Decision Support Systems 165 (2023) 113896

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Fine-grained classification of drug trafficking based on Instagram hashtags

Chuanbo Hu [a], Bin Liu [b], Yanfang Ye [c], Xin Li [a,*]

[a] Lane Department of Computer Science and Electrical Engineering, West Virginia University, PO Box 6109, Morgantown, WV 26506-6109, USA
[b] Department of Management Information System, West Virginia University, PO Box 6025, Morgantown, WV 26506-6025, USA
[c] Department of Computer Science and Engineering, University of Notre Dame, 323B Cushing Hall of Engineering, Notre Dame, IN 46556, USA

ARTICLE INFO

Keywords:
Drug trafficking
Hashtags
Semi-supervised learning
BERT
Graph convolutional network

ABSTRACT

Social networks have become important platforms for the marketing and sale of illicit drugs. Hashtags make it easier for users to engage in drug trafficking, further increasing the risk of drug abuse. However, there are significant challenges in the detection and management of drug trafficking activities. In addition, the rapid legalization of some drugs has required a fine-grained classification of drugs to distinguish them from those that are illegal. Motivated by these observations, in this paper, our aim is to develop a methodology using the latest advances in AI technology to classify hashtags from posts advertising illicit drugs for sale on social networks. We present a semi-supervised deep learning approach to classify hashtags from posts advertising illicit drugs. An elegant combination of Bidirectional Encoder Representations from Transformers (BERT) with Graph Convolutional Network (GCN) allows us to analyze the characteristics (e.g., shipping region and platform self-regulation) of illegal drug trafficking. Our BERT+GCN model achieved the best performance with more than 75% accuracy compared to the other three baseline models. Then, fine-grained hashtags identified are applied to explore the characteristics of drug trafficking. Finally, we report our results for further exploration of shipping regions and self-regulation of drug trafficking on the platform in two analysis scenarios. Our developed approach has shown its effectiveness in detecting hashtags for different types of drugs from illegal drug sellers. Based on hashtag classification, we also provide two case studies that indicate that (1) there are differences in self-regulation for different types of drugs on social media, (2) there are regional differences in the demand for different types of drugs.

1. Introduction

Drug overdoses pose a major threat to public health in the United States. A recent study reported that more than 100,000 people died from drug overdoses during the 12 months ending April 2021 [1]. With overdose deaths increasing 28.5% from the same period a year earlier and almost doubled in the last five years, opioids continue to be the driving cause of drug overdose deaths. Synthetic opioids caused almost 64% of all drug-related overdose deaths, up 49% from the year before, according to the CDC National Center for Health Statistics. Federal data show that deaths by overdose of methamphetamine and other psychostimulants also increased significantly, 48% in the year ending April 2021 compared to the year before.

Illegal online drug trade (a.k.a. drug trafficking), as an important trigger for the drug overdose crisis, has aroused social attention [2,3]. The sale of unapproved and misbranded drugs (such as opioids) has posed increased dangers to consumers who purchase those products over the Internet (e.g., Purdue Pharma. Lawsuit). Unlike drugs approved by the FDA, there has been no FDA evaluation of whether unapproved products are safe and effective for their intended use or if they have dangerous side effects or other safety concerns [4]. Social networks such as Instagram, Facebook, and Twitter have become convenient direct-to-consumer marking tools for online drug trafficking [5–7]. Marijuana, prescription painkillers, Xanax, Molly (MDMA), and lean (codeine syrup mixture) are the most popular drugs on Instagram for sale [8].

Detecting online illicit drug trafficking has become a critical step in combating the online trade of illicit drugs. However, with the rapid proliferation and diversification of the Internet ecosystem, the detection of illegal drug sellers online, including social media posts and dark web providers, has become a challenge [9,10]. In the literature, machine learning methods have been applied to detect advertisements for illicit drug sales and drug dealer accounts on popular social media platforms, including Twitter, Facebook, and Instagram [3,6,8,11]. The development of multimodal fusion technology provides a series of solutions

* Corresponding author.
E-mail address: xin.li@ieee.org (X. Li).

https://doi.org/10.1016/j.dss.2022.113896
Received 18 February 2022; Received in revised form 10 September 2022; Accepted 2 November 2022
Available online 13 November 2022

[12–14]. However, these studies focus on mining the semantic relationship between image objects and are not applicable to drug trafficking detection tasks. Additionally, text information on social networks provides more information on illegal drug trafficking [6], compared to image-based cyber attacks [15–17]. More recently, a deep learning approach was proposed to detect illicit drug trafficking events from posts or comments on posts on Instagram [18]. However, since an ad for illicit drug sales is constantly generated, it is often difficult to dynamically adapt the model to detect an ad for drug dealers.

A promising remedy is to develop a hashtag-based methodology, because hashtags have been widely adopted to indicate a specific topic and spread important information on social networks and microblogging platforms [19]. Similarly, drug dealers use hashtags to promote the marketing and sale of illicit drugs, increasing the risk of drug abuse in the public [6,8]. For the public, hashtags are used as queries on social media platforms (such as Instagram) to search for relevant posts. As such, the detection and management of drug-related hashtags is arguably the most effective way to combat drug trafficking. However, it is challenging to detect drug-related hashtags on social media for the following reasons. First, hashtags by nature are ambiguous in different contexts and evolve to several synonyms to avoid being detected. Second, a single post can contain multiple hashtags, and those hashtags are used for different purposes. For example, some of them are related to drugs, and others are not. Third, regional differences in drug legalization, such as the legalization of marijuana in Canada and some states in the United States [20], require a fine-grained and policy-agnostic classification of drug-related hashtags.

In this paper, we present a system for fine-grained and policy-agnostic classification of drug-related hashtags in posts or comments on Instagram. The system classifies hashtags into fine-grained drug types (e.g., marijuana drugs, opioid drugs, club drugs) so that the classification can be integrated into decision support systems to monitor drug trafficking activities on social media platforms. We formulate it as a semi-supervised learning problem in which a small subset of the hashtags is annotated with labels. To this end, we propose a method that combines bidirectional encoder representations from transformers (BERT) and graph-convolutional network (GCN) to infer the labels for the unlabeled hashtags. On the one hand, GCN is applied to capture the structure relationships among all hashtags so that label information can effectively propagate to unlabeled hashtags. On the other hand, BERT is powerful in capturing the semantic relationships between hashtags from posts or comments. The experimental results show that the proposed method outperforms the baseline methods with an accuracy greater than 75%. Our technical contributions are summarized below.

- We provide a systematic study of the fine-grained classification of drug-related hashtags in posts or comments on Instagram.
- We formulate hashtag classification as a semi-supervised deep learning problem, and we propose a method that combines BERT and GCN to infer the labels for the unlabeled hashtags.
- We demonstrate the effectiveness of the proposed method in hashtag classification using real-world datasets that we collected from Instagram.
- Based on the classification of hashtags, we also provide two case studies that indicate (1) that there are differences in self-regulation for different types of drugs on Instagram, (2) there are regional differences in the demand for different types of drugs.

2. Literature review

Illicit drug networks have evolved along with rapid advances in modern information and communication technology (ICT) [8]. The marriage of illicit supply networks with ICT, including darknet cryptomarkets and surface net social networks, has offered a shared platform for illicit drug trafficking with a lower degree of risk. In addition to anonymity, the ease of online advertising and stealth delivery have made the option of trading illicit drugs through social media platforms such as Instagram more desirable than conventional street markets. Understanding the operations and dynamics of illicit supply networks has presented a great new challenge for both data mining and decision-making communities.

To make informed decisions in combating illicit drug trafficking, early research work has focused primarily on offline data, including offender databases [21,22], law enforcement records of electronic and/or physical surveillance (e.g., wiretap transcripts) [23,24], and transcripts of court proceedings [25]. However, due to rapid advances in ICT in recent decades, newly formed, but exponentially increasing, illicit drug markets have presented new challenges to academics, governments, and law enforcement entities to address the growing online participation in drug trafficking and sales. Sifting out drug-trafficking-related activities from the astronomical amount of social media data is like finding a needle in a haystack.

2.1. Drug trafficking analysis from online data

In the existing literature, there has been limited work on tracking drug abuse and illicit drug trade from online data. Among these existing works, [26] analyzed the time and location patterns of drug use by mining Twitter data. This line of research was recently advanced by [27] for the detection of opioid use disorders (OUD). The network information of the Instagram user timelines was used in [28] to monitor suspicious drug interaction activities; [11,29] analyzed the Instagram data to track and identify drug dealer accounts. More recently, machine learning and natural language processing techniques have been applied to combat prescription drug abuse [30,31] and detect drug dealers [6]. Unlike previous work that used only one mode data (e.g., text data), [8,18] applied multimodal data fusion to identify illicit drug dealers, and [32] combined image and text data to identify the risk of substance use. Our work is different in two ways: First, our research goal is to identify illicit drug dealers, which is more challenging because illicit drug trafficking methods have become sophisticated on social media platforms. Second, our multimodal data fusion is based on a novel quadruple-based data representation (image vs. text, post vs. homepage).

2.2. Social media data mining

Social networks have greatly facilitated the generation and sharing of information through virtual communities and networks. The data associated with popular social media platforms have grown at exponential rates. In recent decades, social media data mining [33] has evolved rapidly. In particular, our work is related to the use of hashtags on social networks that have enabled cross-referencing of content. The use of hashtags has been exploited to study gender differences [34], education and marketing [35], and public health [36]. Unlike images and texts, hashtags are neither registered nor controlled by any user or group of users, making them a more persistent marker for analyzing the trend and users on social networks. To our knowledge, no previous work has been done on the use of hashtags to expedite fine-grained analysis of drug trafficking activities on Instagram. In this work, we will present the first attempt to leverage the latest advances in deep learning for hashtag-based characterization of illegal drugs from Instagram data.

3. Methods

3.1. Problem formulation

In this paper, our objective is to design a system for the fine-grained and policy-agnostic classification of drug-related hashtags in posts or comments on Instagram. The system classifies hashtags into fine-grained drug types (e.g., marijuana drugs, opioid drugs, club drugs) so that the classification can be integrated into decision support systems to monitor

drug trafficking activities on social media platforms. Specifically, let $\mathscr{D} = \{p_1, p_2, \ldots, p_T\}$ be a collection of posts or comments on Instagram and $\mathscr{H} = \{h_1, h_2, \ldots, h_N\}$ be the set of associated hashtags. Let there be a total of K drug types $\{1, 2, \ldots, K\}$. Among the hashtags $\mathscr{H}$, a subset $\mathscr{H}_l = \{h_1, \ldots, h_l\} \in \mathscr{H}$ has been annotated with labels $\{y_1, \ldots, y_l\}$ with each label $y_i \in \{1, 2, \ldots, K\}$, and the rest of the hashtags $\mathscr{H}_u = \{h_{l+1}, \ldots, h_N\} \in \mathscr{H}$ are not labeled. Our goal is to build a machine learning model to classify unlabeled hashtags $\mathscr{H}_u$ by exploring the relationships between all hashtags $\mathscr{H}$ and the associated context information in comments or post data $\mathscr{D}$.

3.2. Overview of the proposed framework

As shown in Fig. 1, our proposed framework consists of three stages: data collection, data processing, and model training. First, we collect and annotate a set of hashtags from drug dealer accounts through the designed data collection and annotation system. Then, a graph network is constructed from drug-related hashtags. Finally, a BERT+GCN model is constructed for the classification of multiple labeled drug-related hashtags.

3.3. Data collection and annotation

Data collection and annotation is an important part of this study. To efficiently collect representative data, we have designed a web scraper to collect posts from Instagram. Fig. 2 shows the architecture of the system for data collection and annotation. As hashtags (e.g., #drug) are widely used for post-search on Instagram, we chose an initial list of hashtags related to the most popular drug names following research by the American Addiction Center [8]. Drug names include marijuana, codeine, MDMA, Xanax, painkillers, mushrooms, LSD, and cocaine.

Then domain experts annotate drug-related hashtags from drug dealer comments based on predefined fine-grained classification schema. The labeling schema contains four non-overlapping drug categories, including marijuana, opioids, club drugs, and other drugs. We design the labeling schema for the following reasons: 1) the rapid legalization of marijuana in the United States makes its management different from other drugs; 2) Opioid (e.g. codeine, fentanyl, and oxycodone) is an important driving factor that causes drug overdose [37]; 3) Most Instagram users are young people aged 18–34 years, these people may have more opportunities to contact club drugs [38] (e.g. MDMA, methamphetamine, and LSD) than other age groups [39]. Furthermore, we uniformly classify the remaining drugs into the fourth category (e.g., Xanax, Adderall, and cocaine).

3.4. Proposed approach to hashtag classification

Since annotating massive label data is time-consuming, we propose applying a semi-supervised graph-based deep learning approach to hashtag classification. To analyze the relationship between drug dealers' hashtags on Instagram, we propose a graph-based deep learning model by combining the pre-trained Bidirectional Encoder Representations of Transformers (BERT) model [40] and Graph Convolutional Networks (GCN) [41]. BERT is a deeply bidirectional and unsupervised language representation that is pre-trained using a plain text corpus. GCN [41] is an effective deep learning architecture to capture node relationships in graph data. On the one hand, GCN is applied to capture the structure relationships among all hashtags so that label information can effectively propagate to unlabeled hashtags. On the other hand, BERT is powerful in capturing the semantic relationships among hashtags from posts or comments.

3.4.1. Hashtag graph construction

Since hashtags are topic indicators generated by users, Instagram comments that share the same hashtags have underlying topics that overlap strongly [42]. For drug trafficking on social media, comments sharing the same drug-related hashtags can indicate that the same type of drug is being sold. Additionally, the hashtags in the same comment can be synonyms or slang terms for the same type of drug.

To represent the relationship between the hashtags of one comment or different comments, we denote by a hashtag graph G = (V, E, W), where V is the set of nodes (hashtags), E is a set of edges connecting the hashtags, and $w \in W$ indicates the weight of the edge $e \in E$. Drug-related hashtag graphs are constructed from hashtags in drug dealers' comments, which serve as a medium to link hashtags. The vertices (V) on a graph represent the unique hashtag in our dataset. As shown in Fig. 3, we have four comments (i.e., C1, C2, C3, and C4) from four drug dealers and there are these hashtags *#lsdtrip* (*H*1), *#lsdtabs* (*H*2), *#acidtrip* (*H*3), *#acid*. Edges (E) are connections that represent the relationships between nodes. To analyze the relationship between hashtags, we create an edge $e \in E$ if a pair of hashtags are mentioned in the same comment. For example, the hashtag *#lsdtrip* was mentioned in the same comment as *#lsdtabs*, we would assume that there is a relationship between the hashtag *#lsdtrip* and *#lsdtabs*, and then we create an edge between these two vertices. Furthermore, if these two hashtags appear n times in the same comments in our data set, the weight $w \in W$ of the edge $e \in E$ between them will be n. In other words, the more times two hashtags are mentioned together, the greater the weight of the edge that connects them.

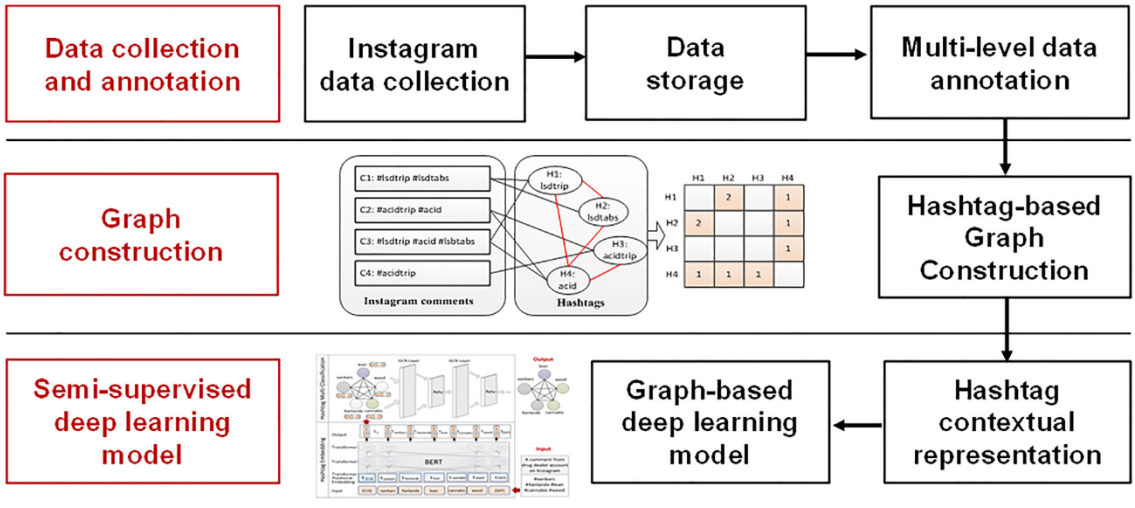

Fig. 1. Overview of our proposed framework for fine-grained classification of drug trafficking with hashtags on Instagram. Our system consists of three stages: 1) data collection and annotation (Section 3.3); 2) Hashtag-based graph construction (Section 3.4.1); and 3) semi-supervised deep learning (Section 3.4.2).

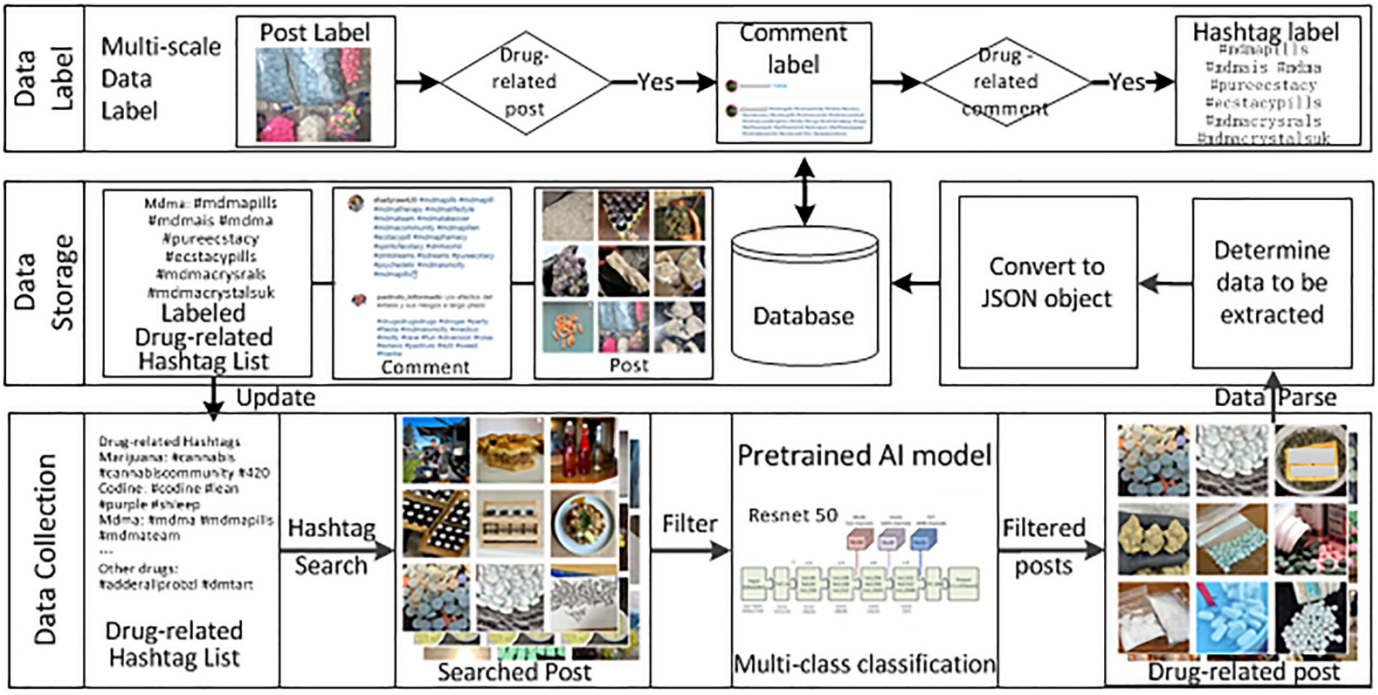

Fig. 2. The architecture of the system for data collection, storage, and calibration.

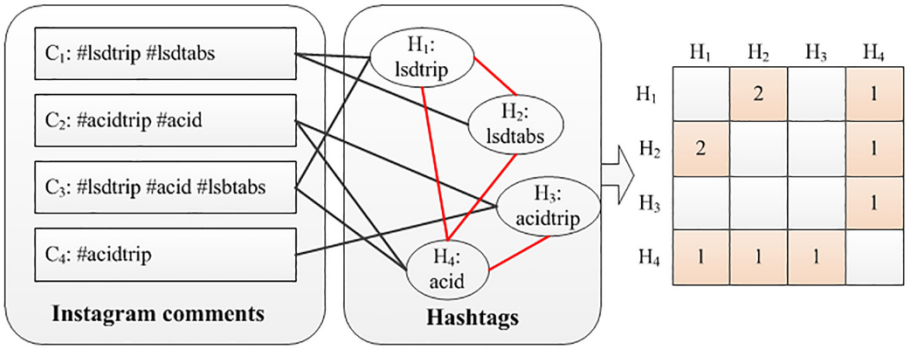

Fig. 3. An illustration of the semantic relationship in Instagram comments. The explicit relationship is a co-occurrence relation between hashtags (marked with red links). The relationship between hashtags can be formulated as a relation graph represented by an adjacency matrix. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4.2. Semi-supervised learning with graph convolutional networks

As shown in Fig. 4, our BERT+GCN model contains two major components: hashtag embedding and hashtag classification. The embedding of hashtags is used to extract text features with contextual relations from captions and comments in posts. Specifically, let $\mathscr{D} = \{p_1, p_2, \ldots, p_T\}$ be a collection of captions or comments from Instagram posts and let $\mathscr{H} = \{h_1, h_2, \ldots, h_N\}$ be the set of associated hashtags. BERT takes the text data $\mathscr{D}$ as input and embeddings of the output for each token in $\mathscr{D}$, including each hashtag in that hashtags $\mathscr{H}$ are associated with $\mathscr{D}$. For a given token (e.g., xanbars, Xanlandx, and lean in Fig. 4), the representation of the input vector is constructed by adding the embedded token, the embedded segmentation, and the embedded position. These token embeddings for hashtags as input were used to extract features using the BERT model. The size of the output features for each hashtag is 1×768. We assign to each corresponding hashtag the output vector of the constructed graph network as the node attribute. As a result, we have the embedding matrix $X = [x_1, x_2, \ldots, x_N] \in R^{N \times C}$ for hashtags $\mathscr{H} = \{h_1, h_2, \ldots, h_N\}$, where $C = 768$ is the number of input features derived from the BERT model.

Since only a small subset of hashtags have been labeled, we apply a semi-supervised learning method based on a graph convolutional network (GCN) [41] to capture the structure relationships between all hashtags so that the label information can be effectively propagated to unlabeled hashtags. Let $A \in R^{N \times N}$ be the adjacency matrix of the hashtag graph $G = (V, E, W)$ constructed in Sec. 3.4.1. The GCN is made up of an input layer, a few hidden propagation layers, and an output layer. Specifically, given the input layer $H^{(0)} = X$ and the adjacency matrix A, the GCN performs layer-wise propagation as follows:

$$H^{(l+1)} = \sigma\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(l)} W^{(l)}\right) \tag{1}$$

where $l = 0, 1, \ldots, L-1$ and $D = \text{diag}(d_1, d_2, \ldots, d_N)$ be a diagonal matrix with $d_i = \sum_{j=1}^{N} A_{ij}$. $W^{(l)}$ is a layer-specific weight matrix to be learned from the data and $\sigma(\cdot)$ denotes an activation function such as ReLU. Note that $H^{(l+1)}$ is the output after the activation operation in the l-th layer. For classification, we perform a softmax operation on the final layer $H^{(L)}$, namely,

$$Z = \text{softmax}\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(L)} W^{(L)}\right). \tag{2}$$

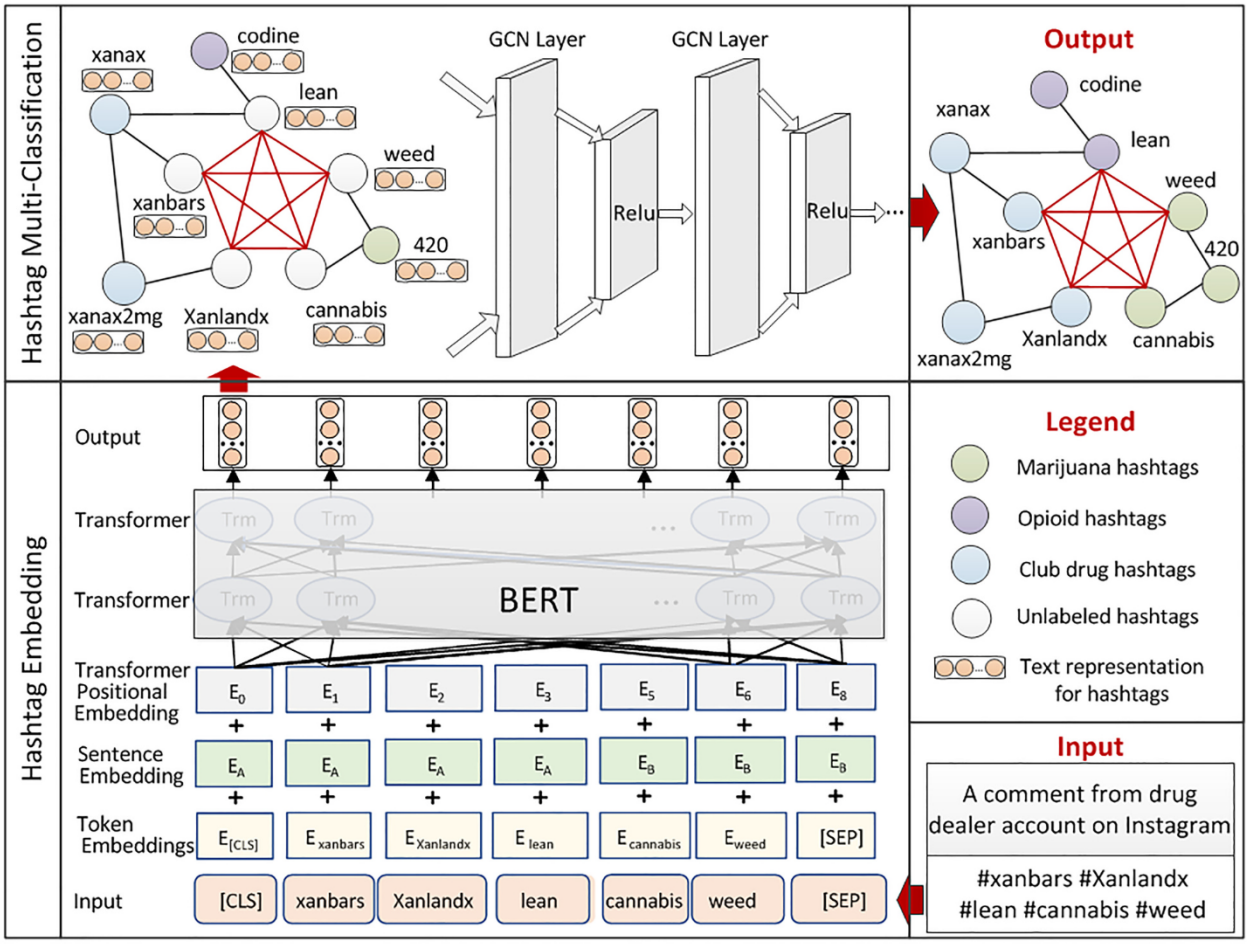

Fig. 4. Overview of the proposed BERT+GCN model. Taking this case as an example, a comment containing several drug-related hashtags as input and the drug category for each hashtag as output. Different colored circles represent hashtags of different drug categories.

The output $Z \in R^{N\times K}$ is the prediction of the label for all hashtags $\mathscr{H} = \{h_1, h_2, \ldots, h_N\}$, and K is the number of classes of drug types being considered. For example, if we set the number of layers as $L = 2$, then the model takes the following form:

$$Z = f(X,A) = softmax\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\,\mathrm{ReLU}\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}XW^{(0)}\right)W^{(1)}\right). \quad (3)$$

In a semi-supervised learning setting where only a portion of the hashtags are labeled, we train the neural network on a supervised target for all nodes with labels to obtain the optimal weights $\{W^{(0)}, W^{(1)}, \ldots, W^{(L)}\}$. Furthermore, for multiclassification tasks, the class imbalance problem is common during training. To address this problem, the focal loss function [43] has been used as the loss function, which can help focus learning on hard misclassification. The loss function on all hashtags labeled is then defined as:

$$L = -\sum_{l\in\mathscr{Y}_l}\sum_{k=1}^{K} y_{lk}\cdot\alpha(1-z_{lk})^{\gamma}log(z_{lk}) \quad (4)$$

where $\mathscr{Y}_l$ is the set of node indices that have labels, $\alpha \in (0,1]$ and $\gamma \geq 0$ are tunable parameters.

4. Experiments

4.1. Experimental setup

Dataset. To verify the feasibility of the proposed method, we collected a total of more than 20,000 Instagram posts over 4 months from April 2020 to August 2020. There were a total of 1022 drug dealer posts, 1956 drug dealer comments (999 of them contain hashtags), and 4240 drug dealer hashtags based on manual annotation. 34.5% of these hashtags have been annotated with four types of drugs, namely, marijuana, opioid drugs, club drugs, and other drugs. The annotation of dataset was carried out by a group of 6 selected auditors. The issue of interrater variability is addressed by some training session: all participants reviewed the tutorial on how to annotate the sample data before working on the real data. The inter-rater / intercoder agreement (kappa) score [44] was 79.59%. The original data set has been pre-processed into a graph dataset, including 7 graphs, 4240 nodes, 96,805 undirected edges, and 45.66% average node degrees. 34.5% of these hashtags have been annotated.

Baselines. To our knowledge, our work is the first study on fine-grained classification of different drug types using hashtags on Instagram. The original dataset has been preprocessed into a graph dataset, 7 graphs, 4240 nodes, 96,805 undirected edges. Based on the graph dataset, we have compared our proposed method with the following text representation methods, including GloVe [45], FastText [46], Komninos [47], Roberta [48] and GPT-2 [49]. Furthermore, three semi-supervised graph-based learning methods have been selected, including graph neural networks (GNNs) [50], GraphSAGE [51], graph attention networks (GATs) [52], to compare the performance of the model with the proposed method.

Training and testing. We used a 10-fold cross-validation to report the results for each model. As only 34.5% of the hashtags (1463 samples) are annotated in our dataset, the labeled data samples are split into 10 folds for a 10-fold cross-validation. All models are trained with the

Adam optimizer algorithm. We set the learning rate at 0.01 and the weight decay at 0.0005. We opt to terminate the training after 500 epochs. All experiments were performed using PyTorch on a workstation with an RTX 2080 GPU. We evaluated the models using accuracy, precision, recall, and the F1 score.

4.2. Results

Table 1 shows the comparison of hashtag classification performances between the proposed model and the baseline approaches. The results show that the proposed approach, namely BERT + GCN, achieves the best performance compared to the other models. The F1 score for the four types of drugs are: marijuana 91.03%, opioid 66.04%, club drugs 77.78%, and other drugs 64.71%. We observe that both hashtag embedding methods (GloVe, Komnimos, FastText, Roberta, GPT2, and semi-supervised learning methods based on BERT) and graph neural networks (GNN, GraphSAGE, GAT, and GCN) would impact classification performances. In particular, BERT is shown to be the best embedding method. The possible reason is that BERT can understand the full context of the word by using the transformer [53] architecture that processes any given word about all other words in a sentence, rather than processing them one at a time (e.g., GloVe, Komninos, and FastText). For comments on ads selling illegal drugs, the name, slang, color, and shape of a certain drug can be mentioned multiple times with different hashtags. BERT can help distinguish between different types of drug-related hashtags by representing comments.

Fig. 5a shows the t-SNE plot the feature map in the last layer of the GCN, and Fig. 5b shows the confusion matrix of the proposed model performance evaluation. We observe that it is relatively easy to distinguish between marijuana and club drugs, but more difficult to distinguish between opioids and other drugs. One possible explanation is that opioid drugs contain many types of drugs (e.g., codeine, fentanyl, hydrocodone, methadone, and morphine), as do other types of drugs (e.g., Xanax, cocaine, and Adderall).

4.3. Impacts of label annotation rate

For semi-supervised learning, the label annotation rate has an important impact on its model performance. To evaluate the impacts of label annotation rate classification performances, we compare the model performance based on all our labels (nearly 35%) with different label rates at 5%, 15% and 25%. As shown in Table 2, we can see that the higher the rate of labeling of the sample, the better the performance of the model. For example, we see more than 20% improvement in the F1 score when we increase the rate of labeling from 5% to 20%.

4.4. Case studies with classified hashtags

The classified hashtags can be used for a wide range of decision support systems to monitor drug trafficking activities on social media platforms. We present two case studies with the classified hashtags.

Table 1
Performance comparison (averaged over all drug types) between the proposed method and the baseline approaches.

Methods	Precision	Recall	Accuracy	F1 score
GloVe + GCN	48.60%	47.11%	49.62%	43.75%
Komnimos + GCN	52.88%	49.63%	52.29%	46.30%
FastText + GCN	59.22%	55.31%	57.25%	53.88%
Roberta + GCN	64.14%	64.08%	65.27%	64.07%
GPT2 + GCN	70.20%	70.54%	71.37%	70.32%
BERT + GNN	58.39%	58.35%	60.31%	57.82%
BERT + GraphSAGE	69.56%	69.42%	70.23%	69.46%
BERT + GAT	**75.56%**	74.61%	75.57%	74.55%
BERT + GCN	74.78%	**75.07%**	**75.95%**	**74.89%**

4.4.1. Case study 1: evaluation of platform self-regulatory on drug trafficking

As an important tool to increase audience engagement, hashtags have been widely used for illegal drug trafficking on social media. Therefore, limiting the visibility of drug-related hashtags could effectively control the spread of illegal drug trafficking. However, currently there is a lack of monitoring mechanisms to determine whether social media platforms are committed to controlling illicit drug trafficking. This case study demonstrates our classified drug-related hashtags can be used to evaluate the self-regulation of drug trafficking on social media platforms. To this end, for each drug-related hashtag in our dataset we collected the most recent 100 posts that contains the hashtag. Each post then is classified whether as a drug trafficking-related post using a high accurate (94.95% accuracy) drug post classification model [18]. The *positive rate*—proportion of drug-trafficking-related posts for each drug-related hashtag—is calculated based on the drug trafficking predictions. The average positive rate of drug-related hashtags is used as a quantitative index to evaluate self-regulation of this type of drugs on social media. The higher the positive rate, the looser self-regulation on social media over this type of drug-related hashtags. Some examples of positive rate for the predicted hashtags are shown in Table 3. Fig. 6 shows the positive rate for each type of drug-related hashtags. We can see that the platform has the loosest self-regulation over marijuana-related hashtags (positive rate more than 40%). Self-regulation of club drugs (including LSD, MDMA, etc.) is looser compared to the other two types of drugs (opioids and other drugs). This observation may be related to the legalization process of different drugs. The commercial sale of recreational marijuana is legalized nationwide in Canada, Thailand, and Uruguay. Therefore, the legalization of marijuana may prompt platforms to loosen the self-regulation of related drug trafficking.

4.4.2. Case study 2: geographic analysis of drug-related hashtags

Beyond drug-related hashtags, there are numerous hashtags about locations (e.g., country, state and city), which might be related to drug shipment in drug trade. These location-related hashtags allow us to explore regional differences in the trading of different drugs. We obtain the predicted drug type for each location-related hashtag based on the proposed model. The drug type of each location is calculated based on the maximum value of the weighted sum of hashtags and their frequency of occurrence. We then map the top drug types in each state to analyze regional differences in drug trafficking in the United States. Fig. 7a shows the geo-mapping of the top drug-related hashtags in each state in the United States. We found that the the *other drugs* type (e.g., xanax, cocaine, and adderall) is the major drug trafficking in the United States (16 states). Club drugs are the second most popular drug (13 states). Opioid drugs are the leading drug trafficking in three states: West Virginia, Kansas, and Oklahoma. The main drug on Instagram sold in Washington is marijuana.

We further compare the identified top drugs in each state using our classification results with drug-related dataset collected from Google Trends. When ground-truth data is unavailable, Google Trends data can be used as a surrogate [54]; in particular, large amount of searches for drug-related keywords tend to reflect local drug demand [55]. Therefore, we assume that there is a certain spatial correlation between the top-drug map derived from our classification results and that derived from Google Trends, which is shown in Fig. 7b. We use Pearson correlation analysis to quantitatively analyze the correlation between the top drugs identified by our methods (Fig. 7a) and those derived from Google Trends (Fig. 7b). The Pearson correlation coefficient is $r = 0.3901$, which indicates there is a moderate correlation between our results and the results of the Google Trends data. Furthermore, to verify the robustness of the proposed method, the evaluation of the influence of different models in this case study is shown in Table 4. The results of the geographical analysis based on the proposed method have the highest positive correlation with the Google Trends data. Combining the experimental results in Table 1 shows that the better the model

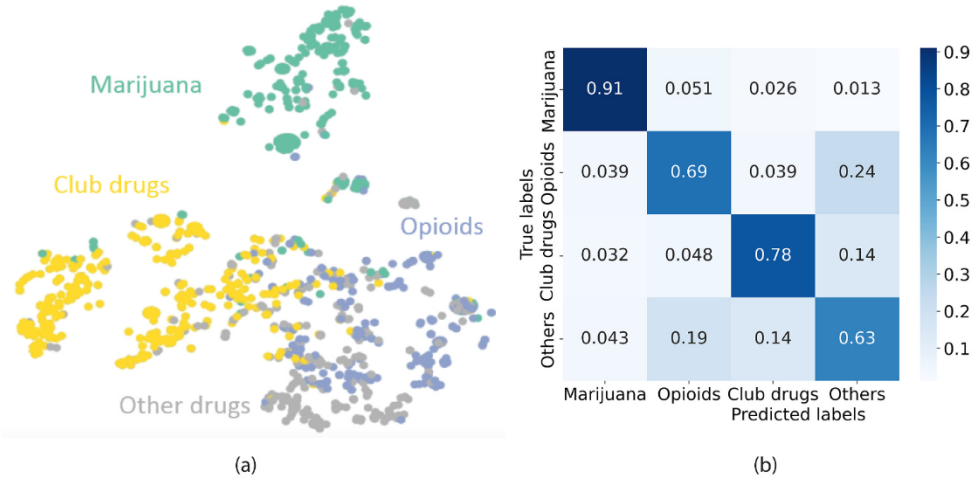

Fig. 5. (a) tSNE visualizations of the feature map in the last layer of the GCN. Different classes are marked with different colors. (b) Visualizations of the confusion matrix for the performance evaluation of the proposed model.

Table 2
Performance comparison of the proposed method with different label rates.

Label rates	Precision	Recall	Accuracy	F1 score
5%	36.74%	47.74%	44.66%	39.10%
15%	64.72%	64.05%	64.89%	63.45%
25%	67.82%	67.73%	68.70%	67.35%
35%	**74.78%**	**75.07%**	**75.95%**	**74.89%**

Table 3
Examples of positive rate for the predicted hashtags.

Category	Hashtag	Positive Rate
Marijuana	#420united	33%
	#420miami	100%
	#weedreup	47%
Opioids	#californialean	6%
	#oxytocin	17%
	#painkiller11	4%
Club drugs	#molly	56%
	#trippymushroom	16%
	#miamilsd	48%
Other drugs	#handlebars	5%
	#coketurkey	21%
	#whitecocaine	10%

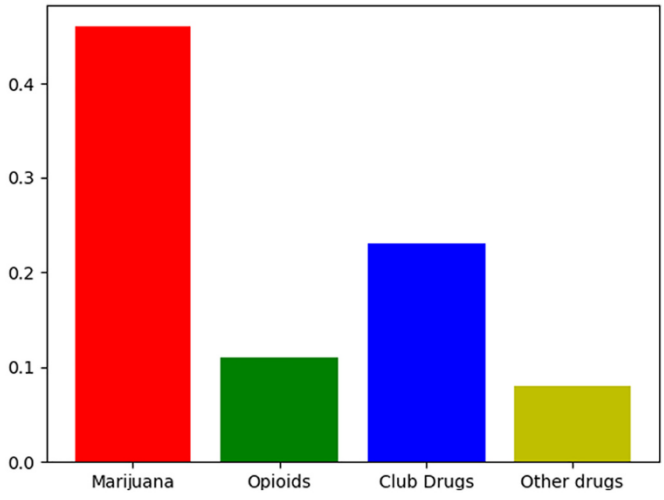

Fig. 6. Average drug trafficking post *positive rate* for each drug-related hashtag.

performance, the higher the correlation between geographic analysis results and Google Trends data.

5. Conclusion and discussions

In this paper, we provided a systematic study on the fine-grained classification of drug-related hashtags associated with different drug types from illegal drug sellers on Instagram. We formulated hashtag classification as a semi-supervised leaning problem, and we proposed a method that combines BERT and GCN to infer the labels for the unlabeled hashtags. We have collected and constructed a hashtag-based dataset from Instagram to support fine-grained drug classification. Based on the constructed dataset, we have reported preliminary experimental results to demonstrate the effectiveness of the proposed approach. Based on the classification of hashtags, we also provided two case studies indicating there are regional differences in the demand for different types of drugs, as well as in self-regulation on social media platform.

First, unlike previous studies that analyzed Instagram posts on illicit drug sales [6,8,11,18], our main finding indicates that many drug dealers on Instagram posted ads for illicit drug sales with drug name hashtags, which can help improve viral attention on user posts and increase public health and safety risks. In other words, reasonable and effective hashtag control is the most effective means of controlling drug trafficking activities for public health and safety risk. The proposed method can find numerous drug-related hashtags that describe the color, form, dosage, how-to-use, and names in different languages (e.g., Spanish) of drugs to avoid platform detection.

Furthermore, we found that differences in the legalization of the four proposed drug types will lead to different levels of activity in drug trafficking based on the fine-grained hashtag classification. For example, marijuana use is legal in 18 states in the United States, and the proportion of drug trafficking in its related hashtag posts is much higher than other illegal drugs. The method we propose can provide an important means for policymakers to monitor whether social media platforms have taken effective measures to control the risk of drug trafficking to public health and safety. Additionally, the fine-grained drug-related hashtag classification scheme designed can provide personalized management based on the evolving drug legalization process.

Finally, we found that these four types of drugs have varying market demands in different regions based on the analysis of the location-based hashtag of drug dealers. The proposed method can provide a valuable

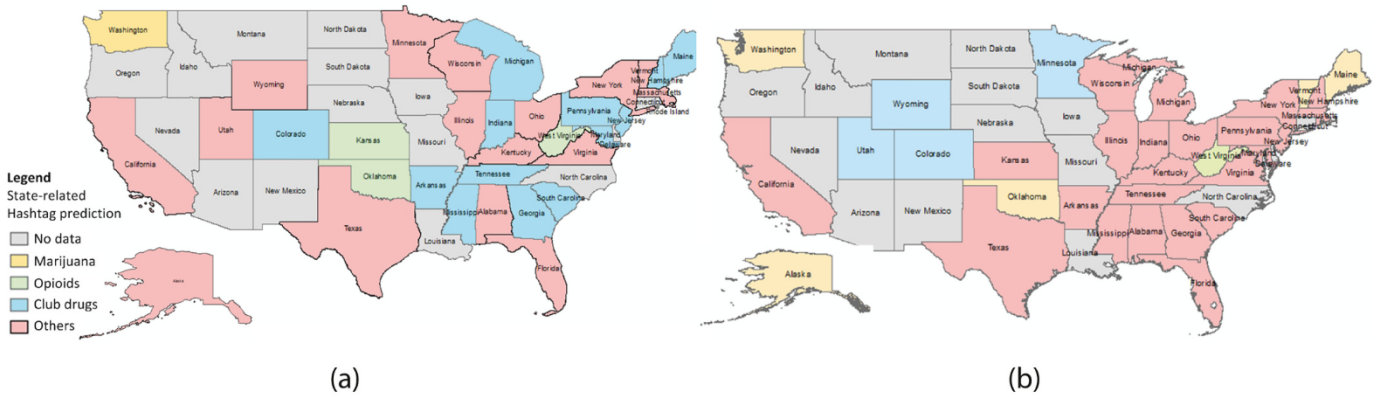

Fig. 7. Geo-mapping of the top drug-related hashtags with out predicted results (a), and with Google Trends data (b).

Table 4
Evaluation of the influence of different model components.

Method	Pearson's r
GloVe + GCN	0.2344
BERT + GNN	0.2764
BERT + GCN (ours)	0.3901

new tool for further exploration of the hashtag for location of shipping in drug trafficking advertisements to understand the variations in the illicit drug market between different regions. We also note that there are some differences in self-regulation of different types of drugs on social media platforms. These findings could provide useful sales information for drug sales risk management and control on social media platforms in the future.

5.1. Public health implications

As social networks have become important platforms for the marketing and sale of illicit drugs, effective detection of illicit drug trafficking online has become critical in combating online trade in illicit drugs. As such, our AI-supported methodology can be used to better monitor drug trafficking activities on social media platforms and to identify illegal online sellers for law enforcement agencies. For example, Twitter prohibits the promotion of drugs and drug paraphernalia; other platforms such as Instagram and Snapchat are also taking steps to curb the exploitation of social media tools by illegal drug dealers. Hashtag-based analysis can effectively facilitate the monitoring and filtering of drug-related content to comply with FDA advertising regulations. Our methodology can also help law enforcement more effectively detect, disrupt, and dismantle illicit drug dealers on social media. Since the unregulated online sale of controlled substances is illegal, our AI-enabled monitoring system can directly report suspicious online sales activities to the DEA and FDA. This system can potentially be connected to the existing prescription drug monitoring program (PDMP) to persistently track prescription drug misuse and abuse.

5.2. Limitations and future work

Our study has several limitations. First, this study was limited to a short period of data collection with 1022 drug dealer posts and 4240 unique hashtags, which is just the tip of the iceberg in many drug advertisements for sale posted on Instagram. Therefore, the results of the study may not be generalizable and necessarily representative of the Instagram drug sales community. In the future, we will work to collect larger data sets to discover more generalized characteristics of drug trafficking on social media. Furthermore, the proposed method can only achieve a fine-grained classification of hashtags, and we did not engage in management plans for these drug-related hashtags. The question of how to manage these types of drugs in a differentiated way is the key issue that needs to be studied in the future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is partially supported by the NSF under grants IIS-2209814, IIS-2203262, IIS-2214376, IIS-2217239, OAC-2218762, CNS-2203261, CNS-2122631, CMMI-2146076, and the NIJ 2018-75-CX-0032. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any funding agencies.

References

[1] A.S. Bennett, T. Townsend, L. Elliott, The covid-19 pandemic and the health of people who use illicit opioids in New York city, the first 12 months, Int. J. Drug Policy 101 (2022), 103554.
[2] K.A. Mack, C.M. Jones, M.F. Ballesteros, Illicit drug use, illicit drug use disorders, and drug overdose deaths in metropolitan and nonmetropolitan areas—United States, Am. J. Transplant. 17 (12) (2017) 3241–3252.
[3] T.K. Mackey, J. Kalyanam, T. Katsuki, G. Lanckriet, Twitter-based detection of illegal online sale of prescription opioid, Am. J. Public Health 107 (12) (2017) 1910–1915.
[4] T. Mackey, J. Kalyanam, J. Klugman, E. Kuzmenko, R. Gupta, 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, J. Med. Internet Res. 20 (4) (2018), e10029.
[5] J. Demant, S.A. Bakken, A. Oksanen, H. Gunnlaugsson, Drug dealing on facebook, snapchat and instagram: a qualitative analysis of novel drug markets in the nordic countries, Drug Alcohol. Rev. 38 (4) (2019) 377–385.
[6] J. Li, Q. Xu, N. Shah, T.K. Mackey, et al., A machine learning approach for the detection and characterization of illicit drug dealers on instagram: model evaluation study, J. Med. Internet Res. 21 (6) (2019), e13803.
[7] J. Liu, A. Bharadwaj, Drug abuse and the internet: evidence from craigslist, Manag. Sci. 66 (5) (2020) 2040–2049.
[8] C. Hu, M. Yin, B. Liu, X. Li, Y. Ye, Detection of Illicit Drug Trafficking Events on Instagram: A Deep Multimodal Multilabel Learning Approach, CIKM, 2021, pp. 3838–3846.
[9] N. Dasgupta, C. Freifeld, J.S. Brownstein, C.M. Menone, H.L. Surratt, L. Poppish, J. L. Green, E.J. Lavonas, R.C. Dart, et al., Crowdsourcing black market prices for prescription opioids, J. Med. Internet Res. 15 (8) (2013), e2810.
[10] J. Pergolizzi Jr., J. LeQuang, R. Taylor Jr., R. Raffa, N. R. Group, The "darknet": the new street for street drugs, J. Clin. Pharm. Ther. 42 (6) (2017) 790–792.

[11] X. Yang, J. Luo, Tracking illicit drug dealing and abuse on instagram using multimodal analysis, ACM Trans. Intellig. Syst. Technol. 8 (4) (2017) 1–15.
[12] P. Zhang, X. Li, X. Hu, J. Yang, L. Zhang, L. Wang, Y. Choi, J. Gao, Vinvl: revisiting visual representations in vision-language models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 5579–5588.
[13] X. Li, X. Yin, C. Li, P. Zhang, X. Hu, L. Zhang, L. Wang, H. Hu, L. Dong, F. Wei, et al., Oscar: Object-semantics aligned pre-training for vision-language tasks, in: European Conference on Computer Vision, Springer, 2020, pp. 121–137.
[14] S. Jia, X. Li, C. Hu, G. Guo, Z. Xu, 3d face anti-spoofing with factorized bilinear coding, IEEE Trans. Circ. Syst. Video Technol. 31 (10) (2020) 4031–4045.
[15] J. Raiyn, et al., A survey of cyber attack detection strategies, Int. J. Secur. Appl. 8 (1) (2014) 247–256.
[16] S. Jia, C. Hu, G. Guo, Z. Xu, A database for face presentation attack using wax figure faces, in: International Conference on Image Analysis and Processing, Springer, 2019, pp. 39–47.
[17] S. Jia, C. Hu, X. Li, Z. Xu, Face spoofing detection under super-realistic 3d wax face attacks, Pattern Recogn. Lett. 145 (2021) 103–109.
[18] C. Hu, M. Yin, B. Liu, X. Li, Y. Ye, Identifying illicit drug dealers on instagram with large-scale multimodal data fusion, ACM Trans. Intellig. Syst. Technol. 12 (5) (2021) 1–23.
[19] L. Pinho-Costa, K. Yakubu, K. Hoedebecke, L. Laranjo, C.P. Reichel, M.D.C. Colon-Gonzalez, A.L. Neves, H. Errami, Healthcare hashtag index development: identifying global impact in social media, J. Biomed. Inform. 63 (2016) 390–399.
[20] M. Hajizadeh, Legalizing and regulating marijuana in Canada: review of potential economic, social, and health impacts, Int. J. Health Policy Manag. 5 (8) (2016) 453.
[21] A. Heber, The networks of drug offenders, Trends Organ. Crime 12 (1) (2009) 1–20.
[22] F.J. Desroches, The Crime that Pays: Drug Trafficking and Organized Crime in Canada, Canadian Scholars' Press, 2005.
[23] M. Natarajan, Understanding the structure of a drug trafficking organization: a conversational analysis, Crime Prevention Studies 11 (2000) 273–298.
[24] M. Natarajan, Understanding the structure of a large heroin distribution network: a quantitative analysis of qualitative data, J. Quant. Criminol. 22 (2) (2006) 171–192.
[25] M. Natarajan, M. Belanger, Varieties of drug trafficking organizations: a typology of cases prosecuted in New York city, J. Drug Issues 28 (4) (1998) 1005–1025.
[26] C. Buntain, J. Golbeck, This is your twitter on drugs: any questions?, in: Proceedings of the 24th International Conference on World Wide Web, 2015, pp. 777–782.
[27] B. Tofighi, Y. Aphinyanaphongs, C. Marini, S. Ghassemlou, P. Nayebvali, I. Metzger, A. Raghunath, S. Thomas, Detecting illicit opioid content on twitter, Drug Alcohol. Rev. 39 (3) (2020) 205–208.
[28] R.B. Correia, L. Li, L.M. Rocha, Monitoring potential drug interactions and reactions via network analysis of instagram user timelines, in: Biocomputing 2016: Proceedings of the Pacific Symposium, World Scientific, 2016, pp. 492–503.
[29] Y. Zhou, N. Sani, C.-K. Lee, J. Luo, Understanding illicit drug use behaviors by mining social media, arXiv preprint arXiv:1604.07096, 2016.
[30] J. Kalyanam, T.K. Mackey, A review of digital surveillance methods and approaches to combat prescription drug abuse, Curr. Addict. Reports 4 (4) (2017) 397–409.
[31] A. Sarker, G. Gonzalez-Hernandez, Y. Ruan, J. Perrone, Machine learning and natural language processing for geolocation-centric monitoring and characterization of opioid-related social media chatter, JAMA Netw. Open 2 (11) (2019) e1914672.
[32] S. Hassanpour, N. Tomita, T. DeLise, B. Crosier, L.A. Marsch, Identifying substance use risk based on deep neural networks and instagram social media data, Neuropsychopharmacology 44 (3) (2019) 487–494.
[33] G. Barbier, H. Liu, Data mining in social media, in: Social Network Data Analytics, Springer, 2011, pp. 327–352.
[34] Z. Ye, N.H. Hashim, F. Baghirov, J. Murphy, Gender differences in instagram hashtag use, J. Hosp. Mark. Manag. 27 (4) (2018) 386–404.
[35] R.G. Dorfman, E.E. Vaca, E. Mahmood, N.A. Fine, C.F. Schierle, Plastic surgery-related hashtag utilization on instagram: implications for education and marketing, Aesthet. Surg. J. 38 (3) (2018) 332–338.
[36] J.-P. Allem, P. Escobedo, K.-H. Chu, T.B. Cruz, J.B. Unger, et al., Images of little cigars and cigarillos on instagram identified by the hashtag# swisher: thematic analysis, J. Med. Internet Res. 19 (7) (2017), e7634.
[37] C.L. Mattson, L.J. Tanz, K. Quinn, M. Kariisa, P. Patel, N.L. Davis, Trends and geographic patterns in drug and synthetic opioid overdose deaths—United States, 2013–2019, Morb. Mortal. Wkly Rep. 70 (6) (2021) 202.
[38] P.M. Gahlinger, Club drugs: Mdma, gamma-hydroxybutyrate (ghb), rohypnol, and ketamine, Am. Fam. Physician 69 (11) (2004) 2619–2626.
[39] D.E. Ramo, C. Grov, K. Delucchi, B.C. Kelly, J.T. Parsons, Typology of club drug use among young adults recruited using time–space sampling, Drug Alcohol Depend. 107 (2–3) (2010) 119–127.
[40] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding, NAACL-HLT(1), 2019.
[41] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, Int. Conf. Learn. Represent. n/a (n/a) (2017).
[42] Y. Wang, J. Liu, J. Qu, Y. Huang, J. Chen, X. Feng, Hashtag Graph Based Topic Model for Tweet Mining, ICDM, 2014, pp. 1025–1030.
[43] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[44] M.L. McHugh, Interrater reliability: the kappa statistic, Biochem. Med. 22 (3) (2012) 276–282.
[45] J. Pennington, R. Socher, C.D. Manning, Glove: Global Vectors for Word Representation, EMNLP, 2014, pp. 1532–1543.
[46] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, T. Mikolov, Fasttext. Zip: compressing text classification models, ICLR (2017).
[47] A. Komninos, S. Manandhar, Dependency based embeddings for sentence classification tasks, in: Proceedings of NAACL-HLT, 2016, pp. 1490–1500.
[48] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: a robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692, 2019.
[49] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI Blog 1 (8) (2019) 9.
[50] J. You, Z. Ying, J. Leskovec, Design space for graph neural networks, Adv. Neural Inf. Proces. Syst. 33 (2020) 17009–17021.
[51] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, Adv. Neural Inf. Proces. Syst. 30 (2017).
[52] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, arXiv preprint arXiv:1710.10903, 2017.
[53] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Proces. Syst. 30 (2017).
[54] H. Choi, H. Varian, Predicting the present with google trends, Econ. Rec. 88 (2012) 2–9.
[55] Y. Carrière-Swallow, F. Labbé, Nowcasting with google trends in an emerging market, J. Forecast. 32 (4) (2013) 289–298.

Chuanbo Hu received the Ph.D. degree in geographic information science from Wuhan University, Wuhan, China, in 2017. He was a Post-Doctoral Research Fellow with The Chinese University of Hong Kong, Hong Kong, from 2017 to 2018. He is currently a Post-Doctoral Scholar with West Virginia University, Morgantown, WV, USA. His research interests include geospatial artificial intelligence (GeoAI), health geography, and data mining.

Bin Liu received the PhD degree from Rutgers University. He is a research scientist at the IBM Thomas J. Watson Research Center. He is interested in data mining and machine learning, and their intersections with healthcare, business analytics, recommender systems, and privacy/security. He has published in premier journals such as the IEEE Transactions on Knowledge and Data Engineering, the ACM Transactions on Knowledge Discovery from Data, the ACM Transactions on Privacy and Security, the ACM Transactions on Intelligent Systems and Technology; and top conferences such as KDD, AAAI, WSDM, and USENIX Security. He currently serves on the editorial board of the Journal of Business Analytics, and has served as a reviewer for many journals, including IEEE TKDE. He has served regularly on program committees of conferences, including KDD, AAAI, CIKM, SDM, and RecSys. He is a member of the ACM and IEEE.

Yanfang Ye is currently the Collegiate Associate Professor of Computer Science and Engineering at the University of Notre Dame, Notre Dame, IN, USA. Her research interests include cybersecurity, data mining, machine learning, and health intelligence. Her proposed techniques by advancing AI and data-driven innovations for malware detection have been incorporated into popular commercial cybersecurity products that protect millions of users worldwide. She has expanded her research on health intelligence with focus on combating opioid crisis and infectious disease outbreaks. She was the recipient of the ACM CIKM 2021 Best Paper Award (Full Paper Track), the CSE Innovation Award in 2020-2021 and CSE Research Award in 2019–2020 at Case Western Reserve University, the NSF Career Award in 2019, the MetroLab Innovation of the Month in May 2020, the IJCAI 2019 Early Career Spotlight, the AICS 2019 Challenge Problem Winner, the SIGKDD 2017 Best Paper Award and Best Student Paper Award (ADS Track), the IEEE EISIC 2017 Best Paper Award, and the New Researcher of the Year Award in 2016–2017 at West Virginia University.

Xin Li received the B.S. degree (Hons.) in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 1996, and the Ph. D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2000. He was a Member of the Technical Staff with Sharp Laboratories of America, Camas, WA, USA, from August 2000 to December 2002. Since January 2003, he has been a Faculty Member of the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA. His research interests include image/video coding and processing. Dr. Li received various best paper awards at image processing and data mining conferences. He was elevated to an Fellow of IEEE in 2017.