A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining

Hong-Han Shuai, Chih-Ya Shen, De-Nian Yang, *Senior Member, IEEE*, Yi-Feng Carol Lan, Wang-Chien Lee, *Member, IEEE*, Philip S. Yu, *Fellow, IEEE*, and Ming-Syan Chen, *Fellow, IEEE*

Abstract—The explosive growth in popularity of social networking leads to the problematic usage. An increasing number of social network mental disorders (SNMDs), such as Cyber-Relationship Addiction, Information Overload, and Net Compulsion, have been recently noted. Symptoms of these mental disorders are usually observed passively today, resulting in delayed clinical intervention. In this paper, we argue that mining online social behavior provides an opportunity to actively identify SNMDs at an early stage. It is challenging to detect SNMDs because the mental status cannot be directly observed from online social activity logs. Our approach, new and innovative to the practice of SNMD detection, does not rely on self-revealing of those mental factors via questionnaires in Psychology. Instead, we propose a machine learning framework, namely, *Social Network Mental Disorder Detection (SNMDD)*, that exploits features extracted from social network data to accurately identify potential cases of SNMDs. We also exploit multi-source learning in SNMDD and propose a new SNMD-based Tensor Model (STM) to improve the accuracy. To increase the scalability of STM, we further improve the efficiency with performance guarantee. Our framework is evaluated via a user study with 3,126 online social network users. We conduct a feature analysis, and also apply SNMDD on large-scale datasets and analyze the characteristics of the three SNMD types. The results manifest that SNMDD is promising for identifying online social network users with potential SNMDs.

Index Terms—Tensor factorization acceleration, online social network, mental disorder detection, feature extraction

1 INTRODUCTION

With the explosive growth in popularity of social networking and messaging apps, online social networks (OSNs) have become a part of many people's daily lives. Most research on social network mining focuses on discovering the knowledge behind the data for improving people's life. While OSNs seemingly expand their users' capability in increasing social contacts, they may actually decrease the face-to-face interpersonal interactions in the real world. Due to the epidemic scale of these phenomena, new terms such as Phubbing (Phone Snubbing) and Nomophobia (No Mobile Phone Phobia) have been created to describe those who cannot stop using mobile social networking apps.

In fact, some social network mental disorders (SNMDs), such as Information Overload and Net Compulsion [53], have been recently noted.[1] For example, studies point out that 1 in 8 Americans suffer from problematic Internet use.[2] Moreover, leading journals in mental health, such as the American Journal of Psychiatry [8], have reported that the SNMDs may incur excessive use, depression, social withdrawal, and a range of other negative repercussions.

Indeed, these symptoms are important components of diagnostic criteria for SNMDs [52] e.g., excessive use of social networking apps—usually associated with a loss of the sense of time or a neglect of basic drives, and withdrawal—including feelings of anger, tension, and/or depression when the computer/apps are inaccessible. SNMDs are social-oriented and tend to happen to users who usually interact with others via online social media. Those with SNMDs usually lack offline interactions, and as a result seek cyber-relationships to compensate. Today, identification of potential mental disorders often falls on the shoulders of supervisors (such as teachers or parents) passively. However, since there are very few notable physical risk factors, the patients usually do not actively seek medical or psychological services. Therefore,

- H.-H. Shuai is with the Department of Electrical Computer Engineering, National Chiao Tung University, No. 1001, University Road, Hsinchu 300, Taiwan. E-mail: hhshuai@nctu.edu.tw.
- C.-Y. Shen is with the Department of Computer Science, National Tsing Hua University, Hsinchu 300, Taiwan. E-mail: chihya@cs.nthu.edu.tw.
- D.-N. Yang is with the Research Center of Information Technology Innovation, Academia Sinica, No. 128, Sec. 2, Academia Road, Taipei 11529, Taiwan. E-mail: dnyang@citi.sinica.edu.tw.
- Y.-F.C. Lan is with the Graduate Institute of Educational Psychology and Counseling, Tamkang University, New Taipei City 251, Taiwan. E-mail: carolyflan@gmail.com.
- W.-C. Lee is with the Department of Computer Science and Engineering, Pennsylvania State University, PA 16801. E-mail: wlee@cse.psu.edu.
- P.S. Yu is with the Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607, and the Institute for Data Science, Tsinghua University, Beijing 100084, China. E-mail: psyu@uic.edu.
- M.-S. Chen is with the Research Center of Information Technology Innovation, Academia Sinica, No. 128, Sec. 2, Academia Road, Taipei 11529, Taiwan, and the Department of Electrical Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan. E-mail: mschen@citi.sinica.edu.tw.

Manuscript received 6 Feb. 2017; revised 14 Sept. 2017; accepted 16 Sept. 2017. Date of publication 25 Dec. 2017; date of current version 1 June 2018. (Corresponding author: Ming-Syan Chen.) Recommended for acceptance by H. Xiong. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TKDE.2017.2786695

1. http://phys.org/news/2015-09-social-media-impacts-mental-well-being.html
2. http://netaddiction.com/faqs/

patients would only seek clinical interventions when their conditions become very severe.

However, a recent study shows a strong correlation between suicidal attempt and SNMDs [37], which indicates that adolescents suffering from social network addictions have a much higher risk of suicidal inclination than non-addictive users. The research also reveals that social network addiction may negatively impact emotional status, causing higher hostility, depressive mood, and compulsive behavior. Even more alarming is that the delay of early intervention may seriously damage individuals' social functioning. In short, it is desirable to have the ability to actively detect potential SNMD users on OSNs at an early stage.

Although previous work in Psychology has identified several crucial mental factors related to SNMDs, they are mostly examined as standard diagnostic criteria in survey questionnaires. To automatically detect potential SNMD cases of OSN users, extracting these factors to assess users' online mental states is very challenging. For example, the extent of loneliness and the effect of disinhibition of OSN users are not easily observable.[3] Therefore, there is a need to develop new approaches for detecting SNMD cases of OSN users. We argue that mining the social network data of individuals as a complementary alternative to the conventional psychological approaches provides an excellent opportunity to *actively identify* those cases at an early stage. In this paper, we develop a machine learning framework for detecting SNMDs, which we call *Social Network Mental Disorder Detection (SNMDD)*.

Specifically, we formulate the task as a semi-supervised classification problem to detect three types of SNMDs [53]: i) Cyber-Relationship Addiction, which shows addictive behavior for building online relationships; ii) Net Compulsion, which shows compulsive behavior for online social gaming or gambling; and iii) Information Overload, which is related to uncontrollable surfing. By exploiting machine learning techniques with the ground truth obtained via the current diagnostic practice in Psychology [53], we extract and analyze the following crucial categories of features from OSNs: 1) social comparison, 2) social structure, 3) social diversity, 4) parasocial relationships, 5) online and offline interaction ratio, 6) social capital, 7) disinhibition, 8) self-disclosure, and 9) bursting temporal behavior. These features capture important factors or serve as proxies for SNMD detection. For example, studies manifest that users exposed to positive posts from others on Facebook with similar background are inclined to feel malicious envy and depressed due to the social comparison [4]. The depression leads users to disorder behaviors, such as information overload or net compulsion. Therefore, we first identify positive newsfeeds and then calculate the profile similarity and relation familiarity between friends. As another example, a parasocial relationship is an asymmetric interpersonal relationship, i.e., one party cares more about the other, but the other does not. This asymmetric relationship is related to loneliness, one of the primary mental factors pushing users with SNMDs to excessively access online social media [5]. Therefore, we extract the ratio of the number of actions to and from friends of a user as a feature. In this paper, the extracted features are carefully examined through a user study.

3. The online disinhibition effect is a loosening (or complete abandonment) of social restrictions and inhibitions that would otherwise be present in normal face-to-face interaction during interactions with others on the Internet.

Furthermore, users may behave differently on different OSNs, resulting in inaccurate SNMD detection. When the data from different OSNs of a user are available, the accuracy of the SNMDD is expected to improve by effectively integrating information from multiple sources for model training. A naïve solution that concatenates the features from different networks may suffer from the curse of dimensionality. Accordingly, we propose an *SNMD-based Tensor Model (STM)* to deal with this multi-source learning problem in SNMDD. Advantages of our approach are: i) the novel *STM* incorporates the SNMD characteristics into the tensor model according to Tucker decomposition; and ii) the tensor factorization captures the structure, latent factors, and correlation of features to derive a full portrait of user behavior. We further exploit CANDECOMP/PARAFAC (CP) decomposition based *STM* and design a stochastic gradient descent algorithm, i.e., *STM-CP-SGD*, to address the efficiency and solution uniqueness issues in traditional Tucker decomposition. The convergence rate is significantly improved by the proposed second-order stochastic gradient descent algorithm, namely, *STM-CP-2SGD*. To further reduce the computation time, we design an approximation scheme of the second-order derivative, i.e., Hessian matrix, and provide a theoretical analysis.

The contributions of this paper are summarized below.

- Today online SNMDs are usually treated at a late stage. To actively identify potential SNMD cases, we propose an innovative approach, new to the current practice of SNMD detection, by mining data logs of OSN users as an early detection system.
- We develop a machine learning framework to detect SNMDs, called *Social Network Mental Disorder Detection*. We also design and analyze many important features for identifying SNMDs from OSNs, such as disinhibition, parasociality, self-disclosure, etc. The proposed framework can be deployed to provide an early alert for potential patients.
- We study the *multi-source learning* problem for SNMD detection. We significantly improve the efficiency and achieve the solution uniqueness by CP decomposition, and we provide theoretical results on nondivergence. By incorporating SNMD characteristics into the tensor model, we propose *STM* to better extract the latent factors from different sources to improve the accuracy.
- We conduct a user study with 3,126 users to evaluate the effectiveness of the proposed SNMDD framework. To the best of our knowledge, this is the first dataset crawled online for SNMD detection. Also, we apply SNMDD on large-scale real datasets, and the results reveal interesting insights on network structures in SNMD types, which can be of interest to social scientists and psychologists.

The rest of this paper is organized as follows. Section 2 surveys the related work. Section 3 presents *SNMDD*, focusing on feature extraction. Section 4 presents the proposed *STM* for multi-source learning and the acceleration of tensor decomposition with the theoretical results. Section 5 reports a user study, various analyses, and the experimental results. Finally, Section 6 concludes this paper.

2 Related Work

Internet Addiction Disorder (IAD) is a type of behavior addiction with the patients addicted to the Internet, just like those

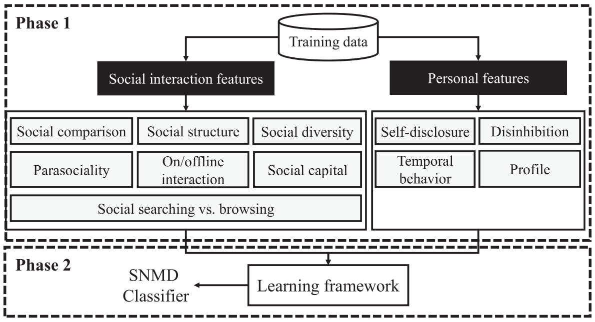


Fig. 1. The SNMDD framework.

addicting to drugs or alcohol [52]. Many research works in Psychology and Psychiatry have studied the important factors, possible consequences, and correlations of IAD [7], [29], [30], [36]. King et al. [30] investigate the problem of simulated gambling via digital and social media to analyze the correlation of different factors, e.g., grade, ethnicity. Baumer et al. [7] report the Internet user behavior to investigate the reason of addiction. Li et al. [36] examine the risk factors related to Internet addiction. Kim et al. [29] investigate the association of sleep quality and suicide attempt of Internet addicts. On the other hand, recent research in Psychology and Sociology reports a number of mental factors related to social network mental disorders. Research indicates that young people with narcissistic tendencies and shyness are particularly vulnerable to addiction with OSNs [6], [14]. However, the above research explores various negative impacts and discusses potential reasons for Internet addiction. By contrast, this paper proposes to automatically identify SNMD patients at the early stage according to their OSN data with a novel tensor model that efficiently integrate heterogeneous data from different OSNs.

Research on mental disorders in online social networks receives increasing attention recently [16], [18], [44]. Among them, content-based textual features are extracted from user-generated information (such as blog, social media) for sentiment analysis and topic detection. Chang et al. [16] employ an NLP-based approach to collect and extract linguistic and content-based features from online social media to identify Borderline Personality Disorder and Bipolar Disorder patients. Saha et al. [44] extract the topical and linguistic features from online social media for depression patients to analyze their patterns. Choudhury et al. [18] analyze emotion and linguistic styles of social media data for Major Depressive Disorder (MDD). However, most previous research focuses on individual behaviors and their generated textual contents but do not carefully examine the structure of social networks and potential Psychological features. Moreover, the developed schemes are not designed to handle the sparse data from multiple OSNs. In contrast, we propose a new multi-source machine learning approach, i.e., STM, to extract proxy features in Psychology for different diseases that require careful examination of the OSN topologies, such as Cyber-Relationship Addiction and Net Compulsion.

Our framework is built upon support vector machine, which has been widely used to analyze OSNs in many areas [28], [47]. In addition, we present a new tensor model that not only incorporates the domain knowledge but also well estimates the missing data and avoids noise to properly handle multi-source data. Caballero et al. [11] estimate the probability of mortality in ICU by modeling the probability of mortality as a latent state evolving over time. Zhao et al. [54] propose a hierarchical learning method for event detection and forecasting by first extracting the features from different data sources and then learning via geographical multi-level model. However, the SNMD data from different OSNs may be incomplete due to the heterogeneity. For example, the profiles of users may be empty due to the privacy issue, different functions on different OSNs (e.g., game, check-in, event), etc. We propose a novel tensor-based approach to address the issues of using heterogeneous data and incorporate domain knowledge in SNMD detection.

3 Social Network Mental Disorder Detection

In this paper, we aim to explore data mining techniques to detect three types of SNMDs [53]: 1) *Cyber-Relationship (CR) Addiction*, which includes the addiction to social networking, checking and messaging to the point where social relationships to virtual and online friends become more important than real-life ones with friends and families; 2) *Net Compulsion (NC)*, which includes compulsive online social gaming or gambling, often resulting in financial and job-related problems; and 3) *Information Overload (IO)*, which includes addictive surfing of user status and news feeds, leading to lower work productivity and fewer social interactions with families and friends offline.

Accordingly, we formulate the detection of SNMD cases as a classification problem. We detect each type of SNMDs with a binary SVM. In this study, we propose a two-phase framework, called *Social Network Mental Disorder Detection*, as shown in Fig. 1. The first phase extracts various discriminative features of users, while the second phase presents a new SNMD-based tensor model to derive latent factors for training and use of classifiers built upon Transductive SVM (TSVM) [19]. Two key challenges exist in design of SNMDD: i) we are not able to directly extract mental factors like what have been done via questionnaires in Psychology and thus need new features for learning the classification models;[4] ii) we aim to exploit user data logs from multiple OSNs and thus need new techniques for integrating multi-source data based on SNMD characteristics. We address these two challenges in Sections 3.1 and 4, respectively.

3.1 Feature Extraction

We first focus on extracting discriminative and informative features for design of SNMDD. This task is nontrivial for the following three reasons.

1. Lack of Mental Features. Psychological studies have shown that many mental factors are related to SNMDs, e.g., low self-esteem [52], loneliness [34]. Thus, questionnaires are designed to reveal those factors for SNMD detection. Some parts of Psychology questionnaire for SNMDs are based on the subjective comparison of mental states in online and offline status, which cannot be observed from OSN logs. For example:

Q1. How often do you feel depressed, moody, or nervous when you are off-line, which goes away once you are back online?

Q2. How often do you prefer the excitement of the Internet to intimacy with your partner?

Consider Q1. The feel of depression and nervousness off-line can not be observed online. To tackle this problem, we have to leverage the knowledge from Psychology, such as

4. Additional issues in feature extraction will be detailed later.

withdrawal or relapse patterns, and exploit some proxy features extracted from online social activity logs to approximate them. For Q2, the preference of excitement of the Internet to intimacy with users' partners is an important question for SNMD detection. As it is difficult to directly observe these factors from data collected from OSNs, psychiatrists are not able to directly assess the mental states of OSN users under the context of online SNMD detection.

2. Heavy Users versus Addictive Users. To detect SNMDs, an intuitive idea is to simply extract the usage (time) of a user as a feature for training SNMDD. However, this feature is not sufficient because i) the status of a user may be shown as "online" if she does not log out or close the social network applications on mobile phones, and ii) heavy users and addictive users all stay online for a long period, but heavy users do not show symptoms of anxiety or depression when they are not using social apps. How to distinguish them by extracting discriminative features is critical.

3. Multi-Source Learning with the SNMD Characteristics. As we intend to exploit user data from different OSNs in SNMDD, how to extract complementary features to draw a full portrait of users while considering the SNMD characteristics into the tensor model is a challenging problem.

To address the first two challenges, we identify a number of effective features as proxies to capture the mental states of users, e.g., self-esteem [52] and loneliness [34].[5] The goal is to distinguish users with SNMDs from normal users. Two types of features are extracted to capture the social interaction behavior and personal profile of a user. Due to the space constraint, some of the above features are presented in Appendix A, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TKDE.2017.2786695. It is worth noting that each individual feature cannot precisely classify all cases, as research shows that exceptions may occur. Therefore, it is necessary to exploit multiple features to effectively remove exceptions.

3.1.1 Social Interaction Features

We first extract a number of *social interaction features* to capture the user behavior on social media.

Social Comparison Based Features (SComp). Although most literature indicates that the majority of the newsfeed updates is positive, recent studies manifest that users who are exposed to positive posts from others on Facebook are inclined to feel envy and depressed due to social comparison [50]. The social comparison leads to SNMDs according to Festinger's theory, which states that many people usually have a strong motivation to evaluate their own opinions and abilities by implicitly or explicitly comparing with others in similar backgrounds, especially when the reference in comparison to the physical world is not specific. The situation becomes increasingly serious because status exchanges among friends are now very convenient via various online social networks.

Envy usually appears after comparisons, and two kinds of envy, i.e., benign envy and malicious envy, exist in Psychology [4]. The experience of benign envy leads to a moving-up motivation aiming at improving one's own position, whereas the experience of malicious envy produces a pulling-down motivation and depression. Malicious envy is incurred from the comparison among close friends with similar backgrounds and states, and it usually leads to SNMDs, such as information overload or net compulsion, because a person in this case usually feels pressure and tends to frequently check the updated status of the corresponding friends. A teenager student in this case may seek online games or gambles as alternatives for acquiring the sense of accomplishment. By contrast, benign envy is usually generated from distant friends with different backgrounds and rarely leads to SNMDs.

Therefore, for malicious envy, we first exploit the existing techniques of emotional signal processing [38] to identify positive newsfeeds and then calculate the profile similarity and relation familiarity between friends. Specifically, let $N_p(i,j)$ and $s(i,j)$ denote the number of positive newsfeeds that user j receives from i and the similarity on backgrounds between user i and j, respectively. For user j, the weighted number of positive newsfeeds based on similarity can be derived as

$$\frac{\sum_{i\in N(j)}[s(i,j)N_p(i,j)]}{\sum_{i\in N(j)} s(i,j)}, \tag{1}$$

where $N(j)$ is the set of neighbors of user j. Moreover, the weighted numbers of positive newsfeeds based on familiarity can also be derived in a similar manner by substituting the similarity function with the familiarity function as an additional proxy feature for the social comparison.

Social Structure Based Features (SS). In Sociology, each person in a social network belongs to one of the following three types of social roles: influential users, structural holes, and normal users. An influential user is the one with a huge degree and many mentions and shares (retweets) [13]. On the other hand, weaker connecting paths between groups are structure holes in OSNs, and researchers have demonstrated that structural holes usually have timely access to important information, e.g., trade trend, job opportunities, which usually leads to social success. Therefore, the users with their roles as structural holes are more inclined to suffer from information overload for newsfeeds because they enjoy finding and sharing new and interesting information to various friends.

According to the above observations, we exploit the state-of-the-art approach [27] to quantify users' tendencies of being structural holes. Specifically, given n users and m communities, let $\mathbf{F}\in\mathbb{R}^{n\times m}$ denote the community indicator matrix, where $\mathbf{f}_{ij}=1$ if a user i is assigned to the jth community, and 0 otherwise. Let $\mathbf{f}^i$ denote the ith row vector of $\mathbf{F}$. By embedding the harmonic function to learn the community indicator matrix, the difference between the value of $\mathbf{f}^i$ and the averaged value of its neighbors $\frac{1}{d_i}\sum_{(v_i,v_j)\in E}\mathbf{f}^j$ is required to be minimized, because neighbors are usually within similar communities. Therefore, the following minimization problem is formulated to detect structural hole spanners

$$\min_{\mathbf{F}} \|\mathbf{F}-\mathbf{D}^{-1}\mathbf{A}\mathbf{F}\|_{2,1} \tag{2}$$

$$\text{s.t. } \mathbf{F}^T\mathbf{F}=\mathbf{I}_m, \tag{3}$$

where $\mathbf{A}$ and $\mathbf{D}$ are respectively the adjacency and degree matrices, and $\|X\|_{2,1}$ is the $\ell_{2,1}$ norm of X, which is the sum of the euclidean norms of the columns of the matrix X. The structural hole spanners correspond to the ones with small $\|\mathbf{f}^i\|_2$. Compared with social capital based features, the structural hole feature considers the community structure (global), while social capital features only examine the ego networks (local). On the other hand, we also extract the network topology based features, i.e., closeness centrality, betweenness centrality, eigenvector centrality, information centrality, flow betweenness, the rush index, as social structure based features

5. The third challenge is addressed in Section 4.

for detecting SNMDs. For example, flow betweenness indicates how much information has been propagated through the node, which relates to information overload. Moreover, eigenvector centrality is a measure of the influence of a node in a network, and the score is similar to the pagerank, i.e., connections to high-scoring neighbors are inclined to increase the score of a node. Therefore, the scores of unpopular users are usually small and correlated to Cyber-Relationship Addiction.

Social Diversity Based Features (SDiv). Researchers have observed that diversity improves the depth of people thinking for both majority or minority [35]. For example, a person with a more diverse background and many friends is less inclined to suffer from SNMDs because she is often supported by friends and thereby rarely feels lonely and isolated (two important factors correlated to SNMDs) [24]. Therefore, the impact of social network diversity is increasingly important and inspires us to incorporate them for effective SNMD detection. Specifically, the diversities of nationality, racial, ethical, religious, and education can be extracted as social diversity based features with Shannon index H as the diversity index, i.e.,

$$H = -\sum_{i=1}^{N_t} p_i \ln p_i, \tag{4}$$

where p_i and N_t are the proportion of users' friends belonging to the ith type of attributes and the total number of types, respectively. The value H increases when the number of types N_t grows. Moreover, Shannon diversity index also increases when there is a more significant *evenness*. In other words, the diversity index is maximized when all type of attributes are of the equal quantities.

Parasocial Relationship (PR). Research shows that the mental factor of loneliness is one of the primary reasons why the users with SNMDs excessively access online social media [5]. As the loneliness of an OSN user is hard to measure, we exploit the parasocial relationship, an asymmetric interpersonal relationship between two people where one party cares more about the other but the other does not, to capture loneliness (as studies show that they are correlated [12]). The feature of parasocial relationship is represented as $|a_{out}|/|a_{in}|$, where $|a_{out}|$ and $|a_{in}|$ denote the number of actions a user takes to friends and the number of actions friends take to the user, respectively.[6] As the ratio increases, the extent of parasocial relationship also grows.

Due to the space constraint, other social interaction features are presented in Appendix A.1, available in the online supplemental material.

3.1.2 Personal Features

Temporal Behavior Features (TEMP). Relapse is the state that a person is inclined to quickly revert back to the excessive usage of social media after an abstinence period, while *tolerance* is the state that the time spent by a person with SNMDs tends to increase due to the mood modification effect.[7] It is worth noting that the above two mental states have been exploited to evaluate clinical addictions [34]. We aim to use them to distinguish *heavy users* and *addictive users* because heavy users do not suffer from relapse and tolerance in use of OSNs. An issue arising here is how to assess relapse and tolerance quantitatively.

6. The actions include like, comment, and post in our work.
7. A patient may need to spend more time on social media to reach the happiness/excitement than before.

It is observed that the use of social media by an SNMD patient is usually in the form of *intermittent bursts* [52]. Therefore, given a stream of a user's activities on an OSN, e.g., "likes", "comments", "posts", we exploit Kleinberg's burst detection algorithm [31], which is based on an infinite Markov model, to detect periods of the user's activities as bursty and non-bursty periods. The bursty period refers to a period during which the activities significantly increase. A bursty period is modeled as a bursty state q_1 in the Markov model, while a non-bursty period is correspondingly modeled as a normal state q_0. The burst detection algorithm finds a state transition sequence $\mathbf{q}$ for each user to divide the corresponding log (stream of activities) into bursty and non-bursty periods. Specifically, let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote a sequence of n time intervals between $n+1$ consecutive activities, with the intervals distributed according to a density function, such as $f_{i_t}(x_t) = \alpha_{i_t} e^{-\alpha_{i_t} x_t}$, where α_{i_t} is either α_0 or α_1, and α_0 and α_1 are parameters that correspond to the normal and burst states, respectively, $\alpha_1 > \alpha_0$. A time interval x_t is in a burst state q_1 if $f_0(x_t) < f_1(x_t)$. Otherwise, it is in a normal state q_0. However, simply deciding the state sequence $\mathbf{q}$ based on this criteria results in numerous small periods. Therefore, a cost $\tau(q_i, q_j)$ is associated with a state transition from q_i to q_j to filter out noises and to ensure that each bursty or non-bursty period is sufficiently long. Therefore, the remaining issue is to find an optimal state-transition sequence $\mathbf{q}$ to minimize the following cost function [31]

$$c(\mathbf{q}|\mathbf{x}) = \sum_{t=1}^{n-1} \tau(q_i, q_{i+1}) + \sum_{t=1}^{n} (-\ln f_{i_t}(x_t)),$$

where $\tau(q_i, q_{i+1}) = 0$ if the state q_i and q_{i+1} is the same. $\tau(q_i, q_{i+1})$ is $\gamma \ln n$ otherwise, where γ is an algorithm parameter larger than 0. Notice that the state sequence that minimizes the cost depends on 1) how easy it is to jump from one state to another and 2) how well it is to comply to the rates of arrivals. After identifying the bursts, we measure their intensity (the number of activities within a burst) and length (the time period of a burst) as the proxy features for *relapse* and *tolerance*, respectively. The $\langle$average, median, standard deviation, maximum, minimum$\rangle$ of both the burst intensity and burst length are included in our feature set, because they capture the characteristic of bursts. For instance, the standard deviation of the burst length for SNMD patients is usually larger than that for heavy users since heavy users constantly use OSNs, whereas the users with SNMDs increase the usage time due to tolerance.

Due to the space constraint, other personal features are presented in Appendix A.2, available in the online supplemental material.

4 Multi-Source Learning with Tensor Decomposition Acceleration

Many users are inclined to use different OSNs, and it is expected that data logs of these OSNs could provide enriched and complementary information about the user behavior. Thus, we aim to explore multiple data sources (i.e., OSNs) in SNMDD, in order to derive a more complete portrait of users' behavior and effectively deal with the data sparsity problem. To exploit multi-source learning in SNMDD, one simple way is to directly concatenate the features of each person derived from different OSNs as a huge vector. However, the above approach tends to miss the correlation of a feature in different OSNs and introduce interference. Thus, we explore tensor techniques which have been used increasingly to model multiple data sources because a tensor can naturally represent multi-source data. We aim to employ tensor decomposition to extract common latent factors from different sources and objects. Based on tensor decomposition on $\mathcal{T}$, we present a *SNMD-based Tensor Model* in previous work [45], which enables $\mathbf{U}$ to incorporate important characteristics of SNMDs, such as the correlation of the same SNMD sharing among close friends.[8] Finally, equipped with the new tensor model, we conduct semi-supervised learning to classify each user by exploiting Transductive Support Vector Machines (TSVM) in Appendix B, available in the online supplemental material. In the following, the problem definition, notation explanation, and brief introduction are first presented for better reading.

4.1 Problem Definition and Notation Explanation

Given D SNMD features of N users extracted from M OSN sources, we construct a three-mode tensor $\mathcal{T} \in \mathbb{R}^{N \times D \times M}$, where each element $t_{ijk} \in \mathcal{T}$ represents the jth feature of user i in source k. The objective here is to extract the latent features for each user with tensor composition from $\mathcal{T}$. Here scalars are denoted by lowercase letters, e.g., u, while vectors are denoted by boldface lowercase letters, e.g., $\mathbf{u}$. Matrices are represented by boldface capital letters, e.g., $\mathbf{U}$, and tensors are denoted by calligraphic letters, e.g., $\mathcal{T}$. The ith row and the jth column of a two-dimensional matrix $\mathbf{U}$ are respectively denoted by $\mathbf{u}_{i:}$ and $\mathbf{u}_{:j}$.

Tucker decomposition and CANDECOMP/PARAFAC decomposition have been widely used for extracting the latent features. In the following, we first briefly introduce Tucker decomposition.

4.2 Tucker Decomposition

Tucker decomposition [32] of a tensor $\mathcal{T} \in \mathbb{R}^{N \times D \times M}$ is defined as follows:

$$\mathcal{T} = \mathcal{C} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}, \tag{5}$$

where $\mathbf{U} \in \mathbb{R}^{N \times R}$, $\mathbf{V} \in \mathbb{R}^{D \times S}$ and $\mathbf{W} \in \mathbb{R}^{M \times T}$ are latent matrices. R, S, and T are parameters to be set according to different criteria [32]. The 1 -mode product of $\mathcal{C} \in \mathbb{R}^{R \times S \times T}$ and $\mathbf{U} \in \mathbb{R}^{N \times R}$, denoted by $\mathcal{C} \times_1 \mathbf{U}$, is a matrix with size $N \times S \times T$, where each element $(\mathcal{C} \times_1 \mathbf{U})_{nst} = \sum_{r=1}^{R} c_{rst} u_{rn}$. Given the input tensor matrix $\mathcal{T}$ that consists of the features of all users from every OSN, Tucker decomposition derives $\mathcal{C}$, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ to meet the above equality on $\mathcal{T}_{ndm}$ for every n, d, and m, where $\mathcal{C}$ needs to be diagonal, and $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ are required to be orthogonal [32]. Matrix $\mathbf{U}$ effectively estimates a deficit feature (e.g., a missing feature value unavailable due to privacy setting) of an OSN from the corresponding feature of other OSNs, together with the features of other users with the similar behavior.

4.3 CP Decomposition

Although Tucker decomposition is flexible and general, it is difficult to interpret the latent features intuitively from the decomposed matrices due to complicated interactions among them. Also, ensuring identifiability is fundamental and important for tensor decomposition. Moreover, the model parameters are encouraged to be uniquely recovered given the observed statistics, i.e., the decomposition yields a unique solution. For Tucker decomposition, the identifiability needs to satisfy complicated criteria, e.g., the structured sparsity and symmetry constraints on the core tensor, and sparsity constraints on the inverse factors of the tensor decomposition [2]. In contrast, the latent features obtained by CANDECOMP/PARAFAC decomposition [25] are much easier to interpret due to the rank-1 component factorization of CP and its intrinsic axis property from parallel proportional profiles. Moreover, Kruskal criterion on the rank of tensors provides a sufficient condition of the identifiability. Most importantly, its computational complexity is much lower than Tucker decomposition, thereby allowing us to analyze SNMDs for large-scale OSNs.

Specifically, CANDECOMP/PARAFAC decomposition of a tensor $\mathcal{T} \in \mathbb{R}^{N \times D \times M}$ is defined as follows:

$$\sum_{r=1}^{R} \mathbf{U}_{:r} \circ \mathbf{V}_{:r} \circ \mathbf{W}_{:r} \approx \mathcal{T}, \tag{6}$$

where $\circ$ denotes the vector outer product, and R is a positive integer representing the dimensionality of $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$, i.e., $\mathbf{U}_{:r} \in \mathbb{R}^{N}$, $\mathbf{V}_{:r} \in \mathbb{R}^{D}$, and $\mathbf{W}_{:r} \in \mathbb{R}^{m}$, for $r = 1, \ldots, R$. The space of variables in CP decomposition is comprised of the elements of $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$. The inner product of third-order tensors $\mathcal{X}$ and $\mathcal{Y}$ is defined as $\sum_i \sum_j \sum_k X_{ijk} Y_{ijk}$. The objective function of CP decomposition is to find $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ such that the decomposition is close to $\mathcal{T}$ (i.e., the difference is minimized). Each element $\mathbf{U}_{:r} \circ \mathbf{V}_{:r} \circ \mathbf{W}_{:r}$ is a rank-one tensor, and $\mathbf{U}_{i:}$ represents the SNMD feature tensor of user i. Compared to Tucker decomposition, the core tensor $\mathcal{C}$ in CP decomposition has been simplified, and thus the number of parameters required to be estimated in Eq. (6) is much smaller. Moreover, the solution is unique in CP decomposition but not unique in Tucker decomposition. Equipped with CP decomposition, the objective function $\mathcal{L}(\mathcal{T}, \mathbf{U}, \mathbf{V}, \mathbf{W})$ is

$$\frac{1}{2}\left\| \mathcal{T} - \sum_{r=1}^{R} \mathbf{U}_{:r} \circ \mathbf{V}_{:r} \circ \mathbf{W}_{:r} \right\|^2 + \frac{\lambda_1}{2} tr(\mathbf{U}^\top \mathbf{L}_a \mathbf{U}) + \frac{\lambda_2}{2} \|\mathbf{U}\|^2, \tag{7}$$

where $tr(\cdot)$ denotes the matrix trace, the Frobenius norm of a tensor $\mathcal{T}$ is defined as $\|\mathcal{T}\| = \sqrt{< \mathcal{T}, \mathcal{T} >}$, and λ_1 and λ_2 are parameters controlling the contribution of each part during the above collaborative factorization. The Laplacian matrix $\mathbf{L}_a$ of the weighted adjacency matrix $\mathbf{A}$ is defined as $\mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is a diagonal matrix with the entries $d_{ii} = \sum_i a_{ij}$. $\mathcal{L}$ first minimizes the decomposition error, i.e., $\|\mathcal{T} - \sum_{r=1}^{R} \mathbf{U}_{:r} \circ \mathbf{V}_{:r} \circ \mathbf{W}_{:r}\|^2$, for $\mathcal{T}$. Moreover, the term that minimizes $\|\mathbf{U}\|^2$ is to derive a more concise latent feature matrix and avoid overfitting. The proposed *STM* is different from the conventional tensor models in the second term of Eq. (7), where important characteristics of SNMDs are incorporated. For example, the probability of finding CR cases around a CR patient is higher than that around a non-CR user due to the loneliness propagation [12]. That is, CR users usually feel lonely and are more likely to establish friendships in cyberspace with other users with similar behavior. Since the nearby nodes with a great quantity of interactions tend to be the same (either CR or non-CR), it is envisaged that the distance between $\mathbf{u}_{i:}$ and $\mathbf{u}_{j:}$ will be small if the edge weight of the edge connecting user i and user j, i.e., $a_{i,j}$ in the adjacency matrix $\mathbf{A}$, is sufficiently large. Therefore, a regularization (smoothing) term $\frac{1}{2} tr(\mathbf{U}^T \mathbf{L}_a \mathbf{U})$ is included in the model to achieve the above goal. Due to the space constraint, the details of deriving $\frac{1}{2} tr(\mathbf{U}^T \mathbf{L}_a \mathbf{U})$ are presented in Appendix C, available in the online supplemental material.

8. Note that D does not capture the social correlations among friends.

4.4 Stochastic Gradient-Descent Algorithm

Notice that CP decomposition is non-convex. For traditional gradient descent algorithms [20], the learning step size η and the initial values on $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ are very sensitive and need to be carefully determined. Otherwise, the algorithm is inclined to diverge, consequently failing to find the decomposition solution. To address this issue, we design a new stochastic gradient-descent algorithm with low computational complexity to guarantee the solution convergence.

We present a stochastic gradient-descent algorithm for CP decomposition of the SNMD-based Tensor Model, namely, *SGD-CP-STM*, to iteratively improve each element in the matrices according to the corresponding gradient. Specifically, let $\mathcal{T}(\cdot,\mathbf{V},\mathbf{W})$ be a matrix obtained from $\mathcal{T}$ by contracting $\mathbf{V}$ and $\mathbf{W}$, i.e.,

$$\mathcal{T}(\cdot,\mathbf{V},\mathbf{W})_{ir} = \sum_j \sum_k \mathcal{T}_{ijk}\mathbf{V}_{jr}\mathbf{W}_{kr}, \tag{8}$$

where $\mathcal{T}(\cdot,\mathbf{V},\mathbf{W}) \in \mathbb{R}^{N\times R}$ (the same as $\mathbf{U}$). The following lemma first derives the gradient of each iteration.

Lemma 1. *The gradient of $\mathcal{L}$ with regard to $\mathbf{U}$, i.e., $\nabla_{\mathbf{U}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W})$, is equal to*

$$-\mathcal{T}(\cdot,\mathbf{V},\mathbf{W}) + \mathbf{U}(\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R) + \lambda_1\mathbf{L}_a\mathbf{U},$$

where $\Gamma(\mathbf{V},\mathbf{W})$ is the Hadamard product of $\mathbf{V}^\top\mathbf{V}$ and $\mathbf{W}^\top\mathbf{W}$, i.e., $\Gamma(\mathbf{V},\mathbf{W})_{ij} = (\mathbf{V}^\top\mathbf{V})_{ij}(\mathbf{W}^\top\mathbf{W})_{ij}$, and $\mathbf{I}_R$ is the identity matrix of size R.

Proof. The objective function $\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W})$ is comprised of three terms, and the derivative of $\frac{\lambda_2}{2}\|\mathbf{U}\|^2$ with regard to $\mathbf{U}$ is $\lambda_2\mathbf{I}_R$. For the first term, the CP gradient can be solved by the following equation according to [1]

$$\begin{aligned}\nabla_{\mathbf{U}}\frac{1}{2}&\left\|\mathcal{T} - \sum_{r=1}^{R}\mathbf{U}_{:r}\circ\mathbf{V}_{:r}\circ\mathbf{W}_{:r}\right\|^2 \\ &= -\mathcal{T}(\cdot,\mathbf{V},\mathbf{W}) + \mathbf{U}\Gamma(\mathbf{V},\mathbf{W}).\end{aligned} \tag{9}$$

For the second term, i.e., $\frac{\lambda_1}{2}tr(\mathbf{U}^\top\mathbf{L}_a\mathbf{U})$, the gradient for $\mathbf{U}$ is

$$\nabla_{\mathbf{U}}\frac{\lambda_1}{2}tr(\mathbf{U}^\top\mathbf{L}_a\mathbf{U}) = \frac{\lambda_1}{2}(\mathbf{L}_a + \mathbf{L}_a^\top)\mathbf{U}. \tag{10}$$

If the weighted adjacency matrix $\mathbf{A}$ is symmetric, Eq. (10) can be further simplified to $\lambda_1\mathbf{L}_a\mathbf{U}$, and $\nabla_{\mathbf{U}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W})$ is equal to

$$-\mathcal{T}(\cdot,\mathbf{V},\mathbf{W}) + \mathbf{U}(\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R) + \lambda_1\mathbf{L}_a\mathbf{U}. \tag{11}$$

The theorem follows. □

Therefore, the stochastic gradient descent algorithm updates $\mathbf{U}$ at the tth iteration as follows:

$$\begin{aligned}\mathbf{U}^{(t)} = \mathbf{U}^{(t-1)} - \eta^{(t)}(&-\mathcal{T}^{(t-1)}(\cdot,\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)}) \\ &+ \mathbf{U}^{(t-1)}(\Gamma(\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)}) + \lambda_2\mathbf{I}_R) + \lambda_1\mathbf{L}_a\mathbf{U}^{(t-1)}).\end{aligned} \tag{12}$$

Based on Eq. (9), the gradient for $\mathbf{V}$ and $\mathbf{W}$ can be derived in the similar way as follows:

$$\begin{aligned}\nabla_{\mathbf{V}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W}) &= -\mathcal{T}(\mathbf{U},\cdot,\mathbf{W}) + \mathbf{V}\Gamma(\mathbf{U},\mathbf{W}) \\ \nabla_{\mathbf{W}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W}) &= -\mathcal{T}(\mathbf{U},\mathbf{V},\cdot) + \mathbf{W}\Gamma(\mathbf{U},\mathbf{V}).\end{aligned}$$

Note that $\mathbf{V}^{(t)}$ and $\mathbf{W}^{(t)}$ are also updated similarly in each iteration.

4.5 Acceleration of Convergence Rate

It has been widely recognized that the performance is poor when a constant learning step size η is adopted. If the learning step size is too large, it is inclined to skip the optimal solution. In contrast, if the learning step size is too small, the convergence rate to the optimal solution becomes unacceptably slow. To avoid the above issue, we further propose a second-order stochastic gradient descent algorithm for CP decomposition, namely, *STM-CP-2SGD*. The second-order stochastic gradient descent (2SGD) considers the second-order information by adaptively assigning the learning rate as the inverse of the Hessian matrix in objective function $\mathcal{L}$ to guide the searching direction. Since the inversion of the full Hessian matrix

$$\begin{bmatrix}\nabla_{\mathbf{U}}\nabla_{\mathbf{U}}\mathcal{L} & \nabla_{\mathbf{U}}\nabla_{\mathbf{V}}\mathcal{L} & \nabla_{\mathbf{U}}\nabla_{\mathbf{W}}\mathcal{L} \\ \nabla_{\mathbf{V}}\nabla_{\mathbf{U}}\mathcal{L} & \nabla_{\mathbf{V}}\nabla_{\mathbf{V}}\mathcal{L} & \nabla_{\mathbf{V}}\nabla_{\mathbf{W}}\mathcal{L} \\ \nabla_{\mathbf{W}}\nabla_{\mathbf{U}}\mathcal{L} & \nabla_{\mathbf{W}}\nabla_{\mathbf{V}}\mathcal{L} & \nabla_{\mathbf{W}}\nabla_{\mathbf{W}}\mathcal{L}\end{bmatrix},$$

is computationally expensive, its block-diagonal parts are alternatively used as an approximation of the inverse of the Hessian matrix [26]. Therefore, we update $\mathbf{U}^{(t)}$ as follows:

$$\mathbf{U}^{(t)} = \mathbf{U}^{(t-1)} - \eta^{(t)}(\nabla^2_{\mathbf{U}^{(t-1)}}\mathcal{L})^{-1}\nabla_{\mathbf{U}^{(t-1)}}\mathcal{L}. \tag{13}$$

To calculate $\nabla^2_{\mathbf{U}}\mathcal{L}$, we take a derivative of Eq. (11) with respect to $\mathbf{U}$ and obtain

$$\nabla^2_{\mathbf{U}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W}) = \frac{d(\mathbf{U}(\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R))}{d\mathbf{U}} + \lambda_1\frac{d(\mathbf{L}_a\mathbf{U})}{d\mathbf{U}}. \tag{14}$$

We find the derivatives by exploiting the relationship between the Kronecker product and the vec operator (vectorizing matrices by stacking its columns) as follows. For the first term in Eq. (14), we have

$$\frac{d(\mathbf{U}(\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R))}{d\mathbf{U}} = \frac{d(\mathrm{vec}(\mathbf{U}(\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R))}{d\mathrm{vec}(\mathbf{U})}. \tag{15}$$

Since $\mathrm{vec}(\mathbf{U}(\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R)) = \mathrm{vec}(\mathbf{I}_N\mathbf{U}(\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R)) = (\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R)^T \otimes \mathbf{I}_N\mathrm{vec}(\mathbf{U})$, where $\otimes$ is Kronecker product, we have

$$\frac{d(\mathbf{U}(\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R))}{d\mathbf{U}} = (\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R)^T \otimes \mathbf{I}_N. \tag{16}$$

We find the second term in Eq. (14) in a similar manner

$$\frac{d(\mathbf{L}_a\mathbf{U})}{d\mathbf{U}} = \frac{d\mathrm{vec}(\mathbf{L}_a\mathbf{U})}{d\mathrm{vec}(\mathbf{U})}. \tag{17}$$

Since $\mathrm{vec}(\mathbf{L}_a\mathbf{U}) = \mathrm{vec}(\mathbf{L}_a\mathbf{U}\mathbf{I}_R) = \mathbf{I}_R \otimes \mathbf{L}_a\mathrm{vec}(\mathbf{U})$, we have

$$\frac{d(\mathbf{L}_a\mathbf{U})}{d\mathbf{U}} = \mathbf{I}_R \otimes \mathbf{L}_a. \tag{18}$$

Substituting Eqs. (16) and (18) into Eq. (14), $\nabla^2_{\mathbf{U}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W})$ is equal to

$$\begin{aligned}&(\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R)^T \otimes \mathbf{I}_N + \lambda_2\mathbf{I}_R \otimes \mathbf{L}_a \\ &= (\Gamma(\mathbf{V},\mathbf{W}) + \lambda_2\mathbf{I}_R)^T \oplus \lambda_2\mathbf{L}_a,\end{aligned} \tag{19}$$

where $\oplus$ is the Kronecker sum. Therefore, we update $\mathbf{U}$ at the tth iteration as follows, i.e., $\mathbf{U}^{(t)}$ is equal to

$$\mathbf{U}^{(t-1)} - \eta^{(t)}((\Gamma(\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)}) + \lambda_2\mathbf{I}_R)^T \oplus \lambda_2\mathbf{L}_a)^{-1}\nabla_{\mathbf{U}^{(t-1)}}\mathcal{L}. \tag{20}$$

Compared with Eq. (12), here the update of $\mathbf{U}$ includes a new term, i.e., $((\Gamma(\mathbf{V},\mathbf{W})+\lambda_2\mathbf{I}_R)^T\oplus\lambda_2\mathbf{L}_a)^{-1}$, representing the adaptive learning step size. Previous studies show that the number of iterations for 2SGD to reach the optimum is much smaller than that of SGD [10]. More specifically, given a second-order convex function, 2SGD requires only one iteration because it derives the optimal step size to the optimum point. To properly choose the $\eta^{(t)}$, several approaches can be applied, e.g., damped Newton method [41]. Here, we employ cross validation on the training dataset to find the initial value $\eta^{(0)}$ with Adagrad accordingly. Moreover, we update $\mathbf{V}^{(t)}$ and $\mathbf{W}^{(t)}$ as follows:

$$\mathbf{V}^{(t)}=\mathbf{V}^{(t-1)}-\eta^{(t)}(\Gamma(\mathbf{U}^{(t-1)},\mathbf{W}^{(t-1)})^T\otimes\mathbf{I})^{-1}\nabla_{\mathbf{V}^{(t-1)}}\mathcal{L},$$
$$\mathbf{W}^{(t)}=\mathbf{W}^{(t-1)}-\eta^{(t)}(\Gamma(\mathbf{U}^{(t-1)},\mathbf{V}^{(t-1)})^T\otimes\mathbf{I})^{-1}\nabla_{\mathbf{W}^{(t-1)}}\mathcal{L}.$$

2SGD, which exploits a Hessian matrix, outperforms SGD because the optimal step size is equal to the inverse of the Hessian matrix when the surface is approximated as a quadratic plane. *STM-CP-2SGD* further utilizes the block diagonal components to approximate the Hessian matrix for acceleration. Notice that if the Hessian matrix is not invertible, a nonnegative diagonal matrix with negligible-valued elements can be added to the original matrix to produce a positive definite matrix [23].

4.6 Theoretical Results

In the following, we first derive the computational complexity of the above two algorithms. Afterward, we prove that *STM-CP-2SGD* always converges to the solution of CP decomposition. More specifically, let $|\mathcal{T}|$ denote the number of nonzero elements in $\mathcal{T}$.

Lemma 2. *The computational complexity of* STM-CP-SGD *is* $O((N+D+M)R^2+N^2R+|\mathcal{T}^{(t)}|R)$.

title>Proof. For $\mathbf{V}$ and $\mathbf{W}$, the complexity of *STM-CP-SGD* is $O((D+M)R^2+|\mathcal{T}^{(t)}|R)$ for each update, where $O((D+M)R^2)$ is for computing $\mathbf{V}\Gamma(\mathbf{U},\mathbf{W})$ and $\mathbf{W}\Gamma(\mathbf{U},\mathbf{V})$, and $O(|\mathcal{T}^{(t)}|R)$ is to find $\mathcal{T}^{(t)}(\mathbf{U},\cdot,\mathbf{W})$ and $\mathcal{T}^{(t)}(\mathbf{U},\mathbf{V},\cdot)$. For $\mathbf{U}$, the complexity of *STM-CP-SGD* is $O(NR^2+|\mathcal{T}^{(t)}|R+N^2R)$ for each update since the time to find $\mathbf{L}_a\mathbf{U}$ is $O(N^2R)$. Therefore, the computational complexity of *STM-CP-SGD* is $O((N+D+M)R^2+N^2R+|\mathcal{T}^{(t)}|R)$. The lemma follows. □

Lemma 3. *The computational complexity of* STM-CP-2SGD *is* $O((N+D+M)R^2+N^2R+|\mathcal{T}^{(t)}|R)$.

Proof. Compared with *STM-CP-SGD*, *STM-CP-2SGD* needs to additionally derive the inverse of the Hessian matrix. To update $\mathbf{V}$, since $(\Gamma(\mathbf{U},\mathbf{W})^T\otimes\mathbf{I})^{-1}=(\Gamma(\mathbf{U},\mathbf{W})^T)^{-1}\otimes\mathbf{I}^{-1}$, we only have to additionally calculate the inverse of $\Gamma(\mathbf{U},\mathbf{W})$, and thus the computational complexity is $O(R^3)$, as well as for the update of $\mathbf{W}$. On the other hand, for the update of $\mathbf{U}$, efficiently calculating the inverse of the Kronecker sum, i.e., $\Gamma(\mathbf{V},\mathbf{W})^T\oplus\mathbf{L}_a\in\mathbb{R}^{RN\times RN}$, has been studied in solving Sylvester equation in control system, and it can be approximated in $O(N^2+R^2)$ [33]. Therefore, the computational complexity is still $O((N+D+M)R^2+N^2R+|\mathcal{T}^{(t)}|R)$. The lemma follows. □

The worst-case computational complexity of *STM-CP-2SGD* is the same as *STM-CP-SGD*. Nevertheless, the computation of *STM-CP-2SGD* is slightly more complicated than that of *STM-CP-SGD* because *STM-CP-2SGD* needs to compute the inverse of the Hessian matrix. Neverthelss, as shown in the experimental results, the convergence rate of *STM-CP-2SGD* is much faster than that of *STM-CP-SGD*. Given the same start point, *STM-CP-2SGD* can adaptively control the learning step size to approach the optimal solution and thus requires much fewer iterations than that of *STM-CP-SGD* with a constant learning step size.

Assume that the Frobenius norm of $\mathcal{T}$ is bounded by a constant, i.e., $\|\mathcal{T}\|\leq C$, we have the following theorem.

Theorem 1. *Given any initial solution or step size,* STM-CP-2SGD *does not diverge.*

Proof. *STM-CP-2SGD* updates $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ by finding the first and second derivatives. Consider the tth update of $\mathbf{U}^{(t)}$. We first fix the values of $\mathbf{V}^{(t-1)}$ and $\mathbf{W}^{(t-1)}$ and set Eq. (11) as zero to find $\mathbf{U}^*$, i.e.,

$$\mathbf{U}(\Gamma(\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})+\lambda_2\mathbf{I}_R)+\lambda_1\mathbf{L}_a\mathbf{U}=\mathcal{T}(\cdot,\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)}).\quad(21)$$

Since Eq. (21) is a Sylvester-type equation, the optimal solution of $\mathbf{U}$ when fixing $\mathbf{V}^{(t-1)}$ and $\mathbf{W}^{(t-1)}$, denoted as $\mathbf{U}^*$, can be derived as

$$\mathcal{T}(\cdot,\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})((\Gamma(\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})+\lambda_2\mathbf{I}_R)^T\oplus\lambda_1\mathbf{L}_a)^{-1},$$

which is the least-square solution by fixing $\mathbf{V}^{(t-1)}$ and $\mathbf{W}^{(t-1)}$. By rearranging Eq. (20), $\mathbf{U}^{(t)}$ can be derived as

$$\begin{aligned}&\mathbf{U}^{(t-1)}-\eta^{(t)}((\Gamma(\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})+\lambda_2\mathbf{I}_R)^T\oplus\lambda_2\mathbf{L}_a)^{-1}\\&(-\mathcal{T}(\cdot,\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})+\mathbf{U}^{(t-1)}(\Gamma(\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})+\lambda_2\mathbf{I}_R)\\&\quad+\lambda_1\mathbf{L}_a\mathbf{U}^{(t-1)})\\&\quad=\mathbf{U}^{(t-1)}+\eta^{(t)}\mathcal{T}(\cdot,\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})((\Gamma(\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})\\&\quad+\lambda_2\mathbf{I}_R)^T\oplus\lambda_2\mathbf{L}_a)^{-1}-\eta^{(t)}\mathbf{U}^{(t-1)}\\&\quad=(1-\eta^{(t)})\mathbf{U}^{(t-1)}+\eta^{(t)}\mathbf{U}^*.\end{aligned}$$

Therefore, $\mathbf{U}^{(t)}$ is a linear combination of $\mathbf{U}^{(t-1)}$ and $\mathbf{U}^*$. When $\mathbf{V}^{(t-1)}$ and $\mathbf{W}^{(t-1)}$ are fixed as constants, according to [39], $\mathbf{U}^*$ is a solution of the least-square problem, and thus

$$\mathcal{L}(\mathcal{T},\mathbf{U}^*,\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})\leq\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)}),\forall\mathbf{U}.$$

Let $\mathbf{U}$ in the right-hand side be 0, and we have $\mathcal{L}(\mathcal{T},\mathbf{U}^*,\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})\leq\frac{1}{2}\|\mathcal{T}\|^2$. Therefore, since the Frobenius norm of $\mathcal{T}$ is bounded, the initial solution of $\mathbf{U}^{(0)}$ is also bounded, and so is $\mathbf{U}^{(t)}$. The updates of $\mathbf{V}^{(t)}$ and $\mathbf{W}^{(t)}$ are similar. The theorem follows. □

Notice that this theorem does not guarantee the convergence[9] of the algorithm since the solutions may oscillate when updating $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ sequentially. Nevertheless, the above property is useful in practice since the solution can be always bounded and does not result in overflow [46].

5 Experimental Results

In this section, we evaluate SNMDD with real datasets. A user study with 3,126 people is conducted to evaluate the accuracy of SNMDD. Moreover, a feature study is performed. Finally, we apply SNMDD on large-scale datasets and analyze the detected SNMD types.

9. To ensure the convergence of 2SGD, more assumptions, e.g., $(\alpha,\gamma,\epsilon,\delta)$-strict saddle with stochastic gradient oracle with the radius at most Q, are required to prove the range to optimal solutions [22], [40], but the above condition may not always exist in practical situations.

TABLE 1
Details of the Datasets

Dataset	Description
FB_US	User profile, the friends of each user, the news feeds created by users with metadata (who likes, who comments, stickers, and geotag), the news feeds users like or comment (stickers also), events (join/decline), joined groups with events, and game posts created by game apps
IG_US	User profile, the followers/followees of each user, the media created by users with metadata (who likes, who comments, and geotag), and the contents users like or comment
FB_L	Anonymized user ID that performs the action, anonymized user ID that receives the action, and timestamp of action creation
IG_L	Anonymized media ID, anonymized ID of the user who created the media, timestamp of media creation, set of tags assigned to the media, number of likes and number of comments received

5.1 Data Preparation and Evaluation Plan

In the following, we detail the preparation of the datasets used in our evaluation.

5.1.1 User Study

We recruit 3,126 OSN users around the world via Amazon Mechanical Turk (MTurk) to obtain data for training and testing the classifiers in SNMDD. The participants include 1,790 males and 1,336 females. Their professions are very diverse, affiliating with universities, government offices, technology companies, art centers, banks, and businesses. Each user is first invited to fill out the standard SNMD questionnaires [3], [53].[10] Then, a group of professional psychiatrists participating in this project assess and manually label the users as *potential SNMD cases* (and their types of SNMDs) or *normal users*.[11] There are 389 users labeled as *SNMD*, including 246 Cyber-Relationship Addiction, 267 Information Overload, and 73 Net Compulsion.[12] The result obtained by the psychiatrists serves as the ground truth for our evaluation. We also crawl the Facebook (denoted as FB_US) and Instagram (denoted as IG_US) data of the participants in the user study for training and testing of our machine learning models (based on features detailed in Section 3.1). All the data are collected with the Facebook and Instagram APIs as listed in Table 1.

In the experiment, we first evaluate the effectiveness of the proposed features, including all features (`All`), social interaction features (`Social`), personal profile features (`Personal`), with a baseline feature `Duration`, i.e., the total time spent online, using TSVM [19] for semi-supervised learning in the user study. The combinations of different features, i.e., `Duration` with `Social` (`D-S`), `Duration` with `Personal` (`D-P`), `Social` with `Personal` (`S-P`), are also presented. We also collect two large-scale datasets, including Facebook (denoted as FB_L) with 63 K nodes, 1.5 M edges, and 0.84 M wall posts [49], and Instagram (denoted as IG_L) with 2 K users, 9 M tags, 1200 M likes, and 41M comments [21]. Note that some proposed features cannot be extracted from certain large-scale datasets, e.g., game posts and stickers are not available in IG_L, which is handled by using the imputation technique [43]. The details of the data crawled from each social media are listed in Table 1.

With labeled (IG_US and FB_US) and unlabeled data (IG_L and FB_L) described above, we perform a 5-fold cross validation, i.e., take 4 folds for training and 1 fold for testing, to evaluate the performance of proposed features using semi-supervised TSVM. A number of supervised approaches, including J48 Decision Tree Learning [51], SVM [15], Logistic Regression, and DTSVM [17] which do not use unlabeled data, are also compared to justify our choice of using TSVM in SNMDD. Next, we compare the proposed *SNMD-based Tensor Model*, implemented by different algorithms, i.e., *STM-Tucker-SGD*, *STM-CP-SGD*, and *STM-CP-2SGD*, with two baseline algorithms. The first baseline algorithm concatenates the features from different networks together (denoted as *CF*), while the second baseline algorithm employs the existing Tucker model (denoted as *Tucker*) that does not incorporate prior knowledge regarding the characteristics of SNMD cases (observed from our analysis). Finally, the effectiveness of each feature is carefully analyzed in Section 5.5.

5.1.2 Large-Scale Experiments

To discover new insights, we apply our semi-supervised SNMDD on IG_L and FB_L to classify their users and then analyze the detected cases of different SNMD types. Notice that the goal of this analysis is exploratory-oriented as we do not have the ground truth for the large datasets. We examine whether friends of SNMD cases tend to be potential SNMD cases as well. Also, we apply community detection on FB_L and IG_L to derive the relationships between different types of SNMD users in their communities. Finally, the average hop distance between the SNMD users of the same type is reported.

5.2 Evaluation of the Proposed Features

In the following, we first evaluate the performance of the proposed features using TSVM. We adopt Accuracy (Acc.) and Area Under Curve (AUC) for evaluation of SNMDD. Moreover, Microaveraged-F1 (Micro-F1) and Macroaveraged-F1 (Macro-F1) are also compared for multiple-label classification. Table 2 summarizes the average results and standard deviations, where the examined feature sets are denoted by self-explained labels.

The results on the IG_US and FB_US datasets in the user study show that `Duration` leads to the worst performance, i.e., the results of accuracy are 34 and 36 percent, and the AUC are 0.362 and 0.379, respectively. Notice that the AUC function can flip the results if the calculated AUC is less than 0.5, i.e., 1-AUC. Here, we do not filp the results to show that `Duration` is in fact a negative predictor in our case because `Duration` cannot differentiate heavy users with addictive users. Using all (`All`) or parts (`Social` or `Personal`) of the proposed features outperforms `Duration` significantly (see Table 2). `All` achieves the best performance (80 and 84 percent accuracy on the IG_US and FB_US datasets, respectively) because SNMDD is able to capture the various features extracted from data logs to effectively detect SNMD cases. The performance of `Personal` and `Social` are comparable, and the integrated feature set `All` outperforms `Personal` and `Social` by at least 15 and 16 percent on IG_US and FB_US in terms of accuracy. Since the F1 measure ignores true negatives, its magnitude is mostly determined by the number of true positives, i.e., large classes dominate small

10. The IRB number of this project is AS-IRB-HS 15003 v.1.

11. They are from California School of Professional Psychology, Taipei City Hospital, Nat'l Taipei Univ., psychiatric clinics, etc.

12. Note that a person may have multiple types of SNMDs simultaneously.

TABLE 2
Different Combinations of Feature Categories for Performance Evaluations on the IG_US and FB_US Datasets

Measure	`Duration`	`Social`	`Personal`	`D-S`	`D-P`	`S-P`	`All`
Instagram							
Acc.	0.34 ± 0.02	0.67 ± 0.01	0.69 ± 0.03	0.63 ± 0.02	0.68 ± 0.02	0.80 ± 0.01	0.80 ± 0.01
AUC	0.36 ± 0.02	0.71 ± 0.02	0.74 ± 0.01	0.69 ± 0.02	0.73 ± 0.02	0.81 ± 0.01	0.81 ± 0.01
Micro-F1	0.33 ± 0.01	0.76 ± 0.01	0.78 ± 0.04	0.74 ± 0.01	0.77 ± 0.03	0.85 ± 0.01	0.85 ± 0.01
Macro-F1	0.33 ± 0.01	0.71 ± 0.01	0.73 ± 0.02	0.69 ± 0.02	0.72 ± 0.02	0.85 ± 0.01	0.85 ± 0.01
p value on AUC	$3.80 \cdot 10^{-8}$	$6.05 \cdot 10^{-5}$	$1.18 \cdot 10^{-5}$	$1.64 \cdot 10^{-6}$	$3.06 \cdot 10^{-6}$	0.76	-
Facebook							
Acc.	0.36 ± 0.01	0.72 ± 0.03	0.73 ± 0.02	0.70 ± 0.03	0.73 ± 0.01	0.84 ± 0.02	0.84 ± 0.02
AUC	0.37 ± 0.01	0.75 ± 0.02	0.77 ± 0.02	0.74 ± 0.01	0.76 ± 0.02	0.86 ± 0.02	0.85 ± 0.01
Micro-F1	0.44 ± 0.04	0.80 ± 0.02	0.81 ± 0.01	0.79 ± 0.02	0.80 ± 0.01	0.90 ± 0.01	0.90 ± 0.01
Macro-F1	0.35 ± 0.02	0.76 ± 0.01	0.77 ± 0.03	0.74 ± 0.01	0.76 ± 0.02	0.91 ± 0.01	0.91 ± 0.01
p value on AUC	$7.01 \cdot 10^{-9}$	$2.35 \cdot 10^{-5}$	$3.06 \cdot 10^{-4}$	$3.90 \cdot 10^{-6}$	$1.46 \cdot 10^{-4}$	0.064	-

classes in microaveraging. As shown in Table 2, Micro-F1 of `Duration`, `Social`, and `Personal` are larger than Macro-F1 using both IG_US and FB_US datasets, indicating that using parts of features performs better on IO and CR (large classes) than NC. In contrast, the performance of SNMDD is almost the same in Micro-F1 and Macro-F1, which indicates its robustness. The results from FB_US are better than those from IG_US because IG_US is sparser, e.g., there are no event and game posts on Instagram. After comparing the results from SNMDD with the ground truth obtained via user study, we observe that some false-positive users are detected as NC, probably because people with NC are more likely to hide their real usage time, e.g., the game logs of some people with NC are hidden. As a result, a few normal users may be incorrectly detected as NC. However, SNMDD generally performs very well for NC due to some effective features. For example, users of NC are usually less parasocial since they are less frequent to interact with friends. Moreover, since the NC users' friends with game benefits usually do not know the NC users' other friends (e.g., colleagues), their clustering coefficients are lower than the normal users. Finally, the performance of `Social` and `Personal` with `Duration` features (`D-S` and `D-P`) are almost the same since SVM finds the best hyperplane to classify the training data and may not take the dimensions that downgrade the results. The results also manifests that the proposed features are robust with SVM. The p-value tests of different features with `All` indicate that `All` is significantly better than `Duration`, `Social`, `Personal`, `D-S`, `D-P` with p values that are much smaller than 0.05, while the performance is close to `S-P`.

5.3 Evaluation of Classification Techniques and STM

In the following, given all the proposed features, we first evaluate TSVM in comparison with some representative supervised learning approaches in SNMDD. As shown in Table 3, the accuracy of semi-supervised TSVM (84.3 percent) outperforms all the supervised algorithms, including ℓ_2-regularized logistic regression (78.6 percent) and ℓ_2-regularized ℓ_2-loss SVM (79.2 percent), since TSVM effectively uses unlabeled data to address the issues of overfitting and data sparsity. The accuracy and AUCs of the single-source supervised learning methods are similar, indicating that the proposed features provide robust information that is not sensitive to the choice of learning algorithms.

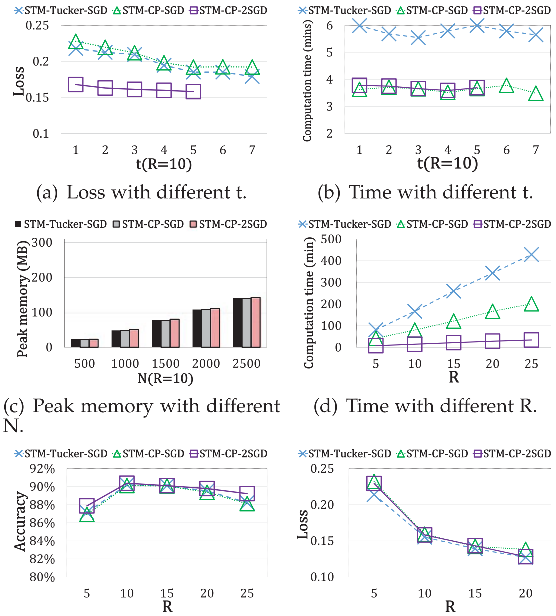
Next, we compare the proposed multi-source *STM-Tucker-SGD*, *STM-CP-SGD*, and *STM-CP-2SGD* with two baselines, i.e., *CF* and *Tucker*, to integrate the features extracted from IG_US and FB_US datasets with TSVM. Table 3 points out that the accuracy and AUC of *STM-CP-2SGD* are 90.4 percent and 0.938, respectively. *STM-CP-2SGD* with the decomposed latent factor matrix **U** can effectively recover important missing features and provide extra latent information to better characterize the users. In contrast, *CF*, which simply concatenates the features from FB_US and IG_US, suffers from the worst accuracy and AUC and is even beated by single-source ℓ_2-regularized ℓ_2-loss SVM. This is because *CF* loses correlations among the features and thereby tends to introduce noises. On the other hand, *STM-CP-2SGD* outperforms the other approaches because it incorporates important characteristics of SNMD and thereby derives more precise and accurate latent features, while the accuracy and AUC of *STM-Tucker-SGD* and *STM-CP-SGD* are almost the same.

5.4 Evaluation of the Proposed Tensor Decomposition Acceleration

In the following, the default dimensionality of **U**, **V** and **W**, threshold ϵ, and the maximum number of iterations are set as 10, 0.001, and 50, respectively. We first compare the loss function through each iteration in Fig. 2a. Note that *STM-CP-2SGD* converges very quickly (always terminates before the 5th iteration). Between *STM-Tucker-SGD* and *STM-CP-SGD*, the loss of *STM-Tucker-SGD* is slightly smaller since the core tensor of *STM-Tucker-SGD* allows more freedom to fit data. On the other hand, *STM-CP-2SGD* outperforms

TABLE 3
Comparisons of SNMDD with Different Classification Techniques

Technique	Acc.	AUC
Single-source (FB)		
J48 Decision Tree Learning	75.4%	0.763
ℓ_1-regularized ℓ_2-loss SVM	79.1%	0.790
ℓ_2-regularized ℓ_2-loss SVM	79.2%	0.791
ℓ_1-regularized logistic regression	78.5%	0.788
ℓ_2-regularized logistic regression	78.6%	0.789
DTSVM	78.5%	0.782
TSVM	**84.2%**	**0.851**
Multi-source (FB+IG)		
CF	76.4%	0.775
Tucker	87.9%	0.892
STM-Tucker-SGD	**90.2%**	**0.933**
STM-CP-SGD	**90.1%**	**0.930**
STM-CP-2SGD	**90.4%**	**0.938**



(a) Loss with different t. (b) Time with different t.

(c) Peak memory with different N. (d) Time with different R.

(e) Accuracy with different R. (f) Loss with different R.

Fig. 2. Comparisons of different datasets.

both *STM-Tucker-SGD* and *STM-CP-SGD* in terms of the convergence rate, as well as the loss function. Fig. 2b compares the running time through each iteration. The results manifest that the running time of *STM-CP-SGD* is the fastest for each iteration, with *STM-CP-2SGD* as the close second (which terminates first). They both significantly outperforms *STM-Tucker*. The overall running time of *STM-CP-2SGD* is the smallest since it requires much fewer iterations. Fig. 2c shows the peak memory usage of different tensor decomposition methods. The memory usage of *STM-CP-2SGD* is slightly greater than that of the others. Notice that the feature tensor and adjacency matrix are sparse, and therefore we use sparse tensor representation for each decomposition to reduce the memory usage.

Moreover, Figs. 2d, 2e, and 2f compare the performance of *STM-Tucker-SGD*, *STM-CP-SGD*, and *STM-CP-2SGD* in terms of running time, accuracy, and loss function with different R, respectively. As shown in Fig. 2d, *STM-CP-2SGD* significantly outperforms the other two in terms of running time for different R, i.e., the running time is at most 9.7 and 18.9 percent of *STM-Tucker-SGD* and *STM-CP-SGD*, respectively, while the accuracy for detecting SNMD is almost the same for different proposed methods as shown in Fig. 2e, which shows the power of acceleration of the proposed *STM-CP-2SGD* without sacrificing accuracy. Moreover, the accuracy of different methods does not increase as R grows since the latent features may overfit the training data and thus do not perform well on testing data. Fig. 2f further shows the loss function with different R. As R increases, the loss functions for different methods all decrease. For Figs. 2e and 2f, although the loss function of *STM-Tucker-SGD* slightly outperforms *STM-CP-2SGD*, the accuracy of *STM-Tucker-SGD*, *STM-CP-SGD*, and *STM-CP-2SGD* are similar. In summary, *STM-CP-2SGD* is the most efficient one without compromising efficiency and accuracy.

5.5 Feature Study

To observe the differences among the three types of SNMDs, Table 4 lists the top-5 discriminative features using

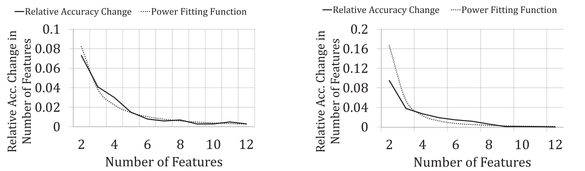
TABLE 4
Top Features and Acc. on the FB_US Dataset

CR	NC	IO	
Parasociality	Game posts	Median of BI	Median of BI
Median of BI	Online/offline ratio	Online/offline ratio	Online/offline ratio
Eigenvector centrality	Parasociality	SD of BL	Parasociality
Online/offline ratio	Shannon index	Sticker number	Sticker number
Sticker number	Social comparison score	Social roles	SD of BL
Acc.: 80.5%	Acc.: 77.6%	Acc.: 82.9%	Acc.: 80.7%

information gain and the corresponding accuracy on the FB_US dataset by TSVM, where CC, BI, BL, and SD respectively denote the clustering coefficient, burst intensity, burst length, and standard deviation. It is worth noting that the number of selfies, an indicator of self-disclosure, is not useful for detecting CR and IO, but it is effective for NC. This is because NC users are usually less socially active, comparing to CR and IO users. Moreover, the online/offline interaction ratio of NC is much higher than the ratios of the other two types, probably because NC users usually show less willingness to join offline activities. In contrast, CR and IO cases prefer to use social media, instead of playing games alone. Moreover, people with compulsive personality are more introverted. In contrast, people with CR usually create virtual bonds to develop pathological relationships for compensation of their (missing) offline relationships. The Shannon index, an indicator of the social diversity, is also useful in detecting NC since the friends of NC are in similar backgrounds, and thus the Shannon index is lower than that of normal users. Moreover, the social comparison score is important for detecting NC cases. This is because when the users with malicious envy see the positive newsfeeds from the friends with similar backgrounds, they may eager to pursue the sense of success, which is much easier to be achieved in online games.

The parasociality, effective for detecting all SNMD types, is especially useful for detecting CR cases. For example, in our user study, we find user A, 21-year-old male, frequently posting news feeds, such as “I’m so bored :(((((...Ahhhhhh!!”, and his cross-dressing photos on his Facebook timeline, more than 3 times a week, which usually get fewer than 5 likes. At the same time, he “likes” a large number of posts from others. SNMDD classifies him as a potential CR case and his questionnaire reveals that he constantly blocks out disturbing thoughts about life and finds himself anticipating when he goes online again.

Burst intensity and length are quite useful for detecting IO cases. For example, user B, 36-year-old male, is detected as IO since the behavior of clicking “likes” fits the pattern of bursts, i.e., the median of his burst intensity is high, equal to 31. His answers to the standard questionnaire reveal that he loses sleep due to late-night access on Facebook to check others’ news feeds. Through interview, user B explains that he cannot stop checking for new posts and e-mails even when all his news feeds and emails are read. Some of his friends reply him: “are you a robot? no sleep needed?!?!!”, indicating that user B is indulged in finding social news. Moreover, social roles are important in detecting IO since the users with IO usually share or like the information from different communities ant thus are inclined to be detected as structural holes.



(a) Relative improvement w.r.t. the number of features.

(b) Relative improvement w.r.t. the number of latent features (*STM*).

Fig. 3. Relative accuracy change with respect to number of features.

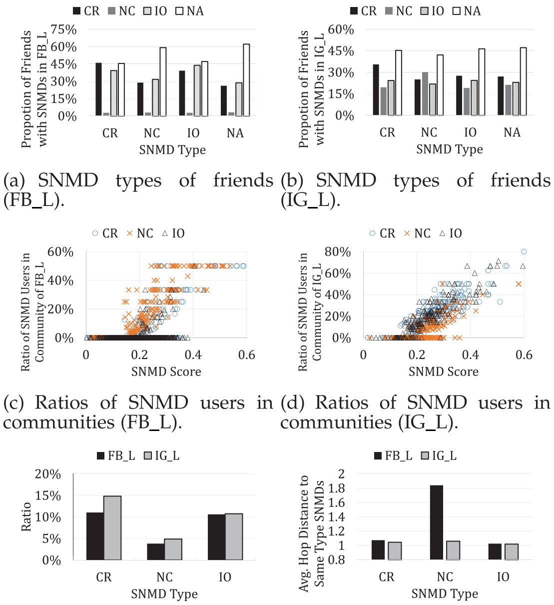
Next, we analyze the importance of different features to our classifiers. χ^2-test is exploited to measure the importance of each feature via SelectKBest of Scikit-Learn. The top 5 important features overall are 1) median of the intensity of bursts, 2) parasociality, 3) online/offline interaction ratio, 4) number of used stickers, and 5) standard deviation of the length of bursts. It is worth noting that TSVM using only these 5 features in SNMDD achieves an accuracy of 80.7 percent for FB_US, close to that of using all features (`All`). In other words, integrating important social and personal features provides good results because effective personal features, e.g., the temporal behavior features, can be used to differentiate the users suffering from withdraw or relapse symptoms and heavy users, while social features capture the interactions among users to differentiate different SNMDs.

Figs. 3a and 3b show the improvement made by adding different features in TSVM on the FB_US dataset and the proposed *STM* on multi-source data (i.e., FB_US and IG_US). The feature selection of TSVM is based on the information gain (the top-5 features mentioned earlier), while the tensor approach automatically extracts important latent features. We observe a diminishing return property on both figures, where the improvement becomes marginal as more features are included. Fig. 3a shows a power fit function ($p(x) = 0.3091x^{-1.92}$) of the curve with $R^2 = 0.9534$. The exponent -1.92 denotes that the improvement by adding nth feature is $n^{-1.92}$ times smaller than that by adding the first feature. On the other hand, the results of the tensor-based approach in Fig. 3b show that the accuracy increment for adding a single feature drops faster ($p(x) = 1.11x^{-2.01}$) since the proposed *STM* can extract much more important and concise features.

5.6 Analysis of SNMD Types in Large Datasets

In this analysis, we first apply the proposed SNMDD framework (with TSVM) on some large-scale OSN datasets, i.e., FB_L and IG_L, to classify their users. In Figs. 4a and 4b, we analyze the detected SNMD cases among the friends of an SNMD user. In Fig. 4a, the leftmost bar indicates that in FB_L, among all CR users, about 45 percent of their friends are also CR users, which is greater than the percentage of other SNMD types. On the other hand, the 8th bar from the left in Fig. 4a indicates that in FB_L, about 59 percent of NC users' friends are NA (non-SNMD users). Figs. 4a and 4b show that, in FB_L and IG_L, CR and IO users have similar friend types. This is because CR and IO cases, by their nature, are similar, i.e., they are both seeking social satisfaction (e.g., relationships and information) from the OSNs. Moreover, among different SNMD cases, CR and IO users are likely to be friends with other CR and IO users. For CR users, this phenomenon has been described as "loneliness propagates" [12].

Furthermore, Infomap community detection [42] is performed on FB_L and IG_L to derive the relationships between different types of SNMD users in their communities. Figs. 4c



(a) SNMD types of friends (FB_L).

(b) SNMD types of friends (IG_L).

(c) Ratios of SNMD users in communities (FB_L).

(d) Ratios of SNMD users in communities (IG_L).

(e) SNMD users.

(f) Hop distance to same type SNMD users.

Fig. 4. Comparisons of different datasets.

and 4d analyze the community structures of SNMD users with different SNMD scores, where each point represents the characteristic of a community. Specifically, each community in the dataset is represented by three different types of points, i.e., CR, NC, and IO. For example, each CR point is represented as $\langle score, ratio \rangle$, where $score$ is the average CR score in that community, and $ratio$ indicates the proportion of CR users in the community. It is similar for each IO/NC point. As Figs. 4c and 4d show, for each SNMD type, when the average SNMD score is higher, it is likely to have more SNMD users in the community. Moreover, there are many communities with large IO scores in IG_L that have IO ratios close to 1. This implies that the users with large IO scores in IG_L are more inclined to form homogeneous groups. At the first glance, one may feel that NC users frequently appear in many communities, and there seems to be a large number of NC users, especially in FB_L (i.e., Fig. 4c). However, after carefully examining these communities, we find that those communities (with large ratios of NC users) are usually very small (usually with the size around 5) because NC users are less-active. On the other hand, in IG_L, when SNMD scores are larger, the ratios of IO users in communities are also larger. This is because IO users can view, like, or follow others in Instagram more easily (not necessary to be friends first).

Fig. 4e compares the ratios of different types of SNMD users identified in FB_L and IG_L. There are more CR users in IG_L probably because CR users seek social supports online to compensate the loneliness in real life. We argue that the Instagram platform makes it easy to freely create social relationships with strangers. In contrast, it is not that easy to create new social relationships on Facebook since the friend requests need to be approved. Finally, Fig. 4f compares the average number of hops from each SNMD user to the nearest user with the same type of SNMDs. The leftmost bar shows that the average hop distance from each CR user to the closest CR user is 1.07 hop, indicating that CR and IO users are close to other same-type users, i.e., average hop distances are within 1.15, where Figs. 4a and 4b also report similar results.

6 Conclusion

In this paper, we make an attempt to automatically identify potential online users with SNMDs. We propose an SNMDD framework that explores various features from data logs of OSNs and a new tensor technique for deriving latent features from multiple OSNs for SNMD detection. This work represents a collaborative effort between computer scientists and mental healthcare researchers to address emerging issues in SNMDs. As for the next step, we plan to study the features extracted from multimedia contents by techniques on NLP and computer vision. We also plan to further explore new issues from the perspective of a social network service provider, e.g., Facebook or Instagram, to improve the well-beings of OSN users without compromising the user engagement.

Acknowledgments

This work is supported in part by the US National Science Foundation through grants IIS-1526499, IIS-1717084, CNS-1626432, and NSFC 61672313 and by the Ministry of Science and Technology in Taiwan through grants 106-2221-E-009-154-MY2, 106-3114-E-009-010, 105-2218-E-002-035, 106-2221-E-001-015-MY2, 105-2628-E-001-001-MY2, 104-2221-E-001-005-MY2, and 105-2218-E-007 -032-MY2.

References

[1] E. Acar, D. M. Dunlavy, and T. G. Kolda, "A scalable optimization approach for fitting canonical tensor decompositions," *J. Chemometrics*, vol. 25, pp. 67–86, 2011.

[2] A. Anandkumar, D. Hsu, M. Janzamin, and S. Kakade, "When are overcomplete topic models identifiable? uniqueness of tensor tucker decompositions with structured sparsity," *J. Mach. Learn. Res.*, vol. 16, pp. 2643–2694, 2015.

[3] C. Andreassen, T. Torsheim, G. Brunborg, and S. Pallesen, "Development of a Facebook addiction scale," *Psychological Rep.*, vol. 110, pp. 501–517, 2012.

[4] H. Appel, J. Crusius, and A. L. Gerla, "Social comparison, envy, and depression on Facebook: A study looking at the effects of high comparison standards on depressed individuals," *J. Social Clinical Psychology*, vol. 34, no. 4, pp. 277–289, 2015.

[5] Y. Baek, Y. Bae, and H. Jang, "Social and parasocial relationships on social network sites and their differential relationships with users' psychological well-being," *Cyberpsychology Behavior Social Netw.*, vol. 16, pp. 512–517, 2013.

[6] D. La Barbera, F. La Paglia, and R. Valsavoia, "Social network and addiction," *Cyberpsychology Behavior*, vol. 12, pp. 628–629, 2009.

[7] E. Baumer, et al., "Limiting, leaving, and (re)lapsing: An exploration of Facebook non-use practices and experiences," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2013, pp. 3257–3266.

[8] J. Block, "Issues of DSM-V: Internet addiction," *Amer. J. Psychiatry*, vol. 165, pp. 306–307, 2008.

[9] A. Bordes, L. Bottou, and P. Gallinari, "SGD-QN: Careful quasi-newton stochastic gradient descent," *J. Mach. Learn. Res.*, vol. 10, pp. 1737–1754, 2009.

[10] L. Bottou, "Stochastic gradient descent tricks," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012.

[11] K. Caballero and R. Akella, "Dynamically modeling patient's health state from electronic medical records: A time series approach," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 69–78.

[12] J. Cacioppo, J. Fowler, and N. Christakis, "Alone in the crowd: The structure and spread of loneliness in a large social network," *J. Personalityt Social Psychology*, vol. 97, pp. 977–991, 2009.

[13] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummad, "Measuring user influence on Twitter: The million follower fallacy," in *Proc. Int. AAAI Conf. Weblogs Social Media*, 2010, pp. 10–17.

[14] K. Chak and L. Leung, "Shyness and locus of control as predictors of internet addiction and internet use," *Cyberpsychology Behavior*, vol. 7, pp. 559–570, 2004.

[15] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," 2001.

[16] C.-H Chang, E. Saravia, and Y.-S. Chen, "Subsonscious crowdsourcing: A feasible data collection mechanism for mental disorder detection on social media," in *Proc. Int. Conf. Advances Social Netw. Anal. Mining*, 2016, pp. 374–379.

[17] F. Chang, C.-Y. Guo, X.-R. Lin, and C.-J. Lu, "Tree decomposition for large-scale SVM problems," *J. Mach. Learn. Res.*, vol. 11, pp. 2935–2972, 2010.

[18] M. Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Proc. Int. AAAI Conf. Weblogs Social Media*, 2013, pp. 1–10.

[19] R. Collobert, F. Sinz, J. Weston, and L. Bottou, "Large scale transductive SVMs," *J. Mach. Learn. Res.*, vol. 7, pp. 1687–1712, 2006.

[20] P. Comon, X. Luciani, and A. L. D. Almeida, "Tensor decompositions, alternating least squares and other tales," *J. Chemometrics*, vol. 23, pp. 393–405, 2009.

[21] E. Ferrara, R. Interdonato, and A. Tagarelli, "Online popularity and topical interests through the lens of Instagram," in *Proc. ACM Conf. Hypertext Social Media*, 2014, pp. 24–34.

[22] R. Ge, F. Huang, C. Jin, and Y. Yuan, "Escaping from saddle points-online stochastic gradient for tensor decomposition," in *Proc. 28th Conf. Learning Theory*, 2015, pp. 797–842.

[23] J. Gill and G. King, "What to do when your Hessian is not invertible: Alternatives to model respecification in nonlinear estimation," *Sociol. Methods Res.*, vol. 32, pp. 54–87, 2004.

[24] S. Graham, A. Munniksma, and J. Juvonen, "Psychosocial benefits of cross-ethnic friendships in urban middle schools," *Child Develop.*, vol. 85, pp. 469–483, 2014.

[25] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis," *UCLA Work. Papers Phonetics*, vol. 16, pp. 1–84, 1970.

[26] K. Hayashi, T. Maehara, M. Toyoda, and K. Kawarabayashi, "Real-time top-R topic detection on twitter with topic hijack filtering," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 417–426.

[27] L. He, C.-T. Lu, J. Ma, J. Cao, L. Shen, and P. S. Yu, "Joint community and structural hole spanner detection via harmonic modularity," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 875–884.

[28] R. Jain and N. Abouzakhar, "A comparative study of hidden Markov model and support vector machine in anomaly intrusion detection," *Int. J. Internet Technol. Secured Trans.*, vol. 2, pp. 176–184, 2013.

[29] K. Kim, et al., "Poor sleep quality and suicide attempt among adults with internet addiction: A nationwide community sample of Korea," *PLoS One*, vol. 12, 2017, Art. no. e0174619.

[30] D. L. King, P. H. Delfabbro, D. Kaptsis, and T. Zwaans, "Adolescent simulated gambling via digital and social media: An emerging problem," *Comput. Human Behavior*, vol. 31, pp. 305–313, 2014.

[31] J. Kleinberg, "Bursty and hierarchical structure in streams," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2002, pp. 91–101.

[32] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, pp. 455–500, 2009.

[33] B. Kågström and P. Poromaa, "Distributed and shared memory block algorithms for the triangular Sylvester equation with sep^{-1} estimators," *SIAM J. Matrix Anal. Appl.*, vol. 13, pp. 90–101, 1992.

[34] L. Leung, "Net-generation attributes and seductive properties of the internet as predictors of online activities and internet addiction," *Cyberpsychology Behavior Soc. Netw.*, vol. 7, pp. 333–348, 2004.

[35] S. S. Levine, E. P. Apfelbaum, M. Bernard, V. L. Bartelt, E. J. Zajac, and D. Stark, "Ethnic diversity deflates price bubbles," in *Proc. Nat. Academy Sci. USA*, vol. 111, pp. 18524–18529, 2014.

[36] D. Li, X. Li, L. Zhao, Y. Zhou, W. Sun, and Y. Wang, "Linking multiple risk exposure profiles with adolescent Internet addiction: Insights from the person-centered approach," *Comput. Human Behavior*, vol. 75, pp. 236–244, 2017.

[37] I.-H. Lin, et al., "The association between suicidality and Internet addiction and activities in Taiwanese adolescents," *Comprehensive Psychiatry*, vol. 55, pp. 504–510, 2014.

[38] K.-L. Liu, W.-J. Li, and M. Guo, "Emoticon smoothed language models for twitter sentiment analysis," in *Proc. AAAI Conf. Artif. Intell.*, 2012, pp. 1678–1684.

[39] T. Maehara, K. Hayashi, and K. Kawarabayashi, "Expected tensor decomposition with stochastic gradient descent," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 1919–1925.

[40] Y. Nesterov and B. T. Polyak, "Cubic regularization of Newton method and its global performance," *Math. Program.*, vol. 108, pp. 177–205, 2006.

[41] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. New York, NY, USA: Academic Press, 1970.

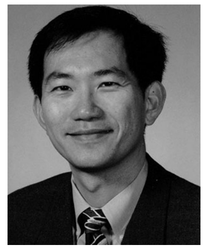
[42] M. Rosvall and C. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Nat. Academy Sci. United States America*, vol. 105, pp. 1118–1123, 2008.
[43] M. Saar-Tsechansky and F. Provost, "Handling missing values when applying classification models," *J. Mach. Learn. Res.*, vol. 8, pp. 1623–1657, 2007.
[44] B. Saha, T. Nguyen, D. Phung, and S. Venkatesh, "A framework for classifying online mental health-related communities with an interest in depression," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 4, pp. 1008–1015, Jul. 2016.
[45] H.-H. Shuai, et al., "Mining online social data for detecting social network mental disorders," in *Proc. Int. Conf. World Wide Web*, 2016, pp. 275–285.
[46] V. Silva and L.-H. Lim, "Tensor rank and the Ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal.*, vol. 30, pp. 1084–1127, 2006.
[47] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, "User-level sentiment analysis incorporating social networks," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 1397–1405.
[48] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279–311, 1966.
[49] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in Facebook," in *Proc. 2nd ACM Workshop Online Social Netw.*, 2009, pp. 37–42.
[50] J. B. White, E. J. Langer, L. Yariv, and J. C. Welch IV, "Frequent social comparisons and destructive emotions and behaviors: The dark side of social comparisons," *J. Adult Develop.*, vol. 13, pp. 36–44, 2006.
[51] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco, CA, USA: Morgan-Kaufmann, 2000.
[52] K. Young, "Internet addiction: The emergence of a new clinical disorder," *Cyberpsychology Behavior*, vol. 1, pp. 237–244, 1998.
[53] K. Young, M. Pistner, J. O'Mara, and J. Buchanan, "Cyber-disorders: The mental health concern for the new millennium," *Cyberpsychology Behavior*, vol. 2, pp. 475–479, 1999.
[54] L. Zhao, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan, "Hierarchical incomplete multi-source feature learning for spatiotemporal event forecasting," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 2085–2094.

Hong-Han Shuai received the BS degree from the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, Taiwan, R.O.C., in 2007, the MS degree in computer science from the National Taiwan University, in 2009, and the PhD degree from the Graduate Institute of Communication Engineering, in 2015. He is an assistant professor with NCTU. His research interests include the area of social network analysis and data mining.

Chih-Ya Shen received the BS and MS degrees from the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 2005 and 2007, respectively, and the PhD degree from the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, in 2013. He is an assistant professor of the Computer Science Department, National Tsing Hua University. His research interests include big data and social network analytics, query processing, and mobile computing.

De-Nian Yang received the BS and PhD degrees from the Department of Electrical Engineering, National Taiwan University, in 1999 and 2004, respectively. He is now a research fellow in the Institute of Information Science, Academia Sinica, Taiwan. His research interests include mobile social networks and mobile multimedia networking. He received K.-T. Li Distinguished Young Scholar Award in ACM Taipei/Taiwan Chapter, Research Exploration Award from Pan Wen Yuan Foundation, and Project for Excellent Junior Research Investigators in NSC, Taiwan. He is a senior member of the IEEE and a member of the ACM.

Yi-Feng Carol Lan received the PsyD degree in clinical psychology from AIU, CSPP. She is currently serving as an adjunct assistant professor with Tamkang University and Tunghai University with research topics focus on psychopathology and culture psychology. Her work experience includes serving as an individual/group psychologist to clients with various psychological disorders. Her counseling specialties include psychological/emotional stress, self accomplishment, interpersonal relationships, self-confidence and inferiority, and issues of gender relationship.



Wang-Chien Lee received the BS degree from National Chiao Tung University, Hsinchu, Taiwan, the MS degree from Indiana University, Bloomington, and the PhD degree from the Ohio State University, Columbus. He is an associate professor of computer science and engineering with Pennsylvania State University, University Park. He is particularly interested in developing data management techniques for supporting complex services in a wide spectrum of networking and mobile environments. He is a member of the IEEE.

Philip S. Yu received the BS degree in electrical engineering from National Taiwan University, the MS and PhD degrees in electrical engineering from Stanford University, and the MBA degree from New York University. He is a distinguished professor in computer science with the University of Illinois at Chicago and also holds the Wexler chair in information technology. Before joining UIC, he was with IBM, where he was manager of the Software Tools and Techniques Group, Watson Research Center. His research interest is on big data, including data mining, data stream, database, and privacy. He has published more than 830 papers in refereed journals and conferences. He holds or has applied for more than 300 US patents. He is a fellow of the ACM and the IEEE. He is the editor-in-chief of the *ACM Transactions on Knowledge Discovery from Data*. He received the IEEE Computer Society 2013 Technical Achievement Award for pioneering and fundamentally innovative contributions to the scalable indexing, querying, searching, mining, and anonymization of big data, the ICDM 2013 10-year Highest-Impact Paper Award, the EDBT Test of Time Award (2014), and the Research Contributions Award from IEEE International Conference on Data Mining (2003).

Ming-Syan Chen received the PhD degree in computer, information, and control engineering from the University of Michigan, Ann Arbor, Michigan. He is now the dean of the College of Electrical Engineering and Computer Science and also a distinguished professor in the EE Department, National Taiwan University. His research interests include databases, data mining, machine learning, and social networks. He is a Fellow of the ACM and the IEEE.