

Identifying Illicit Drug Dealers on Instagram with Large-scale Multimodal Data Fusion

CHUANBO HU, MINGLEI YIN*, BIN LIU*, and XIN LI*, West Virginia University
YANFANG YE, Case Western Reserve University

Illicit drug trafficking via social media sites such as Instagram has become a severe problem, thus drawing a great deal of attention from law enforcement and public health agencies. How to identify illicit drug dealers from social media data has remained a technical challenge due to the following reasons. On the one hand, the available data are limited because of privacy concerns with crawling social media sites; on the other hand, the diversity of drug dealing patterns makes it difficult to reliably distinguish drug dealers from common drug users. Unlike existing methods that focus on posting-based detection, we propose to tackle the problem of *illicit drug dealer identification* by constructing a large-scale multimodal dataset named Identifying Drug Dealers on Instagram (IDDIG). Totally nearly 4,000 user accounts, of which over 1,400 are drug dealers, have been collected from Instagram with multiple data sources including post comments, post images, homepage bio, and homepage images. We then design a quadruple-based multimodal fusion method to combine the multiple data sources associated with each user account for drug dealer identification. Experimental results on the constructed IDDIG dataset demonstrate the effectiveness of the proposed method in identifying drug dealers (almost 95% accuracy). Moreover, we have developed a hashtag-based community detection technique for discovering evolving patterns, especially those related to geography and drug types.

CCS Concepts: • **Information systems** → **Data mining**; • **Applied computing** → **Law, social and behavioral sciences**.

Additional Key Words and Phrases: drug trafficking, drug dealer, Instagram, multimodal data fusion

ACM Reference Format:

Chuanbo Hu, Minglei Yin, Bin Liu, Xin Li, and Yanfang Ye. 2018. Identifying Illicit Drug Dealers on Instagram with Large-scale Multimodal Data Fusion. *J. ACM* 37, 4, Article 111 (August 2018), 23 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

Illegal drug trade (a.k.a. drug trafficking) is a global black market dedicated to the distribution and trade of drugs prohibited by law. Due to the co-evolution of cyberspace and human society, online illicit drug trade has become a major problem, attracting increasingly more attention from both law enforcement and public health agencies. According to a RAND report in 2010 [40], the total size of the illicit drug market in the US was estimated to be in the order of 100 billion dollars per year. Recent studies [44, 76] have shown that popular social media platforms such as Instagram, Twitter, and Facebook have become a convenient direct-to-consumer marketing tool for illegal dealers. Among popular social media platforms, Instagram is a particularly effective tool for advertising

Authors' addresses: Chuanbo Hu, cbhu@whu.edu.cn; Minglei Yin, my0033@mix.wvu.edu; Bin Liu, bin.liu1@mail.wvu.edu; Xin Li, xin.li@mail.wvu.edu, West Virginia University, P.O. Box 6109, Morgantown, WV, 26506-6109; Yanfang Ye, Case Western Reserve University, Dept. of CDS, Cleveland, Ohio, 78229, yanfang.ye@case.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Association for Computing Machinery.

0004-5411/2018/8-ART111 \$15.00

<https://doi.org/10.1145/1122445.1122456>



Fig. 1. A Drug-related Post with Multiple Patterns of Illicit Drug Dealing. Drug dealers can engage in illicit drug trade by direct posting (pattern 1) or commenting (pattern 2).

illicit drugs due to their photo-sharing features. Since the majority of Instagram users are young people, including teenagers, it is important to tackle the problem of illicit drug trade on Instagram.

In this paper, focusing on Instagram, we study the problem of *illicit drug dealer identification*, which is a critical step to combat illicit drug trafficking on social media platforms. There are several technical challenges facing the detection of illicit drug dealers from Instagram data: (1) Ambiguity. It is often difficult to distinguish an illicit drug dealer from a drug abuser account. Even for domain experts, different people could reach different conclusions, which make the results inconsistent [76]. Illicit drug trafficking methods have become sophisticated in social media platforms such as Instagram. As shown in Fig. 1, drug dealers can engage in illicit drug trade by either posting or commenting. Furthermore, how to distinguish illicit drug dealers from legal regular consumers or drug abuser accounts has remained one of the open problems [76]. (2) Heterogeneity. The data sources related to illicit drug trade involve both images and text; moreover, the modes of communication adopted by illicit drug dealers are diverse - posting vs. commenting, hashtag vs. homepage, and diversity of drug-related images. How to systematically pull together these heterogeneous information has remained a long-standing open problem [73]. (3) Scalability. Even though the population of drug dealer accounts is relatively small, the total size of social media data is enormous. How to efficiently mine large-scale and temporally varying social media data calls for innovative technical solutions at the system level [77].

The motivation behind this work is largely two-fold. First, recognizing the importance of data and the rapid growth of Instagram community [21], we argue that it is necessary to construct an update-to-date and large-scale database to support the research related to illicit drug trade. When compared against the pioneering work in this field [76], we believe Instagram-based detection of illicit drug dealers can benefit from a more ambitious effort of data collection. Moreover, we advocate the public release of such a database to serve as a benchmark for supporting the emerging field of research on illicit supply networks [32]. Second, rapid advances in machine learning, especially deep learning [20] and multimodal data fusion [42] provide a rich weaponry of advanced computational tools for automatically extracting knowledge from training data without feature engineering. How

to leverage these latest advances in drug dealer identification calls for an integrated approach. Meanwhile, the issue of dealing with missing information in multimodal data (e.g., drug dealers often intentionally hide some information for evading law enforcement) has remained an open problem [42].

In this work, we propose to tackle the problem of drug dealer identification by constructing a large-scale dataset called Identifying Drug Dealers on Instagram (IDDIG), including over 2,000 posts, nearly 4,000 user homepages as well as multimodal data sources (text and images, posts and homepage). In particular, the importance of biography information at the homepage on social media data mining has remained an underexplored topic [30, 35] and not been studied in previous works on drug-dealer detection [44, 76]. To construct such a large-scale dataset, we have designed an automatic data crawling system for Instagram that jointly uses hashtag and image information to guide the data collection. A user-friendly data annotation platform is also developed to support manual labeling of multimodal data, which is inevitably a tedious process. With labeled data (ground-truth for machine learning), we have built a brand new drug-dealer identification system by leveraging the latest advances in deep learning including Bidirectional Encoder Representations from Transformers (BERT) [10] based text classification, ResNet-based [28] image classification, as well as feature-level multimodal data fusion [16]. The key contributions of this paper are summarized as follows.

- The construction of a large-scale dataset called IDDIG for drug-dealer identification. We have designed an automatic hashtag-based data crawling system and a user-friendly data annotation system to support large-scale data collection. For the first time, we have collected over 1,400 drug dealers' accounts (positive examples) and over 2,000 non-drug-dealer accounts (negative examples). The newly constructed IDDIG dataset will be made publicly available to support the research related to illicit drug trade.
- The development of a fully automatic and highly accurate system for drug dealer identification. Our new system is built upon a novel quadruple-based data representation (image vs. text, post vs. homepage) and several latest advances in deep learning including natural language processing (e.g., BERT model [10]), deep image classification (e.g., ResNet model [28]) and feature-level multimodal data fusion (e.g., [26]). For the first time, we have addressed the issue of missing modalities in multimodal data fusion.
- The application of joint post and homepage identification to real-world Instagram data for identifying illicit drug dealers and drug rings (i.e., community detection). We have achieved an overall accuracy of almost 95% on our test dataset with a balanced training dataset, and demonstrated a graceful performance degradation as the training dataset gradually becomes unbalanced. The developed system has also demonstrated its potential for detecting drug-trafficking communities online, which could facilitate the disruption of illicit drug trade by law enforcement.

The rest of this paper is organized as follows. In Section 2, we briefly review related works on social media data mining and user profiling to contextualize the proposed approach. In Section 3, we describe the construction of our new large-scale dataset in detail, including data crawling and data annotation subsystems. In Section 4, we present a new drug-dealer identification system by jointly exploiting the post and homepage information. In Section 5, we address the issue of community detection based on hashtags and graph clustering. We report our experimental results in Sec. 6 and make several concluding remarks in Sec. 7.

2 RELATED WORK

2.1 Social Media Data Mining

In the past decades, social media has greatly facilitated the generation and sharing of information via virtual communities and networks. Data associated with popular social media platforms such as Facebook and Twitter have grown at exponential rates. Accordingly, social media data mining [3, 69–71, 77] has rapidly evolved into an emerging field with a wide range of applications from online surveys [57] and public health monitoring [50, 74] to online marketing [56, 65] and recommendation systems [23, 48]. Various new mining techniques such as user profiling and community detection have been developed for social media data specifically.

In particular, our work is related to user profiling in social media that classifies users into different labels from their social media data such as posted texts and images [34, 45]. Such a collection of online identities has found successful applications in personalized recommendation [33] and marketing analysis [34]. Most recently, deep learning-based multimodal data fusion has been studied for user profiling in [13]. Conceptually analogous to the grouping of people in the physical world, online identities tend to form various virtual communities. Community detection and analysis [3, 77] have been widely studied in the literature of social media data mining. Both member-based and group-based community detection algorithms have been developed; both the evolution and evaluation of virtual communities have been investigated.

2.2 Drug Abuse and Dealing Analysis

As far as we know, there has been limited work on tracking drug abuse and illicit drug trade from online data. Among these existing works, [5] analyzed the time and location patterns of drug use by mining Twitter data; network information of Instagram user timelines was used in [8] to monitor suspicious drug interaction activities; [85] and [76] analyzed Instagram data for tracking and identifying drug dealer accounts. More recently, machine learning and natural language processing techniques have been applied to combat prescription drug abuse [39, 63, 64] and detect illicit drug dealers [44, 81]. Different from previous work that only used one modality data (i.e., text data), our work is built upon multimodal data fusion for illicit drug dealer identification. There are some works that combines both image and text data for identifying substance use risk [27]. Our work is different in twofold: first, our research goal is illicit drug dealer identification, which is more challenging due to the factor that illicit drug trafficking methods have become sophisticated in social media platforms. Second, our multimodal data fusion is built upon a novel quadruple-based data representation (image vs. text, post vs. homepage).

2.3 Illicit Drug Use Detection Systems

In our own previous work, we have developed an automatic opioid user detection system called *AutoDOA* using Twitter data in [12]. In *AutoDOA*, to model the users and posted tweets as well as their rich relationships, we first construct a structured heterogeneous information network (HIN) [67]. Then we have taken a meta-path based approach to formulate similarity measures over users and aggregate different similarities using Laplacian scores. This work was further extended in [11] into *HinOPU* by constructing an ensemble of classifiers to combine different predictions, which improves the accuracy for opioid user detection. Most recently, we have developed a novel and intelligent system named *uStyle-uID* leveraging both writing and photography styles for drug trafficker identification on darkweb [79]. At the core of this new system is attributed heterogeneous information network (AHIN) [46] which elegantly integrates both writing and photography styles along with the text and photo contents, as well as other supporting attributes (i.e., drug dealer and drug product information) and various kinds of relations. However, it is usually hard to detect

Table 1. Existing Drug-related Datasets from Social Media.

Study	Year	Data source	Drug user/Drug dealer	# of positive posts	# of positive user accounts	Public available
[86]	2016	Instagram	Drug user	2,362	406	No
[76]	2017	Instagram	Drug dealer	1,260	27	No
[52]	2018	Twitter	Drug dealer	692	/	No
[29]	2019	Twitter	Drug user	280	/	No
[44]	2019	Instagram	Drug user and dealer	1,228	237	No
IDDIG (Ours)	2020	Instagram	Drug dealer	1,022	1,406	Yes

illicit drug trafficking activities from Instagram data based on these existing methods due to the following reasons. First, Instagram is an image-oriented social media platform, which is different from text-oriented ones such as Twitter. Therefore, it is difficult to directly transfer *AutoDOA* [12] from Twitter to Instagram due to the variation of modality. Second, *uStyle-uID* [79] is developed for darkweb, whose characteristics differ dramatically from those of clearweb - e.g., it is generally difficult for the public to access darkweb; but once it is accessed, it is relatively easy to detect illicit drug trafficking. By contrast, detecting illicit drug trafficking on clearweb is like finding a needle in a haystack. Most existing Instagram-based detection methods can not comprehensively detect constantly evolving drug trafficking patterns or activities. For example, [44, 76] can detect illegal drug trafficking such as Pattern 1 in Fig. 1 (direct posting) with good performance but fail to detect Pattern 2 (indirect commenting).

3 IDDIG DATASET CONSTRUCTION

Data collection has played an increasingly important role in machine learning and data mining research [72]. In this section, we first review several existing datasets related to drug trafficking research. Then we present a new large-scale dataset constructed from Instagram; toward this objective, we will discuss our efforts on data crawling and data annotation, respectively.

3.1 Current State-of-the-Art

Existing detection methods have collected several databases for drug-dealer and drug-user detection. Tab. 1 summarizes the statistics of these datasets for illicit drug dealing tracking. In [86], a dataset containing 2,362 posts and 406 drug users (including drug dealers) was introduced. It also contained drug users' followers and the accounts they follow. 176 users contain geolocation tags in their posts. Based on the post and homepage information on Instagram, a total of 1,260 drug-related posts and 27 drug dealer accounts were collected in [76] for drug dealer detection. Each account in the dataset contains detailed information, including post content, temporal information of each post, and the number of followers/followees. In [52], 692 drug-related tweets were identified as being associated with illegal online marketing and sale of prescription opioids to detect and report illicit online marketing and sales of controlled substances via Twitter. A recent study [29] collected a dataset (with 280 drug-related) for analysis of drug-abuse risk behavior on Twitter. Recently, 1,228 drug dealer posts and 267 drug dealer accounts were collected by [44] and images of different types of drugs were fine-grained and categorized into five different categories.

Although great effort has been devoted to drug-related data collection, there exist the following limitations of existing datasets: (1) The size of these datasets are not large enough to support deep learning-based approaches. For example, the number of drug dealer or user account instances is all less than 500. This is limited by that drug dealers or users make up a tiny fraction of Instagram's user accounts [76], and many heterogeneous data (e.g., text and images) need to be collected,

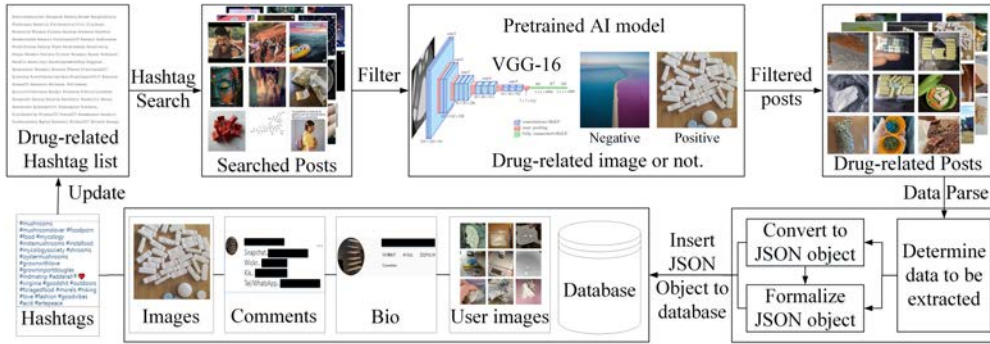


Fig. 2. Automatic and Iterative Hashtag-based Multimodal Data Collection from Instagram.

increasing the cost of storage and management [27]. (2) Due to the privacy protection, the user account number, face related images and contact information (e.g., other chat app user ID, see the black blocks in Fig. 1) are not allowed to be made public. Consequently, no dataset is publicly available to support the research related to drug dealer detection as of today.

3.2 Multi-model Data Collection Scheme on Instagram

To fill in this gap, we have made a great effort on data collection and construction of a large-scale dataset in this project. Existing web crawling techniques (e.g., [58]) do not meet the requirements of this project due to the following technical challenges. First, it is desirable to develop an automatic system capable of identifying drug-related content without human intervention. This is in sharp contrast to previous works (e.g., [76]) that count on humans for data collection, which are difficult to scale up. Second, the designed data crawling system needs to be versatile in that it supports multimodal (including both text and images) and multisource (e.g., posts and comments) collection.

To achieve this objective, we have developed an automatic system capable of iteratively collecting multimodal data, as shown in Fig. 2. The rationale underlying our data crawling system is still based on *hashtag*-based search [19]. Hashtags on Instagram can help users extend their reach, engage their audience, which can be attached to posts and become clickable phrases and topics with the # placed in front of them. However, unlike [76] working with a fixed collection of hashtags, we propose a data crawling algorithm that iteratively expands the pool of hashtags for scaling up our search. Such expansion of hashtags is guided by an intelligent pretrained AI model (VGG-16 [68]) designed for drug image classification. By treating drug-related hashtags and images as a pair of peer hidden variables, our iterative crawling system aims at refining and updating the collected multimodal data in an Expectation-Maximization (EM)-like manner. The detailed description of our data collection system consists of the following four components.

(1) Drug-related hashtags collection. A total of 200 drug-related hashtags have been manually collected by domain experts using the hashtag search API [17]. These hashtags contain various types of drugs, such as 3,4-methylenedioxy-methamphetamine (MDMA), Lysergic acid diethylamide (LSD), codeine, painkiller and xanax etc., which are widely trafficked on Instagram. We have used this set of hashtags as the initial starting point of our data collection.

(2) Drug-related post detection. We search each post (which includes an image and comments) with each drug-related hashtag as input. A VGG-16 based binary classification model [68] is pre-trained to detect drug-related posts from the accompanying image information. The image-based dataset for model pretraining contains various types of drug-related images, which are sources of

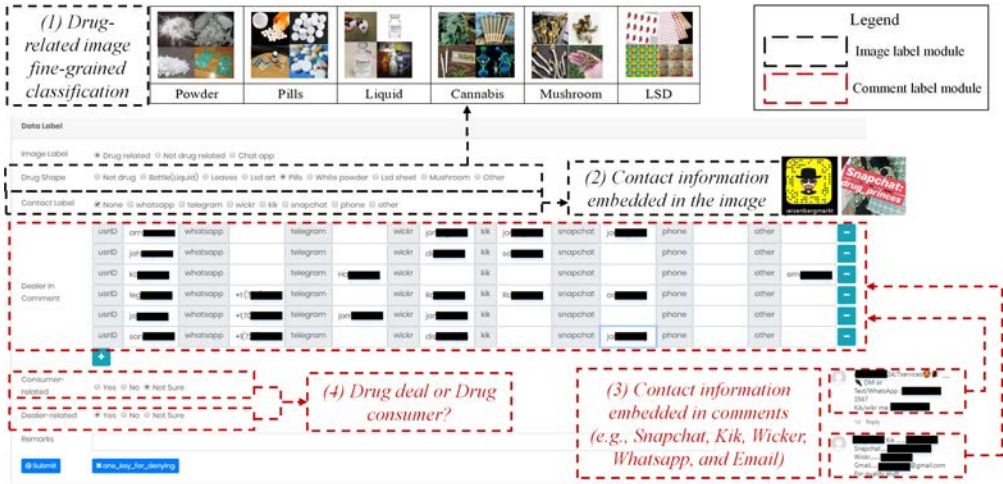


Fig. 3. User Interface of Web-based IDDIG Annotation Platform.

Bing image search API (similar to Google image search API adopted in [76]). If an image of a post is detected by the model as being drug-related (positive), we save its link for further processing.

(3) Drug-related data collection. The detected posts were converted and formalized into a universal Json object [53] to facilitate storage and retrieval. As post comments are sources from several user accounts, we saved each post-related information (including posted images and comments) and homepage information about the user who commented on the post. The user’s homepage information includes both the bio and images from the user’s latest 10 posts (we refer to it as Homepage image in the following). Totally, 10,000 potential posts and 23,034 user homepage information were collected as the initial dataset.

(4) Drug-related hashtag update. New hashtags from each detected post can be added into the list of drug-related hashtags. We have also recorded the frequency of each hashtag to track the most frequent ones. The system uses the new hashtag (which has the highest frequent counts) in the next iteration until the amount of collected data reaches a prespecified threshold (in this study, we have set the threshold to be 1000 drug-dealer accounts).

When compared with previous work [76], we note that the key difference lies in our emphasis on posted comments and biography data (Instagram bio) instead of hashtags and captions as salient features for drug-dealer identification.

3.3 User Interface of Data Annotation Platform

The collected data can not be used as training data without proper annotation (manual labeling). As the collected posts often contain rich drug trafficking information (e.g., drug types, contact information, sales areas, and so on), we have designed a data annotation platform for drug dealer detection, as shown in Fig. 3. The annotation platform mainly contains two modules: the image labeling module and comment labeling module. The image labeling module is designed to label image-related information, including fine-grained classification of drug forms (e.g., powder, pills, liquid, cannabis, mushroom, and LSD as shown in Fig. 3) and contact information (e.g., user ID in the chat app, such as Snapchat, wickr, Kik, What’s up, telegram or email address) embedded in the image. The comment labeling module is designed to label comment-related information, including

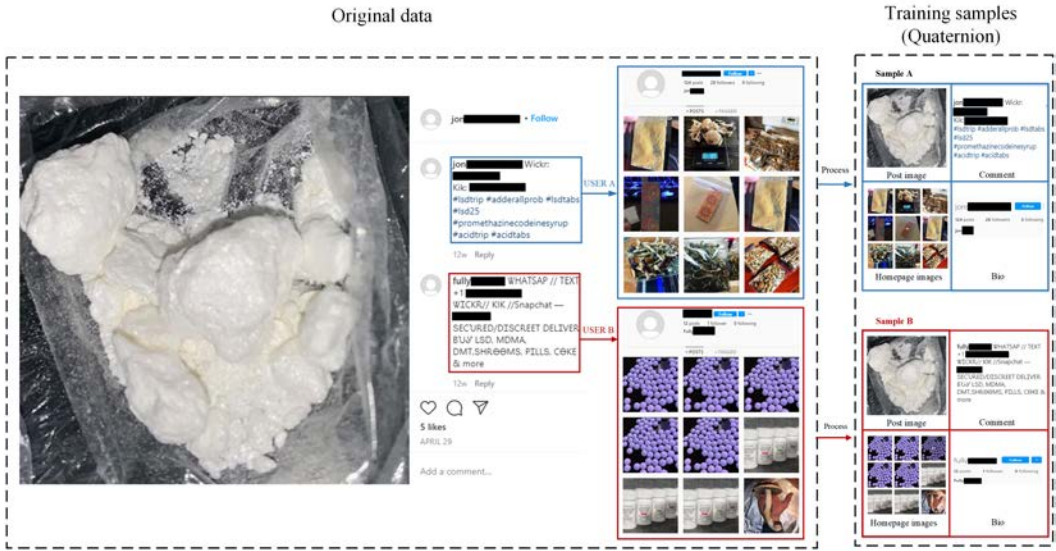


Fig. 4. The Conversion Process from Raw Data in IDDIG Dataset to Quadruple-based Training Samples.

Table 2. Details of the constructed IDDIG Dataset.

	Type	Number	Missing rate
Positive	# of posts	1,022	NA
	# of unique user accounts	1,406	NA
	# of Posted Image	2,815	0%
	# of Posted Comment	2,722	3.30%
	# of Homepage Bio	1,239	55.98%
	# of Homepage Images	12,061	57.15%
Negative	# of posts	1,090	NA
	# of unique user accounts	2,485	NA
	# of Posted Image	3,206	0%
	# of Posted Comment	3,206	0%
	# of Homepage Bio	2,663	16.90%
	# of Homepage Images	24,883	22.39%

contact information embedded in the comment, and whether it is related to a drug dealer or a drug consumer.

In addition, for each post, we have also designed a button to check the homepage of users who comment on the post. Based on the post information and homepage information, the domain expert will add an annotation to the post indicating whether it contains a drug dealer account or not. When multiple drug dealer accounts comment on a post, we process the information into multiple data, as shown in Fig. 4. It takes on average 3-to-5 minutes to label each post; using a crowd-sourcing approach, we have divided the task of data annotation among 10 participants. Overall, we have spent around 400 hours on labeling all collected Instagram data.

In summary, a total of 2,112 posts have been annotated and organized for Identifying Drug Dealer on Instagram (IDDIG). As shown in Tab. 2, IDDIG contains 3,206 negative samples and 2,815

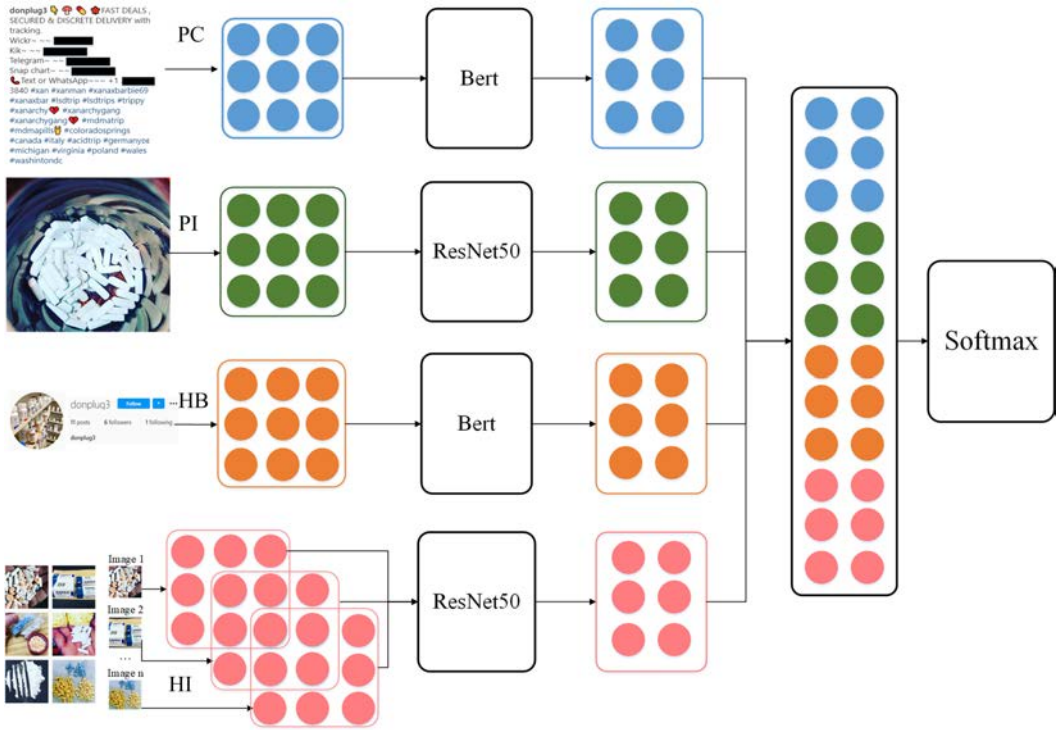


Fig. 5. The framework of our proposed multimodal data fusion approach to illicit drug dealer identification. Our model takes in posted comment (PC), posted image (PI), homepage bio (HB), and homepage images (HIs) associated with Instagram users, and combines multimodal information using a quadruple-based fusion strategy to identify an illicit drug dealer.

positive samples. Each sample was characterized by a quadruple-based data structure, including *Posted Comment (PC)*, *Posted Image (PI)*, *Homepage Bio (HB)*, and *Homepage Image (HI)*. Among 2,815 positive samples, 1,022 positive samples are pattern 1 (trafficking by direct posting) and the other 1,793 positive samples are pattern 2 (trafficking by indirect commenting) as shown in Fig. 1. In addition, Tab. 2 shows that more than 50% positive samples miss homepage bio and homepage images. Most of these missing cases are from pattern 2 samples. To protect the privacy, we manually erase users’ personal information to achieve the purpose of protecting user privacy. Our proposed dataset is publicly available¹.

4 QUADRUPLE-BASED MULTIMODAL FUSION FOR ILLICIT DRUG DEALER IDENTIFICATION

4.1 Problem Formulation and System Overview

Given the background in Section 3, we can formally define the problem of *illicit drug dealer identification* as follows: our aim is to build an effective approach to identify illicit drug dealers from normal users on Instagram. For each user, we have her/his posted comment (PC), posted image (PI), homepage bio (HB), and homepage image (HI). Then each user n is represented with quadruple $u^n = \langle PC^n, PI^n, HB^n, HI^n \rangle$. Let $c^n = 1$ denotes that user N is an illicit drug dealer, $c^n = 0$

¹Tentative downloading link for IDDIG dataset: <https://data.mendeley.com/>.

otherwise. Given a set of n training examples $\mathcal{D} = \{(u^1, c^1), (u^2, c^2), \dots, (u^N, c^N)\}$, we aim at building a predictive model $f : u \rightarrow c$ to identify illicit drug dealers.

We formulate the problem of illicit drug dealer identification as a binary classification problem, namely, the output of our model is the predicted probability \hat{c}^n of a user n being an illicit drug dealer. We have adopted the cross-entropy loss, which has been widely applied in deep learning-based classification [80]. Specifically, the cross-entropy loss function is defined by

$$\ell(\Theta) = - \sum_{n=1}^N c^n \log \hat{c}^n. \quad (1)$$

Fig. 5 shows the framework of our proposed multimodal data fusion approach to illicit drug dealer identification. Our model takes in posted comments (PC), posted images (PI), homepage bio (HB), and homepage images (HI) associated with Instagram users. A quadruple-based fusion strategy is proposed to combine the multimodal information, and the combined feature representation is exploited to identify illicit drug dealers. In this section, we will first discuss the single modality and then elaborate on the strategy of multimodal data fusion.

4.2 Drug Dealer Identification based on Single Modality

4.2.1 Text-based Identification. Text-based information such as hashtags and captions have been used to differentiate drug-related posts from non-drug-related posts in [76]. It has been shown that there is a clear difference between these two classes of posts; but meanwhile it is easy for Instagram platform administrators to automatically detect these drug-related posts from drug dealer. However, increasingly more drug dealers tend to avoid detect by commenting under some hot or drug-related posts (see pattern 2 in Fig. 1). Many cases such cases are collected in our database. As mentioned above, we argue that posted comments and bio information are often more reliable than tags and captions. Therefore, we propose to develop a new text-based classifier based on comments and bio data as follows. To justify the above claim, we have empirically found that comments and biography data posted on Instagram almost always contain many drug-related hashtags, contact information, sales destinations, and payment methods. Such findings are consistent with the common sense that trade-related information, no matter being legal or illicit, has shared characteristics because they are supposed to facilitate the communication between dealers and consumers [44]. Therefore, we have hand-picked posted comments (including the captions associated with posted images) and bio information at the homepage as the two most salient features in text classifier. Unlike [76], we propose to leverage the latest advances in natural language processing into the task of text-based drug-dealer identification.

There have been several deep learning-based models with promising performance in text classification, such as CNN [38], bi-directional long short-term memory (BiLSTM) [84], convolutional long short-term memory (CLSTM) [83], and Bidirectional Encoder Representations from Transformers (BERT) [10]. Among these models, BERT, as a state-of-art text classification model, can consider the full context of a word by looking at the words that come before and after it—particularly useful for understanding the intent behind search queries. BERT uses large-scale corpus to create pretrained language models and fine-tune pretrained models for specific tasks. Therefore, we use the BERT-based model in our method for drug dealer detection. The text information in our IDDIG dataset—namely, comments and bio, are taken as input, respectively, to fine-tune the BERT-based model. The BERT model used is the BERT-based (12-layer, 768-hidden, 12-heads, 110M parameters). To obtain the sentence-level representation, we extract the token embedding of the last layer and compute the mean vector to get the final feature representation of 768 dimensions.

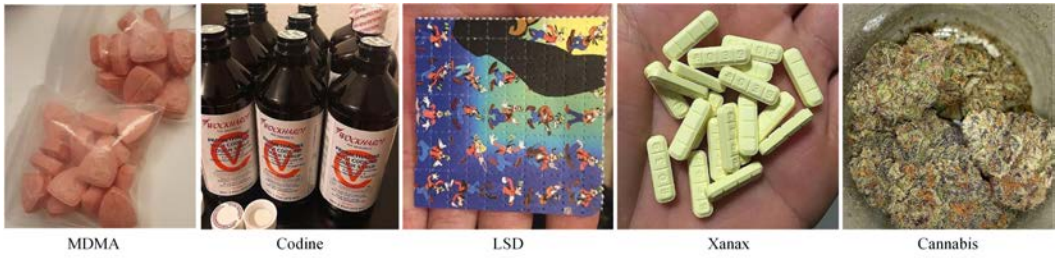


Fig. 6. Sample Post Images from IDDIG Dataset (note that they contain diverse drug-related content, from pills and weeds to syrup and powder).

4.2.2 Image-based Identification. In previous work [76], image-based information has been found more reliable than text-based one for the task of drug-related post recognition. Then, the problem of dealer account detection was solved by extracting features from drug-related posts. Thanks to the construction of IDDIG dataset at a large scale, it is possible to directly exploit image information for the task of drug-dealer identification. Similar to text data, images including both post images and homepage images often contain rich information, such as drug types (as shown in Fig. 6) and contact information of drug dealers (e.g., user ID from Wickr, Snapchat and telegram). Unlike text data, both drug dealers and regular consumers could post similar images containing drug-related content. Therefore, it is plausible to conjecture that image-based information is less reliable than text-based one for the task of drug-dealer identification.

As reported later in Sec. 6, such conjecture has been confirmed by our experimental results. We have used the widely used ResNet-50 [28] for the task of image-based user classification. ResNet-50 is a pretrained deep learning model for image classification with excellent generalization performance on various recognition tasks [18, 36]. Thanks to the identity function introduced to the network, the gradient calculation in back-propagation can flow more effectively, which helps alleviate the notorious vanishing gradient problem. Using the collected IDDIG dataset, we have fine-tuned the pretrained ResNet-50 model on our post images and homepage images, respectively, for drug-dealer detection. The images have been cropped to $224 \times 224 \times 3$, and horizontally flipped randomly for data augmentation. We take these processed images as the input and obtain feature vectors in the latent 2048-dimensional space. To implement transfer learning with fine-tuning [59], we have removed the last predicting layer of the pretrained model and replaced them with task-specific prediction layers (drug-dealer vs. non-drug-dealer).

4.3 Joint Multi-modal Optimization for Drug-Dealer Identification

Multimodal data fusion [42] refers to a class of techniques combining information from multiple modalities for improved performance. In previous work [76], a decision-level (late) fusion strategy was adopted because training a joint model for feature-level (early) fusion faced several technical challenges such as weak correlation in social media data and lack of paired training data. However, as shown in [22], feature-level fusion has some advantages over decision-level fusion, at least in theory. Meanwhile, new multimodal data fusion techniques including feature reduction and concatenation [26], compact and factorized bilinear pooling [18, 37, 47] have been developed for the applications of multimodal biometric recognition and visual question answering.

In this work, we propose to leverage these latest advances in multimodal data fusion and explore their potential of feature-level fusion for drug-dealer identification. To the best of our knowledge, this will be the first systematic study of multimodal data fusion, including both protocol design

and fusion strategies for drug-dealer identification. We have structured our study into two parts. In the first part, we will present three protocols - namely, multimodal post-based identification, multimodal homepage-based identification, and joint quadruple-based optimization identification. In the second part, we will discuss three competing strategies for multimodal data fusion - feature concatenation [26], bilinear pooling [47], and factorized bilinear pooling [18, 37].

4.3.1 Fusion Protocols. Drug dealers usually post both text and images to promote drugs for sale, as well as comments related to product advertisements, sales destinations, and payment methods. Meanwhile, texts and images can be posted either as comments or at the homepage of an Instagram user. Such multimodal and multisource distribution of collected data distinguishes our quadruple-based data representation from previous works. As shown in Fig. 4, quadruple-based data representation allows us to come up with three different fusion protocols, as follows.

1. Multimodal data fusion

This is the scenario that has been considered in previous works [76, 86]. In addition to feature-level fusion, our approach differs from [76, 86] by exploiting the multimodal information contained in both the post and the homepage of Instagram users [30, 35]. The homepage of drug dealers often contains an even richer set of multimodal data than the post, which could serve as supplementary advertising tools. In previous work [76], homepage statistical data, including percentages of drug-related posts, temporal patterns, relational information, and evidence of transactions are combined to identify a drug dealer account. Instead of using the statistical data derived from the homepage, we propose to fuse the bio text and the last ten post images in homepages directly for drug dealer account identification. For multimodal data fusion at the post level, we have used Resnet-50 and BERT to extract 2048-dimensional image-based and 768-dimensional text-based feature vectors, respectively. For multimodal data fusion at the homepage level, we first extract a sequence of ten image-based features from the ten homepage images, respectively. Then the average of these ten features is used as the final image-based feature representation. If the number of homepage images is less than 10, the features are averaged according to the number of homepage images. If homepage images are missing, the image matrix will be filled with 0 to participate in the multimodal fusion training. Similar to post-level fusion, the averaged 2048-dimensional image-based feature vector will be concatenated with a 768-dimensional text-based feature vector to generate a composite 2816-dimensional vector or combined with text-based feature vector by the bilinear pooling method. The fused feature vector is finally fed into the softmax classifier with cross-entropy loss function in Eq. (1) for classifying the drug dealer account.

2. Multi-source data fusion

The collection of multimodal data from different sources (i.e., post vs. homepage) allows us to conduct another line of research on information fusion. Instead of fusing across modalities, it is possible to fuse the information of the same modality but across different sources. Such multisource data fusion [78] has been studied in the field of remote sensing before; under the framework of drug-dealer identification, we believe that multisource data fusion offers complementary insight to multimodal data fusion. Specifically, such comparative studies could help answer the following questions: 1) which modality is more important - image vs. text? 2) which source is more reliable - posts vs. homepage? We conjecture that texts are more important than images due to the intrinsic ambiguity in pictorial representation; posts are more reliable than homepages because posts are often directly used as the advertising tool by drug dealers. Unlike multimodal data fusion, the feature vectors associated with the same modality but from different sources always have the same dimensionality. It is often more convenient to implement FFT-based compact bilinear pooling [15] for a pair of equal-length feature vectors. Note that either multimodal or multisource data fusion

only reflects one perspective that can be biased; it is desirable to jointly exploit them, as we will elaborate next.

3. Joint quadruple-based fusion

In the real world, drug dealers often need to advertise for drugs while evading being caught by law enforcement. Accordingly, the information we have collected about drug dealers from Instagram is often incomplete - i.e., not all quadruple elements in the proposed IDDIG dataset are available for the same account. Such missing data (a.k.a. missing modality) constraint makes it difficult to use only two-element posts or homepages for accurately identifying drug dealers. The method in [76] first fused multimodal post data to filter drug-related posts, and then detected drug dealer accounts based on multimodal homepage fusion. However, this method often fails to distinguish illicit drug dealers from legal regular consumers or drug abuser accounts.

Multimodal data fusion with missing data has been recently studied for deep multimodal encoding in [49]. Inspired by the imputation strategy, in this work, we have designed a joint optimization method for quadruple-based information fusion, as shown in Fig. 5. The features of post-image, comments, bio, and homepage images are first extracted using BERT and ResNet-50 models, respectively. Then the four elementary features are either concatenated into a 5632-dimensional feature vector (the missing element will correspond to an imputation of all zero subvectors) or combined by bilinear pooling operations (note that bilinear pooling can naturally tolerate one missing element and output the available element as the output vector). This way, our quadruple-based fusion can tolerate nine out of a total of 16 different missing patterns. The concatenated or fused features are finally fed into the Softmax for drug dealer detection.

4.3.2 Fusion Strategies. The strategies of multimodal fusion have been extensively studied in the literature (e.g., [2, 42]). Fusion of textual and visual information has been particularly the focus of research on visual question answering [15] and sentiment analysis [7]. When feature vectors associated with textual and visual information have different dimensions, concatenating them becomes the easiest solution, even though such strategy suffers from the potential “curse of dimensionality” [61]. Our experience suggests that such an ad-hoc strategy fits well with the powerful deep learning framework when computational resources are not scarce. A computationally more efficient approach is not to concatenate but to combine multiple feature vectors (assuming normalized to the same length) into a composite vector without the change of dimensionality. This line of research has led to a flurry of bilinear pooling methods, including compact bilinear pooling [15] and factorized bilinear pooling [18].

In this study, we have conducted a comprehensive comparative study of four competing fusion strategies: 1) **Feature concatenation** [26, 82]. Direct concatenation of two feature vectors into a longer one (e.g., $2048 + 768 = 2816$). 2) **Bilinear pooling** [47]. Bilinear pooling was introduced in [47] to provide a robust image representation for fine-grained image classification. In bilinear pooling, two feature vectors are fused by an outer product (or Kroneker product for matrices) -i.e., we can take element-wise interactions between a pair of feature vectors into account as follows:

$$\mathbf{Z} = \sum_{(i,j \in \mathbb{S})} \mathbf{x}_i \mathbf{y}_j^T \quad (2)$$

where $\{\mathbf{x}_i | \mathbf{x}_i \in \mathbb{R}^p, i \in \mathbb{S}\}$, $\{\mathbf{y}_j | \mathbf{y}_j \in \mathbb{R}^q, j \in \mathbb{S}\}$ are two feature vectors, \mathbb{S} is the set of spatial locations (combinations of rows and columns), and $\mathbf{Z} \in \mathbb{R}^{p \times q}$ is the fused feature descriptor. Similar to the kernel expansion, it allows the consideration of all pairwise interactions (e.g., $2048 * 768 = 1572864$). It can be seen that the size of bilinear feature descriptor can be large, which makes it computationally infeasible. 3) **Compact bilinear pooling** [15]. In view of the explosion of dimensionality, a more compact solution to bilinear pooling is developed based on Fast Fourier Transform (FFT). It requires

the two features have identical lengths (e.g., a 768-dimensional vector needs to be resampled to a 2048-dimensional one). 4) **Factorized bilinear coding** [18]. This is one of the latest advances in computationally efficient bilinear pooling. To generate more compact representations, it is possible to employ a factorized bilinear coding (FBC) strategy to more efficiently integrate the features from multimodal data.

The basic idea behind FBC is to encode the features into sparse representations and to learn a dictionary B with k atoms that can be *factorized* into low-rank matrices to capture the sparsity structure of observation data. More specifically, let the dictionary $B = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k] \in \mathbb{R}^{pq \times k}$, FBC proposes to factorize each dictionary atom $\mathbf{b}_l \in \mathbb{R}^{pq}$ ($1 \leq l \leq k$) into $\mathbf{U}_l \mathbf{V}_l^\top$, where $\mathbf{U}_l \in \mathbb{R}^{p \times r}$ and $\mathbf{V}_l \in \mathbb{R}^{q \times r}$ are learnable low-rank matrices. This way, the original bilinear feature $\mathbf{x}_i \mathbf{y}_j^\top$ can be reconstructed by $\sum_{l=1}^k c'_v \mathbf{U}_l \mathbf{V}_l^\top$, with $\mathbf{c}_v \in \mathbb{R}^k$ being the FBC code, and c'_v being the l -th element of \mathbf{c}_v ($1 \leq v \leq N$, N is the number of pairs in \mathbb{S}).

The sparsity-based FBC encodes two input features $(\mathbf{x}_i, \mathbf{y}_j)$ into \mathbf{c}_v by solving the following optimization problem:

$$\min_{\mathbf{c}_v} \left\| \mathbf{x}_i \mathbf{y}_j^\top - \sum_{l=1}^k c'_v \mathbf{U}_l \mathbf{V}_l^\top \right\|^2 + \lambda \|\mathbf{c}_v\|_1 \quad (3)$$

where λ is the Lagrangian multiplier controlling the trade-off between the reconstruction error and the sparsity. The FBC code \mathbf{c}_v can be obtained by the classical LASSO method [75] which produces

$$\begin{cases} \mathbf{c}'_v = \mathbf{P}(\tilde{\mathbf{U}}^\top \mathbf{x}_i \circ \tilde{\mathbf{V}}^\top \mathbf{y}_j), \\ \mathbf{c}_v = \text{sign}(\mathbf{c}'_v) \circ \max((\text{abs}(\mathbf{c}'_v) - \frac{\lambda}{2}), 0). \end{cases} \quad (4)$$

where \circ denotes the Hadamard product, $\mathbf{P} \in \mathbb{R}^{k \times rk}$ is a binary matrix with only elements in the row l , columns $((l-1)r+1)$ to (lr) being "1", and $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are the transformations of \mathbf{U} and \mathbf{V} are in the form of

$$\begin{cases} \tilde{\mathbf{U}}^\top = [\tilde{\mathbf{U}}_l^\top] = [\frac{1}{r} \cdot \mathbf{I}((\mathbf{q}_l \mathbf{1}_{rk}^\top) \circ \mathbf{U}^\top)] \in \mathbb{R}^{rk \times p} \\ \tilde{\mathbf{V}}^\top = [\tilde{\mathbf{V}}_l^\top] = [\frac{1}{r} \cdot \mathbf{I} \mathbf{V}^\top] \in \mathbb{R}^{rk \times q} \end{cases} \quad (5)$$

where $\mathbf{I} \in \mathbb{R}^{r \times rk}$ is an all "1" matrix, \mathbf{q}_l is the l -th column of $\mathbf{Q} = ((\mathbf{P}(\mathbf{U}^\top \mathbf{U} \mathbf{P}^\top \cdot \mathbf{V}^\top \mathbf{V} \mathbf{P}^\top))^{-1} \mathbf{P})^\top$.

5 UNSUPERVISED LEARNING FOR DRUG-DEALER COMMUNITY DETECTION

Community detection in a social network refers to the problem of identifying a cluster of connected vertices forming fairly independent compartments in a graph or a network [54, 55]. Detecting drug-dealer communities (a.k.a. drug trafficking rings) in the real world has been studied by social scientists and criminologists for decades [1]. Finding the community of drug dealers and their customers from social media data such as Instagram has only been recently considered (e.g., [44]). By exploring how the hashtags are connected to Instagram, we make the first step toward revealing the hashtag culture, which might lead to useful hints for online community detection. In other words, the complicated relational characteristics of Instagram hashtag can be discovered through data mining and visualization techniques, allowing a deeper understanding of drug-dealer communities/clusters found across hashtags.

To demonstrate the feasibility of this idea, we have designed an unsupervised learning method using NetworkX [25] - a graph object and a Python language package developed for exploiting and analyzing networks. NetworkX has been widely used for operations on large-scale real-world graphs, such as graphs with over 10 million nodes and 100 million edges [24]. For the IDDIG dataset, we have collected from Instagram, more than 800 posts in our dataset all contain at least one mention of drug-related hashtag (e.g., #pills, #psychedelic, #mdmathrapy, #mdmatrip, #lsdtrip,

Table 3. Text-based Classification for Drug Dealer Detection

Performance	Accuracy	Precision	Recall	F1 score
PC only				
CNN[41]	84.38%	69.12%	95.05%	80.03%
BiLSTM[6]	83.42%	94.00%	75.45%	83.71%
CLSTM[83]	85.12%	70.16%	95.90%	81.04%
BERT[10]	87.17%	93.67%	75.98%	83.90%
HB only				
CNN[41]	84.51%	67.32%	97.82%	79.75%
BiLSTM[6]	84.04%	73.46%	89.42%	80.66%
CLSTM[83]	84.51%	93.47%	70.76%	80.55%
BERT[10]	84.38%	76.52%	93.05%	83.98%

and #sdl). All unique hashtags in the corpus represent nodes in the graph. Edges were formed between hashtags if they were mentioned together in the same post. This yielded an undirected drug-related graph, which was then analyzed by NetworkX. For the task of community detection, we can calculate various graph attributes such as betweenness centrality [4], adjacency matrix [60], clustering coefficient [43], and vertex similarity [14]. The identified communities in a network are often visualized by a tool, so-called Sunburst plot (please refer to Fig. 8 for illustrative examples).

6 EXPERIMENTAL RESULTS

In this section, we first report the performance of single modality methods on our own IDDIG dataset for drug-dealer identification. Different multimodal fusion schemes (protocols and strategies) were then compared and analyzed. Finally, we present our experimental results of drug-dealer community detection using NetworkX and analyzed the characteristics of drug-dealer community on Instagram.

6.1 Experimental Setup

The IDDIG dataset used in our experiments contains 3,206 negative samples and 2,815 positive samples. As shown in Tab. 1, the size of positive user accounts is much larger than all existing datasets. We have split the whole IDDIG dataset into 70% training set and 30% testing set. For post and homepage images, we have adopted a ResNet50 to extract the feature vectors of length 2048; for posted comments and bio texts, we have used BERT pretrained model to extract a 768-dimensional vector. We have trained the proposed multimodal fusion model using the popular Adam optimization algorithm [62] with a mini-batch size of 10 based on IDDIG. The following parameters are adopted in our setting: learning rate $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$. We opt to terminate the training after 50 epochs. All experiments are conducted using PyTorch on a workstation with one RTX 2080 GPU.

6.2 Drug-dealer Identification Results

6.2.1 Single modality based Results for Drug-Dealer Detection. To evaluate the performance of single modality classification at this stage by training and testing on the proposed IDDIG dataset. We have implemented four models (i.e., CNN[41], BiLSTM[6], CLSTM [83] and BERT[10]) for text classification and compare their performance based on the proposed IDDIG dataset. Four different performance metrics (accuracy, precision, recall, and F1 score) are reported in Tab. 3. It can be observed that 1) BERT has the best performance than any other text classification model with

Table 4. Image-based Classification for Drug Dealer Detection

Performance	Accuracy	Precision	Recall	F1 score
PI only				
VGG16 [68]	70.72%	63.44%	70.94%	66.98%
ResNet50 [28]	69.69%	62.61%	68.47%	65.41%
ResNet152 [28]	73.61%	65.96%	76.35%	73.61%
DenseNet121 [31]	72.78%	67.49%	67.49%	67.49%
HI only				
VGG16 [68]	83.25%	74.94%	93.05%	83.02%
ResNet50 [28]	84.04%	76.11%	92.90%	83.67%
ResNet152 [28]	83.58%	75.09%	93.80%	83.41%
DenseNet121 [31]	83.37%	75.93%	93.61%	83.27%

83.90% F1 score on comment and 83.98% F1 score on bio; this is not surprising because BERT represents the current state-of-the-art in natural language processing. 2) comment-based and bio-based textural information have similar discrimination power because they have both achieved over 80% F1 score. 3) two LSTM models (BiLSTM[6] and CLSTM [83]) have demonstrated strikingly different precision/recall results, which reflect the trade-off between these two objectives.

For image-based classification, we have trained four models(i.e., VGG-16 [68], ResNet-50 [28], ResNet-152 [28], and DenseNet121 [31]) to classify our post images and homepage images. As shown in Tab. 4, the best performance for post image classification is achieved by ResNet-152 with over 73% accuracy and 73% F1 score and the best performance for homepage image classification is achieved by ResNet-50 with over 84% accuracy and 83% F1 score. When compared with homepage images, we observe that the performance of image-based classification is noticeably lower for post images. Such an experimental finding is consistent with our common sense because homepage is often a more reliable source than post for drug-dealer identification (note that we have used ten images in the homepage). In addition, our experimental results show that homepage image-based classifications have achieved comparable performance to text-based (closer to bio-based than post-based). Previous study [76] has demonstrated that accounts with a higher value of drug-related percentage are more likely to be drug dealers. Our study further corroborates such findings thanks to numerous drug-related images posted at the homepage, which justifies the benefit of our data collection from the homepage.

6.2.2 Fusion based Results for Drug-Dealer Detection. To compare and analyze different multimodal data fusion methods, we have designed three fusion experimental schemes including data-source-based fusion, data-type-based fusion, and quadruple-based fusion. We will first report our experimental results in correspondence with the three protocols discussed in Sec. 4.3.1. Then, we will compare different fusion strategies, as covered in Sec. 4.3.2.

(1) Multi-modal data fusion: In a previous study [76], it has been shown that the fusion of image and text data can improve the performance of drug-related post recognition. Since we have collected multimodal data from different sources (post vs. homepage), we have conducted the fusion experiments at both post-level and homepage level. The experiment results are as shown at the top row in Table 5 have confirmed the benefit of multimodal data fusion in terms of significant performance improvement. Moreover, our results have demonstrated that post-based fusion may have the better performance with over 94% accuracy and 94% F1 Score than homepage-based fusion. One possible explanation for such findings is that textual information plays a more dominating role in drug-dealer identification than visual information. For example, contact information along

Table 5. Comparison of Different Fusion Protocols (top-down): Multi-modal Data Fusion, Multi-source Data Fusion, and Quadruple-based Fusion.

Performance		Accuracy	Precision	Recall	F1 score
Post-level fusion	PI + PC	94.95%	93.47%	95.17%	94.31%
Homepage-level fusion	HI + HB	91.29%	88.20%	92.60%	90.35%
Text-based fusion	PC + HB	93.68%	90.44%	95.77%	93.03%
Image-based fusion	PI + HI	92.95%	87.57%	97.89%	92.44%
Quadruple-based fusion (Decision-level fusion)	PI + PC + HI + HB	94.48%	96.92%	90.33%	93.51%
Quadruple-based fusion (Feature-level fusion)	PI + PC + HI + HB	95.55%	96.41%	93.35%	94.86%

such as user ID is often sufficient because only drug dealers prefer financial incentives to privacy concerns.

Table 6. Comparison of Experimental Results between Feature Concatenation and Various Bilinear Pooling Methods.

Performance	Accuracy	Precision	Recall	F1 score
Bilinear Pooling	94.02%	95.40%	90.79%	93.03%
Compact Bilinear Pooling	93.62%	92.88%	92.60%	92.74%
Factorized Bilinear Pooling	92.02%	93.85%	87.61%	90.63%
Concat (ours)	94.95%	93.47%	95.17%	94.31%

(2) Multi-source data fusion: An alternative fusion protocol is across different sources instead of modalities. We have implemented text-based feature fusion with 768 + 768 dimensions by concatenating two BERT model output vectors based on the comment and bio in IDDIG dataset. We have also implemented image-based feature fusion with 2048 + 2048 dimensions by concatenating two ResNet model output vectors based on the post-image and homepage images. It can be seen that text-based feature fusions produce better classification results with over 93% accuracy and 93% F1 score than image-based feature fusion, as shown in the middle row in Tab. 5. It is interesting to note that the recall performance is notably better than the precision performance for image-based fusion. Such low-precision and high-recall observation is consistent with the nature of visual information - i.e., it is too liberal for the task of identifying drug dealers. By contrast, textual information is more conservative and favorable to the precision metric as adopted by a related work [44]. This is because text data on Instagram do contain more direct drug-trafficking related information, including various uses of drug-related hashtags, evidence of transactions and payment methods.

(3) Quadruple-based fusion: Similar to the previous work [22], we have implemented and compared feature-level fusion (ours) with decision-level fusion (adopted by [76]) on the IDDIG dataset. In our implementation, a linear weighting method is used for decision-level fusion (similar to [76]). Without any prior knowledge, we have selected 25% weights for post-image-image-based classifier, comment-based classifier, bio-based classifier, and homepage image-based classifier, respectively. The performance comparison result of the two fusion methods (decision-level vs. feature-level) is shown at the bottom row in Table 5. It demonstrated that the proposed feature-level fusions in this paper have better accuracy and F1 score results than decision-level fusion. Feature-level fusions

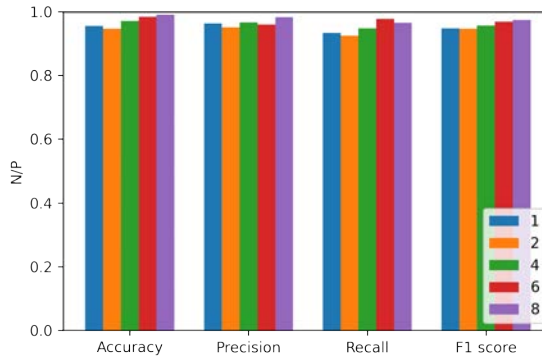


Fig. 7. Impacts of Negative/Positive Ratio on performances of illicit drug dealer identification.

have some advantages over decision-level fusion strategies, in that the process of decision-making cannot increase the amount of information contained in the input features (due to data processing inequality [9]).

(4) Different feature-level fusion strategies: As discussed in Sec. 4.3.2, information fusion at the feature level can be implemented by several strategies. As of today, there still lacks a rigorous theory for evaluating and comparing these competing strategies. Therefore, we have opted to take an empirical approach in this study. Four competing fusion strategies have been implemented: feature concatenation [26], bilinear pooling [47], compact bilinear pooling [15], and factorized bilinear pooling[18]. As shown in Tab. 6, feature concatenation has achieved the overall best performance (only falls behind bilinear pooling on the precision metric). Such findings suggest that brute-force concatenation is capable of outperforming more delicate pooling as long as the increase of dimensionality remains manageable.

6.3 Impact of negative-positive ratio

In the real world scenario, the portion of drug dealers on Instagram is small, which means the dataset can be highly imbalanced. To make our dataset closer to the real world distribution (the number of negative samples is larger than positive), we have collected many nondrug dealers and their associated data on Instagram. Such data augmentation allows us to study the impact of negative-positive ratio on the identification performance of our system. Specifically, we have added different number of negative samples to compare and evaluate the robustness of quadruple-based fusion method. The ratio of negative data over positive data (N/P) is set to 2:1, 4:1, 6:1 and 8:1, and the identification results are shown in Fig. 7. We found our model is not sensitive to negative-positive ratios in the testing range (note that a much larger ratio than 8 might better characterize the challenge of “finding a needle in a haystack” in practice).

6.4 Drug-related community detection and visualization

The constructed IDDIG dataset also allows us to discover drug-related community structures, which might shed novel insights on the dynamics and evolution of illicit drug trade in the real world. In our experiment, we have first conducted a hashtag-based community detection algorithm using the existing tool NetworkX [25]. The default parameter setting has been used (e.g., at most 10 most frequent nodes in each cluster will be preserved). Fig. 8 a) shows the sunburst plot of

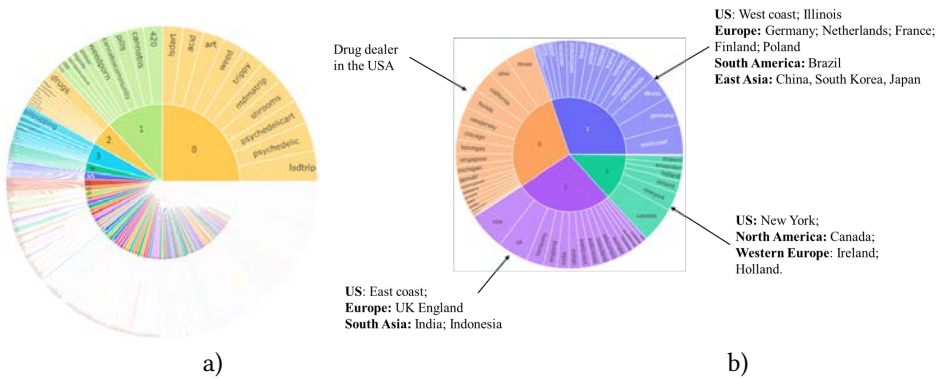


Fig. 8. Sunburst Plots of Community Detection from IDDIG Dataset. a) Network Structure Discovered by Drug-type Hashtag; b) Community Structure Detected by Geography Information Using Hashtag Xanax.

IDDIG dataset organized by different drug types. It can be observed that the class of Lysergic acid diethylamide (LSD) represents the most popular community on Instagram (accounting for almost 25% of hashtags). This is a novel discovery, supporting the strong correlation with the young age of Instagram users, such as teenagers. Based on the hashtag “#Xanax”, we have further analyzed the community structure based on the available geography information (i.e., place-related hashtags). Fig. 8 b) shows the sunburst plot of Xanax hot spots discovered from IDDIG data. We note that the states of Texas, Ohio, and California are the top three ranked in terms of drug-dealer locations. In the meantime, Instagram data also reveals worldwide connections of illicit drug trade between US and other countries in Europe, South America, and Asia.

7 CONCLUSION

Concluding Remarks. In this study, we have collected and constructed a large-scale dataset (IDDIG) from Instagram data to support the research related to drug-dealer identification. Our dataset includes both textual and visual information contained in posted comments as well as at the homepage. An automatic hashtag-based iterative data crawling system and a user-friendly interactive web-based data annotation system were presented. Our data crawling and annotation systems allow us to build a dataset with thousands of positive and negative samples. Based on the constructed IDDIG dataset, we have developed a quadruple-based data representation for drug-dealer identification. BERT and ResNet models are adopted to extract text-based and image-based features from quadruple-based data representation, respectively. We have conducted a comprehensive study of multimodal data fusion for the identification of drug dealers and verified the performance improvement due to quadruple-based fusion. The overall accuracy of our drug-dealer identification has reached almost 95% on the IDDIG dataset, according to our experimental results. Furthermore, we have detected the drug dealer account community by mining hashtag culture. The experiments show that our proposed approach is effective and able to detect drug dealer accounts with high accuracy thanks to data fusion across modalities and sources.

Limitations and Future Research. There are a few limitations in this work and interesting future research directions. First, missing modalities are common in the dataset. We adopted an ad-hoc imputation strategy for the missing modality problem in this work. An interesting future work is to explore more efficient fusion strategies with incomplete modalities. Generative adversarial network (GAN)-based approaches have shown promising performance on handling missing modalities in [66]. However, how to address this issue under the framework of multimodal data fusion has

remained open [51]. Second, although we have assessed the impact of negative-positive ratio on performances of illicit drug dealer identification, in real practice, the negative-positive ratio can be much more negatively skewed. The model can be better evaluated with negative-positive ratio that approximates the real ratio on Instagram. Third, robust detection of drug-related communities such as drug ring especially how to establish the connection between darkweb and clearweb is a challenging problem, which will be left for future studies.

ACKNOWLEDGMENTS

This work is partially supported by the NSF under grants IIS-2107172, IIS-2027127, IIS-2040144, CNS-2034470, IIS-1951504, CNS-1940859, CNS-1814825 and OAC-1940855, and the DoJ/NIJ under grant NIJ 2018-75-CX-0032.

REFERENCES

- [1] Patricia A Adler. 1993. *Wheeling and dealing: An ethnography of an upper-level drug dealing and smuggling community*. Columbia University Press.
- [2] Pradeep K Atrey, M Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S Kankanhalli. 2010. Multimodal fusion for multimedia analysis: a survey. *Multimedia systems* 16, 6 (2010), 345–379.
- [3] Geoffrey Barbier and Huan Liu. 2011. Data mining in social media. In *Social network data analytics*. Springer, 327–352.
- [4] Ulrik Brandes. 2001. A faster algorithm for betweenness centrality. *Journal of mathematical sociology* 25, 2 (2001), 163–177.
- [5] Cody Buntain and Jennifer Golbeck. 2015. This is your Twitter on drugs: Any questions?. In *Proceedings of the 24th international conference on World Wide Web*. 777–782.
- [6] Tao Chen, Ruifeng Xu, Yulan He, and Xuan Wang. 2017. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications* 72 (2017), 221–230.
- [7] Xingyue Chen, Yunhong Wang, and Qingjie Liu. 2017. Visual and textual sentiment analysis using deep fusion convolutional neural networks. In *2017 IEEE International Conference on Image Processing (ICIP)*. 1557–1561.
- [8] Rion Brattig Correia, Lang Li, and Luis M Rocha. 2016. Monitoring potential drug interactions and reactions via network analysis of instagram user timelines. In *Biocomputing 2016: Proceedings of the Pacific Symposium*. World Scientific, 492–503.
- [9] Thomas M Cover. 1999. *Elements of information theory*. John Wiley & Sons.
- [10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [11] Yujie Fan, Yiming Zhang, Yanfang Ye, and Xin Li. 2018. Automatic Opioid User Detection from Twitter: Transductive Ensemble Built on Different Meta-graph Based Similarities over Heterogeneous Information Network. 3357–3363.
- [12] Yujie Fan, Yiming Zhang, Yanfang Ye, Xin Li, and Wanhong Zheng. 2017. Social media for opioid addiction epidemiology: Automatic detection of opioid addicts from twitter and case studies. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1259–1267.
- [13] Golnoosh Farnadi, Jie Tang, Martine De Cock, and Marie-Francine Moens. 2018. User profiling through deep multimodal fusion. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 171–179.
- [14] Santo Fortunato. 2010. Community detection in graphs. *Physics reports* 486, 3-5 (2010), 75–174.
- [15] Akira Fukui, Dong Huk Park, Daylen Yang, Anna Rohrbach, Trevor Darrell, and Marcus Rohrbach. 2016. Multimodal compact bilinear pooling for visual question answering and visual grounding. *arXiv preprint arXiv:1606.01847* (2016).
- [16] Jing Gao, Peng Li, Zhikui Chen, and Jianing Zhang. 2020. A survey on deep learning for multimodal data fusion. *Neural Computation* 32, 5 (2020), 829–864.
- [17] Yuqi Gao, Jitao Sang, Tongwei Ren, and Changsheng Xu. 2017. Hashtag-centric immersive search on social media. In *Proceedings of the 25th ACM international conference on Multimedia*. 1924–1932.
- [18] Zhi Gao, Yuwei Wu, Xiaoxun Zhang, Jindou Dai, Yunde Jia, and Mehrtash Harandi. 2020. Revisiting Bilinear Pooling: A Coding Perspective.. In *AAAI*. 3954–3961.
- [19] Frédéric Godin, Viktor Slavkovikj, Wesley De Neve, Benjamin Schrauwen, and Rik Van de Walle. 2013. Using topic models for twitter hashtag recommendation. In *Proceedings of the 22nd International Conference on World Wide Web*. 593–596.
- [20] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.
- [21] Shannon Greenwood, Andrew Perrin, and Maeve Duggan. 2016. Social media update 2016. *Pew Research Center* 11, 2 (2016).

- [22] Ajith H. Gunatilaka and Brian A. Baertlein. 2001. Feature-level and decision-level fusion of noncoincidentally sampled sensors for land mine detection. *IEEE transactions on pattern analysis and machine intelligence* 23, 6 (2001), 577–589.
- [23] Ido Guy, Naama Zwerdling, Inbal Ronen, David Carmel, and Erel Uziel. 2010. Social media recommendation based on people and tags. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. 194–201.
- [24] A Hagberg and D Conway. 2010. Hacking social networks using the Python programming language. *Sunbelt 2010, Riva del Garda, Italy* (2010).
- [25] Aric Hagberg, Pieter Swart, and Daniel S Chult. 2008. *Exploring network structure, dynamics, and function using NetworkX*. Technical Report. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
- [26] Mohammad Haghighat, Mohamed Abdel-Mottaleb, and Wade Alhalabi. 2016. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. *IEEE Transactions on Information Forensics and Security* 11, 9 (2016), 1984–1996.
- [27] Saeed Hassanpour, Naofumi Tomita, Timothy DeLise, Benjamin Crosier, and Lisa A Marsch. 2019. Identifying substance use risk based on deep neural networks and Instagram social media data. *Neuropsychopharmacology* 44, 3 (2019), 487–494.
- [28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [29] Han Hu, NhatHai Phan, Soon A Chun, James Geller, Huy Vo, Xinyue Ye, Ruoming Jin, Kele Ding, Deric Kenne, and Dejing Dou. 2019. An insight analysis and detection of drug-abuse risk behavior on Twitter with self-taught deep learning. *Computational Social Networks* 6, 1 (2019), 10.
- [30] Yuheng Hu, Lydia Manikonda, and Subbarao Kambhampati. 2014. What we instagram: A first analysis of instagram photo content and user types. In *Eighth International AAAI conference on weblogs and social media*.
- [31] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.
- [32] Caitlin E Hughes, David A Bright, and Jenny Chalmers. 2017. Social network analysis of Australian poly-drug trafficking networks: How do drug traffickers manage multiple illicit drugs? *Social Networks* 51 (2017), 135–147.
- [33] Chia-Chuan Hung, Yi-Ching Huang, Jane Yung-jen Hsu, and David Kuan-Chun Wu. 2008. Tag-based user profiling for social media recommendation. In *Workshop on Intelligent Techniques for Web Personalization & Recommender Systems at AAAI*, Vol. 8. 49–55.
- [34] Kazushi Ikeda, Gen Hattori, Chihiro Ono, Hideki Asoh, and Teruo Higashino. 2013. Twitter user profiling based on text and community mining for market analysis. *Knowledge-Based Systems* 51 (2013), 35–47.
- [35] Jin Yea Jang, Kyungsik Han, Patrick C Shih, and Dongwon Lee. 2015. Generation like: Comparative characteristics in instagram. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 4039–4042.
- [36] Shan Jia, Chuanbo Hu, Guodong Guo, and Zhengquan Xu. 2019. A database for face presentation attack using wax figure faces. In *International Conference on Image Analysis and Processing*. Springer, 39–47.
- [37] Shan Jia, Xin Li, Chuanbo Hu, Guodong Guo, and Zhengquan Xu. 2020. 3D Face Anti-Spoofing with Factorized Bilinear Coding. *IEEE Transactions on Circuits and Systems for Video Technology* (2020).
- [38] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188* (2014).
- [39] Janani Kalyanam and Tim K Mackey. 2017. A review of digital surveillance methods and approaches to combat prescription drug abuse. *Current Addiction Reports* 4, 4 (2017), 397–409.
- [40] Beau Kilmer, Susan S Sohler Everingham, Jonathan P Caulkins, Gregory Midgette, Rosalie Liccardo Pacula, Peter Reuter, Rachel M Burns, Bing Han, and Russell Lundberg. 2014. How big is the US market for illegal drugs? (2014).
- [41] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014).
- [42] Dana Lahat, Tülay Adalı, and Christian Jutten. 2015. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proc. IEEE* 103, 9 (2015), 1449–1477.
- [43] Andrea Lancichinetti and Santo Fortunato. 2009. Community detection algorithms: a comparative analysis. *Physical review E* 80, 5 (2009), 056117.
- [44] Jiawei Li, Qing Xu, Neal Shah, and Tim K Mackey. 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 (2019), e13803.
- [45] Rui Li, Shengjie Wang, Hongbo Deng, Rui Wang, and Kevin Chen-Chuan Chang. 2012. Towards social user profiling: unified and discriminative influence model for inferring home locations. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1023–1031.
- [46] Xiang Li, Yao Wu, Martin Ester, Ben Kao, Xin Wang, and Yudian Zheng. 2017. Semi-supervised clustering in attributed heterogeneous information networks. In *Proceedings of the 26th International Conference on World Wide Web*. 1621–1629.

- [47] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. 2015. Bilinear cnn models for fine-grained visual recognition. In *Proceedings of the IEEE international conference on computer vision*. 1449–1457.
- [48] Guang Ling, Michael R Lyu, and Irwin King. 2014. Ratings meet reviews, a combined approach to recommend. In *Proceedings of the 8th ACM Conference on Recommender systems*. 105–112.
- [49] Zuozhu Liu, Wenyu Zhang, Shaowei Lin, and Tony QS Quek. 2017. Heterogeneous sensor data fusion by deep multimodal encoding. *IEEE Journal of Selected Topics in Signal Processing* 11, 3 (2017), 479–491.
- [50] David D Luxton, Jennifer D June, and Jonathan M Fairall. 2012. Social media and suicide: a public health perspective. *American journal of public health* 102, S2 (2012), S195–S200.
- [51] Mengmeng Ma, Jian Ren, Long Zhao, Sergey Tulyakov, Cathy Wu, and Xi Peng. 2021. SMIL: Multimodal Learning with Severely Missing Modality. *arXiv preprint arXiv:2103.05677* (2021).
- [52] Tim Mackey, Janani Kalyanam, Josh Klugman, Ella Kuzmenko, and Rashmi Gupta. 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 (2018), e10029.
- [53] Kazuaki Maeda. 2012. Performance evaluation of object serialization libraries in XML, JSON and binary formats. In *2012 Second International Conference on Digital Information and Communication Technology and its Applications (DICTAP)*. 177–182.
- [54] Mark EJ Newman. 2006. Modularity and community structure in networks. *Proceedings of the national academy of sciences* 103, 23 (2006), 8577–8582.
- [55] Mark EJ Newman and Michelle Girvan. 2004. Finding and evaluating community structure in networks. *Physical review E* 69, 2 (2004), 026113.
- [56] Liqiang Nie, Xuemeng Song, and Tat-Seng Chua. 2016. Learning from multiple social networks. *Synthesis lectures on information concepts, retrieval, and services* 8, 2 (2016), 1–118.
- [57] Brendan O’Connor, Ramnath Balasubramanian, Bryan R Routledge, and Noah A Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth international AAAI conference on weblogs and social media*.
- [58] Christopher Olston and Marc Najork. 2010. *Web crawling*. Now Publishers Inc.
- [59] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.
- [60] Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, and Ploutarchos Spyridonos. 2012. Community detection in social media. *Data Mining and Knowledge Discovery* 24, 3 (2012), 515–554.
- [61] Tomaso Poggio, Hrushikesh Mhaskar, Lorenzo Rosasco, Brando Miranda, and Qianli Liao. 2017. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review. *International Journal of Automation and Computing* 14, 5 (2017), 503–519.
- [62] Sashank J Reddi, Satyen Kale, and Sanjiv Kumar. 2019. On the convergence of adam and beyond. *arXiv preprint arXiv:1904.09237* (2019).
- [63] Abeed Sarker, Annika DeRoos, and Jeanmarie Perrone. 2020. Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework. *Journal of the American Medical Informatics Association* 27, 2 (2020), 315–329.
- [64] Abeed Sarker, Graciela Gonzalez-Hernandez, Yucheng Ruan, and Jeanmarie Perrone. 2019. Machine learning and natural language processing for geolocation-centric monitoring and characterization of opioid-related social media chatter. *JAMA network open* 2, 11 (2019), e1914672–e1914672.
- [65] David Meerman Scott. 2015. *The new rules of marketing and PR: How to use social media, online video, mobile applications, blogs, news releases, and viral marketing to reach buyers directly*. John Wiley & Sons.
- [66] Chao Shang, Aaron Palmer, Jiangwen Sun, Ko-Shin Chen, Jin Lu, and Jinbo Bi. 2017. VIGAN: Missing view imputation with generative adversarial networks. In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 766–775.
- [67] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and S Yu Philip. 2016. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* 29, 1 (2016), 17–37.
- [68] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [69] Xuemeng Song, Zhao-Yan Ming, Liqiang Nie, Yi-Liang Zhao, and Tat-Seng Chua. 2016. Volunteerism tendency prediction via harvesting multiple social networks. *ACM Transactions on Information Systems (TOIS)* 34, 2 (2016), 1–27.
- [70] Xuemeng Song, Liqiang Nie, Luming Zhang, Mohammad Akbari, and Tat-Seng Chua. 2015. Multiple social network learning and its application in volunteerism tendency prediction. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 213–222.
- [71] Xuemeng Song, Liqiang Nie, Luming Zhang, Maofu Liu, and Tat-Seng Chua. 2015. Interest inference via structure-constrained multi-source multi-task learning. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

- [72] Stefan Stieglitz, Milad Mirbabaie, Björn Ross, and Christoph Neuberger. 2018. Social media analytics—Challenges in topic discovery, data collection, and data preparation. *International journal of information management* 39 (2018), 156–168.
- [73] Yizhou Sun and Jiawei Han. 2012. Mining heterogeneous information networks: principles and methodologies. *Synthesis Lectures on Data Mining and Knowledge Discovery* 3, 2 (2012), 1–159.
- [74] Rosemary Thackeray, Brad L Neiger, Amanda K Smith, and Sarah B Van Wagenen. 2012. Adoption and use of social media among public health departments. *BMC public health* 12, 1 (2012), 1–6.
- [75] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 1 (1996), 267–288.
- [76] Xitong Yang and Jiebo Luo. 2017. Tracking illicit drug dealing and abuse on Instagram using multimodal analysis. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, 4 (2017), 1–15.
- [77] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. 2014. *Social media mining: an introduction*. Cambridge University Press.
- [78] Jixian Zhang. 2010. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion* 1, 1 (2010), 5–24.
- [79] Yiming Zhang, Yujie Fan, Wei Song, Shifu Hou, Yanfang Ye, Xin Li, Liang Zhao, Chuan Shi, Jiabin Wang, and Qi Xiong. 2019. Your style your identity: Leveraging writing and photography styles for drug trafficker identification in darknet markets over attributed heterogeneous information network. In *The World Wide Web Conference*. 3448–3454.
- [80] Zhilu Zhang and Mert R Sabuncu. 2018. Generalized cross entropy loss for training deep neural networks with noisy labels. *arXiv preprint arXiv:1805.07836* (2018).
- [81] Fengpan Zhao, Pavel Skums, Alexander Zelikovsky, Eric L Sevigny, Monica Haavisto Swahn, Sheryl M Strasser, Yan Huang, and Yubao Wu. 2020. Computational approaches to detect illicit drug ads and find vendor communities within social media platforms. *IEEE/ACM transactions on computational biology and bioinformatics* (2020).
- [82] Qinghai Zheng, Jihua Zhu, Zhongyu Li, Shanmin Pang, Jun Wang, and Yaochen Li. 2020. Feature concatenation multi-view subspace clustering. *Neurocomputing* 379 (2020), 89–102.
- [83] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. 2015. A C-LSTM neural network for text classification. *arXiv preprint arXiv:1511.08630* (2015).
- [84] Peng Zhou, Zhenyu Qi, Suncong Zheng, Jiaming Xu, Hongyun Bao, and Bo Xu. 2016. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. *arXiv preprint arXiv:1611.06639* (2016).
- [85] Yiheng Zhou, Numair Sani, Chia-Kuei Lee, and Jiebo Luo. 2016. Understanding illicit drug use behaviors by mining social media. *arXiv preprint arXiv:1604.07096* (2016).
- [86] Yiheng Zhou, Numair Sani, and Jiebo Luo. 2016. Fine-grained mining of illicit drug use patterns using social multimedia data from Instagram. In *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 1921–1930.