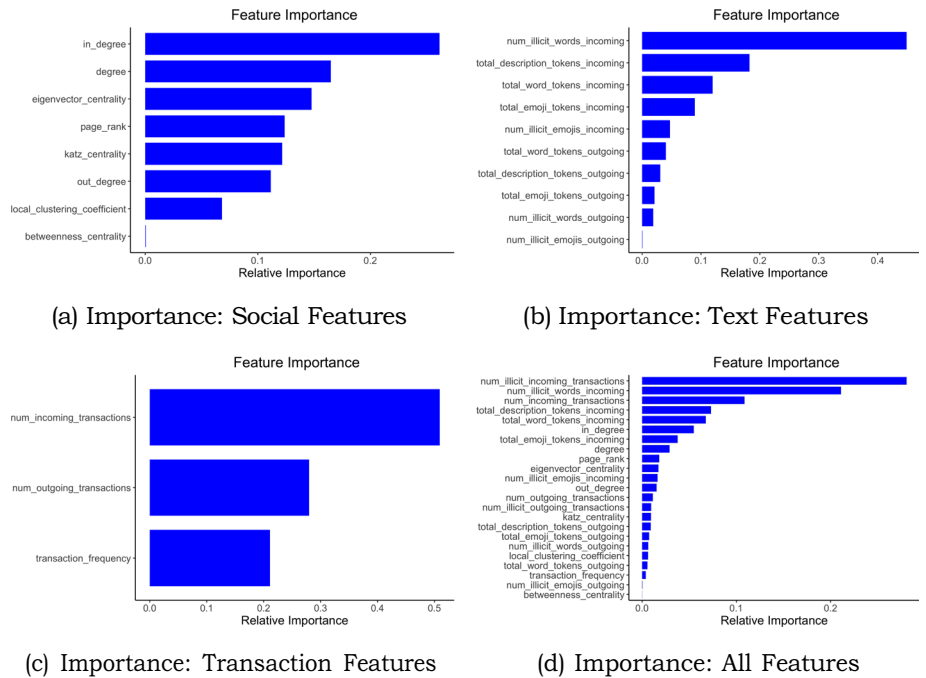


Fig. 3 We present the feature importance for the social, text, and transactional data separately, as well as for all features combined, for the heavy drug trafficker Venmo users. The importance score measures the contribution of each feature to the performance of the random forest classifier. The higher the score, the more important the feature is for predicting illicit activity

References

- Bakken, S. A. (2021). Drug dealers gone digital: Using signalling theory to analyse criminal online personas and trust. *Global Crime*, 22(1), 51–73.
- Barone, R., & Masciandaro, D. (2011). Organized crime, money laundering and legal economy: Theory and simulations. *European Journal of Law and Economics*, 32, 115–142.
- Bhargava, H. K., Rubel, O., Altman, E. J., et al. (2020). Platform data strategy. *Marketing Letters*, 31, 323–334.
- Bodoh-Creed, A., Boehnke, J., & Hickman, B. (2018). Using machine learning to predict price dispersion. Tech. rep., Working paper.
- Bogdanov, P., Mongiovi, M., & Singh, A. K. (2011). Mining heavy subgraphs in time-evolving networks. In *2011 IEEE 11th international conference on data mining* (pp. 81–90). IEEE.
- Bracci, A., Boehnke, J., ElBahrawy, A., et al. (2022). Macroscopic properties of buyer–seller networks in online marketplaces. *PNAS Nexus*, 1(4), pgac201.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Van de Bunt, H., Siegel, D., Zaitch, D., et al. (2014). The social embeddedness of organized crime. Mimeo, Buntain, C., & Golbeck, J. (2015). This is your twitter on drugs: Any questions? In *Proceedings of the 24th international conference on World Wide Web* (pp. 777–782).
- Chen, H., Chung, W., Xu, J. J., et al. (2004). Crime data mining: A general framework and some examples. *Computer*, 37(4), 50–56.
- Cherian, R., Westbrook, M., Ramo, D., et al. (2018). Representations of codeine misuse on instagram: Content analysis. *JMIR Public Health and Surveillance*, 4(1), e8144.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge Academic.
- Cook, K. S., Emerson, R. M., Gillmore, M. R., et al. (1983). The distribution of power in exchange networks: Theory and experimental results. *American Journal of Sociology*, 89(2), 275–305.
- Demant, J., Bakken, S. A., & Hall, A. (2020). Social media markets for prescription drugs: Platforms as virtual mortars for drug types and dealers. *Drugs and Alcohol Today*, 20(1), 36–49.
- Demant, J., Bakken, S. A., Oksanen, A., et al. (2019). Drug dealing on facebook, snapchat and instagram: A qualitative analysis of novel drug markets in the nordic countries. *Drug and Alcohol Review*, 38(4), 377–385.
- Desroches, F. J. (2005). *The crime that pays: Drug trafficking and organized crime in Canada*. Canadian Scholars' Press.
- Du, P. Y., Ebrahimi, M., Zhang, N., et al. (2019). Identifying high-impact opioid products and key sellers in dark net marketplaces: An interpretable text analytics approach. In *2019 IEEE international conference on intelligence and security informatics (ISI)* (pp. 110–115). IEEE.
- Embar-Seddon, A. (2002). Cyberterrorism: Are we under siege? *American Behavioral Scientist*, 45(6), 1033–1043.
- Framis, A., & Morselli, C. (2014). Illegal networks or criminal organizations: Structure, power, and facilitators in cocaine trafficking structures. In *Crime and networks* (pp. 131–147).
- Framis, A. G. S., & Regadera, S. F. (2017). Static and dynamic approaches of a drug trafficking network. In *Crime prevention in the 21st century: insightful approaches for crime prevention initiatives* (pp. 187–211).
- Freeman, R. B. (1999). The economics of crime. *Handbook of Labor Economics*, 3, 3529–3571.
- Hagedorn, J. M. (1994). Homeboys, dope fiends, legits, and new jacks. *Criminology*, 32(2), 197–219.
- Haupt, M. R., Cuomo, R., Li, J., et al. (2022). The influence of social media affordances on drug dealer posting behavior across multiple social networking sites (SNS). *Computers in Human Behavior Reports*, 8(100), 235.
- Holt, T. J., & Bossler, A. M. (2015). *Cybercrime in progress: Theory and prevention of technology-enabled offenses*. Routledge.
- Hu, C., Yin, M., Liu, B., et al. (2021). Detection of illicit drug trafficking events on instagram: A deep multimodal multilabel learning approach. In *Proceedings of the 30th ACM international conference on information & knowledge management* (pp 3838–3846).
- Hu, C., Liu, B., Ye, Y., et al. (2023). Fine-grained classification of drug trafficking based on instagram hashtags. *Decision Support Systems*, 165(113), 896.
- Kleemans, E. R. (2007). Organized crime, transit crime, and racketeering. *Crime and Justice*, 35(1), 163–215.
- Kleemans, E. R., & Van de Bunt, H. G. (1999). The social embeddedness of organized crime. *Transnational Organized Crime*, 5(1), 19–36.
- Klein, M. W., & Crawford, L. Y. (1967). Groups, gangs, and cohesiveness. *Journal of Research in Crime and Delinquency*, 4(1), 63–75.
- Klein, M. W., & Maxson, C. L. (2010). *Street gang patterns and policies*. Oxford University Press.
- Levitt, S. D., & Venkatesh, S. A. (2000). An economic analysis of a drug-selling gang's finances. *The Quarterly Journal of Economics*, 115(3), 755–789.

- Li, J., Xu, Q., Shah, N., et al. (2019). A machine learning approach for the detection and characterization of illicit drug dealers on instagram: Model evaluation study. *Journal of Medical Internet Research*, *21*(6), e13,803.
- Lim, M., Abdullah, A., & Jhanjhi, N. (2020). Data fusion-link prediction for evolutionary network with deep reinforcement learning. *International Journal of Advanced Computer Science and Applications*, *11*(6), 335–342.
- Lossio-Ventura, J.A., & Bian, J. (2018). An inside look at the opioid crisis over twitter. In *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)* (pp. 1496–1499). IEEE.
- Loupos, P., & Nathan, A. (2019). The structure and evolution of an offline peer-to-peer financial network. In *International workshop on complex networks* (pp. 113–122). Springer.
- Mackey, T. K., Kalyanam, J., Katsuki, T., et al. (2017). Twitter-based detection of illegal online sale of prescription opioid. *American Journal of Public Health*, *107*(12), 1910–1915.
- Mackey, T., Kalyanam, J., Klugman, J., et al. (2018). Solution to detect, classify, and report illicit online marketing and sales of controlled substances via twitter: using machine learning and web forensics to combat digital opioid access. *Journal of Medical Internet Research*, *20*(4), e10,029.
- Malm, A., & Bichler, G. (2011). Networks of collaborating criminals: Assessing the structural vulnerability of drug markets. *Journal of Research in Crime and Delinquency*, *48*(2), 271–297.
- McGloin, J. M. (2005). Policy and intervention considerations of a network analysis of street gangs. *Criminology & Public Policy*, *4*(3), 607–635.
- Meng, T., Jing, X., Yan, Z., et al. (2020). A survey on machine learning for data fusion. *Information Fusion*, *57*, 115–129.
- Morselli, C. (2010). Assessing vulnerable and strategic positions in a criminal network. *Journal of Contemporary Criminal Justice*, *26*(4), 382–392.
- Morselli, C., & Giguere, C. (2006). Legitimate strengths in criminal networks. *Crime, Law and Social Change*, *45*(3), 185–200.
- Morselli, C., Paquet-Clouston, M., & Provost, C. (2017). The independent's edge in an illegal drug distribution setting: Levitt and Venkatesh revisited. *Social Networks*, *51*, 118–126.
- Moyle, L., Childs, A., Coomber, R., et al. (2019). # drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs. *International Journal of Drug Policy*, *63*, 101–110.
- Natarajan, M. (2000). Understanding the structure of a drug trafficking organization: A conversational analysis. *Crime Prevention Studies*, *11*, 273–298.
- Natarajan, M. (2006). Understanding the structure of a large heroin distribution network: A quantitative analysis of qualitative data. *Journal of Quantitative Criminology*, *22*(2), 171–192.
- Papachristos, A. V. (2006). Social network analysis and gang research: Theory and methods. In *Studying youth gangs* (pp 99–116).
- Pergolizzi, J., Jr., LeQuang, J., Taylor, R., Jr., et al. (2017). The “darknet”: The new street for street drugs. *Journal of Clinical Pharmacy and Therapeutics*, *42*(6), 790–792.
- Potamias, R. A., Siolas, G., & Stafylopatis, A. G. (2020). A transformer-based approach to irony and sarcasm detection. *Neural Computing and Applications*, *32*(23), 17,309–17,320.
- Qian, Y., Zhang, Y., Ye, Y., et al. (2021). Distilling meta knowledge on heterogeneous graph for illicit drug trafficker detection on social media. *Advances in Neural Information Processing Systems*, *34*, 26,911–26,923.
- Rawat, R., Mahor, V., Chouhan, M., et al. (2022). Systematic literature review (SLR) on social media and the digital transformation of drug trafficking on darkweb. In *International conference on network security and blockchain technology* (pp. 181–205), Springer.
- Reuter, P., & Haaga, J. (1989). The organization of high-level drug markets: An exploratory study.
- Reuter, P., MacCoun, R., Murphy, P., et al. (1990). Money from crime: A study of the economics of drug dealing in Washington, DC. Rand Santa Monica, CA.
- Roy, A., Paul, A., Pirsiavash, H., et al. (2017). Automated detection of substance use-related social media posts based on image and text analysis. In *2017 IEEE 29th international conference on tools with artificial intelligence (ICTAI)* (pp. 772–779). IEEE.
- Sequeira, R., Gayen, A., Ganguly, N., et al. (2019). A large-scale study of the twitter follower network to characterize the spread of prescription drug abuse tweets. *IEEE Transactions on Computational Social Systems*, *6*(6), 1232–1244.
- Storrod, M. L., & Densley, J. A. (2017). ‘going viral’ and ‘going country’: The expressive and instrumental activities of street gangs on social media. *Journal of Youth Studies*, *20*(6), 677–696.
- Tofighi, B., Aphinyanaphongs, Y., Marini, C., et al. (2020). Detecting illicit opioid content on twitter. *Drug and Alcohol Review*, *39*(3), 205–208.

- Tremblay, P., Talon, B., & Hurley, D. (2001). Body switching and related adaptations in the resale of stolen vehicles. Script elaborations and aggregate crime learning curves. *British Journal of Criminology*, 41(4), 561–579.
- Van der Hulst, R. C. (2009). Introduction to social network analysis (SNA) as an investigative tool. *Trends in Organized Crime*, 12, 101–121.
- Von Lampe, K. (2015). *Organized crime: Analyzing illegal activities, criminal structures, and extra-legal governance*. Sage Publications.
- Wright, A. P., Jones, C. M., Chau, D. H., et al. (2021). Detection of emerging drugs involved in overdose via diachronic word embeddings of substances discussed on social media. *Journal of Biomedical Informatics*, 119(103), 824.
- Yang, X., & Luo, J. (2017). Tracking illicit drug dealing and abuse on instagram using multimodal analysis. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(4), 1–15.
- Zaitch, D. (2002). *Trafficking cocaine: Colombian drug entrepreneurs in the Netherlands* (Vol. 1). Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.