5th International Conference on Research Developments in Applied Science, Engineering & Management (AEM- 2019)

Mahratta Chamber of Commerce, Industries and Agriculture ,Pune (India) AEM- 2019

23rd February 2019, www.conferenceworld.in ISBN: 978-93-87793-72-9

Detecting Stress Based on Social Interactions in Social Networks

Prof. S.P. Khandait, Akash. S. Ganvir, Sunny. M. Sharma

Head of Dept. of Information & Technology, KDK College of Engineering, Nagpur-440009
Students of Dept. Information Technology, KDK College Of Engineering, Nagpur-440009

Abstract- Psychological wellness influences a noteworthy level of the world's population. The stress investigation of emotional wellness phenomena in openly accessible social networking sites like twitter and social timeline. A set of stress-related textual, visual, and social attributes from various aspects are first defined and then propose a novel hybrid model. The work has demonstrated the utility of online social information for contemplating despondency, be that as it may, there have been limited assessments of other mental well being conditions .It is not easy to access the user posts on their social timeline page. In order to obtain the user data from social timeline, system have to get the access token from social timeline. The system will also help to Recommending users with different links for psychological counseling centers, soft music or articles to help release their stress according to users' stress level. Results show that the suggested model had improve the detection performance by 6-9 percent in F1-score the number of social structures of sparse connections of stressed users is around 14 % greater than that of de-stressed users, indicating that the social structure of stressed users' network tend to be less connected and complicated than that of non-stressed users.

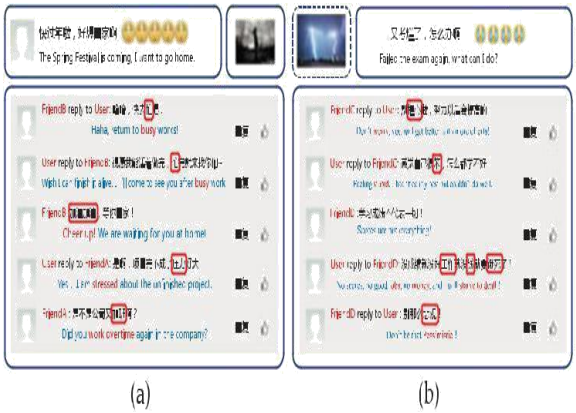
Keywords: Stress detection, social media, micro-blog, access tokens, and face-book.

1 INTRODUCTION:

The Rise of Social World is Changing People's Life, as Well as Research in Healthcare . With the development of social media like Twitter and facebook , more people are willing to post their daily events and moods, and interact with strangers through the social networks. As these social media data timely reflect users' real-life states and emotions in a timely manner, it offers new opportunities measuring, modeling, and mining user's behavior patterns through the large-scale social networks and such social information can find its theoretical basis in stress research. For example, [7] found that stressed users are more likely to be socially less active, and more recently, there have been research efforts on harnessing social networking data for developing mental healthcare tools. Thus, there is importance to detect stress before it turns into severe problems. However, traditional techniques are actually reactive, which are usually labor-consuming, timely and hysteretic. First, tweets are limited to a maximum of 140 characters on social platforms like Twitter and users do not always express their stressful states directly in tweets. Second, users with high psychological stress may exhibit low

 ISBN: 978-93-87793-72-9

activeness on social networks, as reported by a recent study in Pew Research Center.



(a) (b)

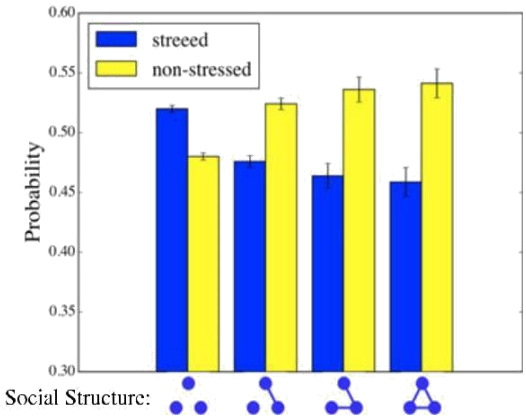
These phenomena become subject to inherent data sparsity and ambiguity problem, which may hurt the performance of tweeting content based stress detection performance. For illustration, Sina Weibo tweet example in Fig. 1. The tweet contains only 14 characters, describing that the user wished to go home for the Spring Festival. Although no stress is revealed from the tweet itself, from the follow-up interactive comments made by the user, we can find that the user is actually stressed from work. Thus, can't simply rely on user tweet.

1.1. OUR WORK:

Inspired by psychological theories, we first define a set of attributes for stress detection from tweet-level and user-level aspects respectively: 1) tweet-level attributes from con-tent of user's single tweet, and 2) user-level attributes from user's weekly tweets. The tweet attributes are mainly enclosed with linguistic, visual, and social attention attributes extracted from a single-tweet's text, image, and attention list. The user-level attributes however enlisted as:

(a) posting behavior attributes as summarized from a user's weekly tweet postings; and (b) social interaction attributes extracted from a user's social interactions with friends.

The social interaction attributes can further described as: (i) social interaction content attributes derived from the content of users social interactions; and (ii) social inter-action structure attributes extracted from the structures of users'



social interactions with friends. Both the structure and content we derived from the user social media timeline helps us to describe and get to the result whether the user tweet helps you to identify the stress.To maximally leverage the user-level information as well as tweet-level content information, we propose a novel hybrid model of factor graph model combined with a convolutional neural network (CNN).. The overall steps are as follows: 1) we first design a convolutional neural network (CNN) with cross auto-encoders (CAE) to generate user-level content attributes from tweet-level attributes; and 2) we define a partially labeled factor graph (PFG) to combine user-level social interaction attributes, user-level posting behavior attributes and the learnt user-level content attributes for stress detection. Evaluating the model as well as the different attributes on a real-world dataset from Sina Weibo. Experimental

results show that by exploiting the users' social interaction attributes, the proposed model can improve the detection performance (F1-score) by 6-9 percent over that of the state-of-art methods. This indicates that the proposed attributes can serve as good cues in tackling the data scarcity and ambiguity problem. Moreover, the pro-posed model can also efficiently combine tweet content and social interaction to enhance the stress detection performance. This stress detection module can detect the user stress effectively and efficiently. Whereas the data analysis and data tempering plays great role while detecting the user stress.

2. RELATED WORK:

Psychological stress detection is related to the topics of sentiment analysis and emotion detection. Research on Tweet-Level Emotion Detection in Social Networks. Computer-aided detection, analysis, and application of emotion, especially in social networks, have drawn much attention in recent years [8], [9], [2], [4], [5]. Relationships between personality traits and psychological stress can be an interesting issue to consider [10]. For ex.[1] providing evidence that daily stress can be reliably recognized based on behavioral metrics from users mobile activity. Studies on social media based emotion are at the tweet level, using classic classification approaches and text-based linguistic features, classifying the emotion categories into four types, i.e., angry, disgusting, joyful, and sad. Fan et al[9],studied the emotion propagation problem in social networks, and found that anger has a stronger correlation among different users than joy, indicating that negative emotions could spread more quickly and broadly in the network. As stress is considered as a negative emotion, this can help us in combining the social influence detection. However, these work leverage the textual contents in social networks. Data in social networks is usually composed of sequential and inter-connected items from modalities and diverse sources, making cross-media data.

3. PROBLEM FORMULATION

Before presenting our problem statement, let's first define some necessary notations. Let V be a set of users on a social network, and let jV j denote the total number of users. Each user v_i 2 V posts a series of tweets, with each tweet containing text, image, or video content; the series of tweets contribute to users social interactions on the social network.

Definition 1 (Stress state). The stress state y of user v_i 2 V at time t is represented as a triple ðy; v_i; tÞ, or briefly y^t_i. In the study, a binary stress state y^t_i 2 f0; 1g is considered, where y^t_i ¼ 1 indicates that user v_i is stressed at time t, and y^t_i ¼ 0 describes that the user is non-stressed at time notated as t, which can be identified from expressions in user tweets or clearly identified by user. Let Y^t be the set of stress states of users at time t.

Definition 2 (Time varying user-level attribute matrix).Each user in V is associated with a set of attributes A. Let X^t be a jV j jAj attribute matrix at time t, in which every row $\mathbf{x}^t_i$ corresponds to a user, each column corresponds to an attribute, and an element $x^t_{i;j}$ is the jth attribute value of user v_i at time t.

A user-level matrix attributes define user specific features, and can be defined in different ways. This study considers user-level content attributes,

statistical attributes, and social interaction attributes.

Definition 3 (Time-varying edge set). Users are linked by edges of certain types. Let E^t V V C be a set of edges between users at time t. Three varieties of edges can be considered in the study. For an edge e ¼ ðvi; vj; cÞ 2 E^t, c ¼ 0 indicates that vi follows or is followed by vj at time t, c indicates that there are positive words in comments between user vi and vj at time t, and c ¼ 2 indicates that there are negative words in comments between them at time t

TABLE 1

Summary of Tweet-Level Attribute

Category	Short Name	#	Description
Linguistic	Positive & Negative Emotion Words	2	Number of positive and negative emotion words
	Positive & Negative Emoticons	2	Number of popular positive and negative emoticons, e.g., and
	Punctuation Marks & Associated Emotion Words	4	To signify the intensity of emotion four typical punctuation mark: ('!', '?', '...', '.') are considered.
	Degree Adverbs & Associated Emo- tion Words	2	In examples "{I feel a little bit sad}" and "{I feel terribly sad}" 'sad' expresses different negative feelings. We use 1-3 to represent neutral moderate, and severe degree of positive emotions, and the minus to represent the negative ones.
Visual	Five-color theme	15	A combination of five dominant colors in HSV color space indicating main color distribution of images, has been revealed to be important or human emo- tions by psychology and art theories.
	Saturation	2	The mean value of saturation and its contrast.
	Brightness	2	The mean value of brightness and its contrast.
	Warm/Cool color	1	Ratio of cool colors with hue ([0-360]) in the HSV space in [30 100].
	Clear/Dull color	1	Ratio of colors with brightness ([0-1]) and saturation < 0.6.
Social	Social Attention	3	Number of comments, retweets, and likes

Definition 4 (Time-varying attribute augmented network). An attribute-augmented network at time t is comprised of four elements, including 1) a user set V^t, 2) an edge set E^t, 3) a user-level attribute matrix set X^t, and a stress state set for all users Y^t at time t, denoted as G^t ¼ ðV^t; E^t; X^t; Y^tÞ. Given a sequence of labeled time varying attribute augmented networks at different times, our goal is to learn a model that can best fit the relationships among users' stress states, user-level attributes, and users social linkage to detect users stress states with the model.

Problem 1(Psychological stress detection). Given a series of T partially labeled time-varying attribute-augmented net- works

fG^t ¼ ðV_L^t; V_U^t ; E^t; Y_L^tÞ j t 2 f1; 2; . . . ; T gg,

V_L^t is a set of users

with labeled stress states Y_L^t at time t, and V_U^t is a set of un-labeled users to learn a function.

f:fG^1; G^2; . . . G^T g ! fY_U^1; Y_U^2; . . .Y_U^T g

4. ATTRIBUTES CATEGORIZATION AND DEFINITION

To address the problem of stress detection, we first define two sets of attributes to measure the differences of the stressed and non-stressed users on social media platforms:

1) tweet-level attributes from a user's single tweet;

2) user-level attributes summarized from a user'.

4.1. Tweet-Level Attributes

Tweet-level attributes describe the linguistic and visual con-tent, as well as social attention factors (being liked, commented, and re-tweeted) of a single tweet.For linguistic attributes, we take the most commonly used linguistic features in sentiment

4.1. Tweet-Level Attributes

Tweet-level attributes describe the linguistic and visual con-tent, as well as social attention factors (being liked, commented, and re-tweeted) of a single tweet.

For linguistic attributes, we take the most commonly used linguistic features in sentiment analysis research. We adopt

LTP [4]—A Chinese Language Platform—to perform lexical analysis, e.g., tokenize and lemmatize, and then explore the use of a Chinese LIWC dictionary—LIWC2007 [4], to map the words into +ve/-ve emotions. LIWC 2007 is a dictionary which categorizes words based on their

4.2 User-Level Attributes

Comparing to tweet-level extracted from a tweet, user-level attributes are extracted from a list of users tweets in a pre-defined sampling period. On one hand, psychological stress usually results from cumulative events or mental states. On the other hand users express their chronic stress in a series of

tweets. Besides, the aforementioned social interaction patterns of users in a specified time also contain useful information for stress detection.

psychological meanings, so we can classify words into different categories, e.g., +ve/-ve emotion words, degree adverbs. We have also tested other linguistic resources including NRC[5] and HowNet,[6] and found that the performances were relatively the same, so we adopted the commonly used LIWC2007 dictionary for experiments. Furthermore, we extract linguistic attributes of emoticons (e.g., ☺ and ☹) and punctuation marks ('!', '?', '...', '.' Every emoticon in square brackets according to Weibo (e.g., they use [haha] for "laugh"), so we can map the keyword in square brackets to find the emoticons. Twitter adopts

Unicode as the representation for all emojis [Table1]

Moreover, as aforementioned, the information in tweets is limited and sparse, As we need to integrate more complementary information to the tweets, e.g., users social interactions.

Thus, designed user-level attributes can provide a macro scope of a users stress states, and avoid noise/missing data. Hereby, we define user attributes from two aspects to measure the differences between stressed and non-stressed states depending on users weekly tweet postings: 1) user-level posting behavior attributes from the users weekly tweet postings; and 2) user-level

social interaction attributes from the user's social interactions beneath his weekly postings.

5. MODEL FRAMEWORK

Two challenges exist in stress detection.

1) How to extract user attributes from users tweeting series and dealing with the problem of absence of modality in the tweets?

2) How to fully leverage social interaction, including structure patterns and interaction content, for stress detection. To tackle these challenge. A novel hybrid model by combining a factor graph model (FGM) with a convolutional neural network (CNN), As CNN is capable of learning unified latent features from multiple modalities, and factor graph model is best at modeling the correlations. We will first introduce the architecture of our model, and then describe the stress detection described modules.

Category	Short Name	#	Description
Social Engagement		3	The numbers of @-mentions, @-retweets, and @-replies in weekly tweet postings, indicating one's social interaction activeness with friends.
Posting Behavior		2	The numbers of tweets posted in hours with a 24-dimensional vector.
Tweeting time		4	Categorize users' tweets into mainly four types based on general cate-gories of social media platforms: (1) Image tweets (tweets containing images); (2) Original tweets (tweets that are originally authored and posted by the user); (3) Information query tweets (tweets that ask questions or ask for help); (4) Information sharing tweets (tweets that contain outside hyperlinks). We use a 4-dimensional vector of the numbers of tweets in the above 4 types respectively to quantify the tweeting type attribute.
Tweeting type		4	
Tweeting linguistic style		10	Adopt 10 categories from LIWC that are related to daily life, social events, e.g., personal pronouns, home, work, money, religion, death, health, ingestion, friends, and family. We extract words from users' weekly tweet postings, and use a 10-dimensional vector of numbers of words in the 10 categories
Content Style	Words	10	A 10-dimensional integer vector, with each value representing the number of words from social interaction content of users weekly tweet postings in each word category from LIWC;
	Emoticons	2	A 2-dimensional integer vector with each value representing the num-ber of positive and negative emoticons (e.g., and) in tweets.
Social Interaction	Stressed Neighbor Count	1	The number of the user's stressed neighbors.
Social Influence	Strong-tie Count	1	The number of stressed neighbors with strong tie.
	Weak-tie Count	1	The number of stressed neighbors with weak tie.
	Follower Count	1	The number of the user's followers.
	Fans Count	1	The number of the user's fans.
Social Structure			Representing the structure distribution of the user's interacted friends, where each element refers to the existence of the corresponding struc-ture in Fig. 6

5.1 Architecture

Fig. 3 shows the architecture of our model. There are three varieties of information that we can use as the initial inputs, i.e., tweet-level attributes, user-level posting behavior attributes, and user-level social interaction attributes, whose detailed computation will be described later. Solution through the following two key components: First, we design a CNN with cross auto-encoders (CAE) to generate user-level interaction content attrib-utes from tweet-level attributes

The CNN has been found to be effective in learning stationary local attributes for series like images [3] and audio.

The model consists of two parts: (i) CNN(conventional Neural Network) & (ii) FGM(Factor Graph Modeling). The CNN is used to generate user-level attributes by convolution with CAE filters as input to the FGM (Factor Graph Modeling). Take user labeled with a red star as example. Tweet-level attributes of the user are implemented through a convolution with CAE to form the user-level attributes.

FGM, attribute factors relates user-level attributes to connect the stress state of multiple users. Dynamic factors connect stress state of a user timely.

corresponding stress states. Social factors

Taking the user labeled with a red star in Fig. 3. We extract attributes from each and every tweet of the user to form tweet-level attributes as shown in the cylinders. Different colors represent different modalities and blank represents modalities that are not available in the tweet. The tweet-level attributes in the cylinder are fed to cross auto-encoders. The CAEs are embedded in a CNN [2] that will integrate attributes from CAE to the aggregated user-level content attributes by pooling each attribute map. The user-level posting behavior attributes, user level content attributes, and user-level social interaction attributes together form the user-level attributes.

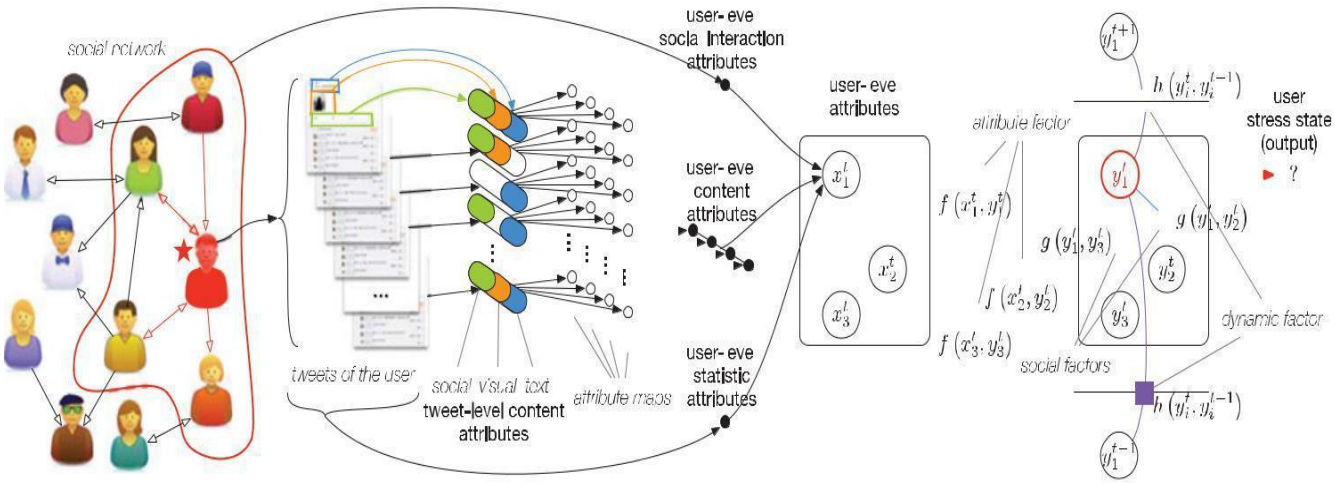


Fig. 3. Architecture of stress level detection in social media interaction

6. EXPERIMENTAL SETUP:

In the following experiments, we first train and test our model on the large-scale FACEBOOK dataset DB1. We then test our model on the other 3 datasets to show effectiveness of the model on different data sources or varieties ground truth labeling methods. For all of our analysis, we use FIVE fold cross validation with over TEN randomized experimental runs. Comparison

Methods: We compare the following classification methods for user-level psychological stress detection with our FGM+CNN model Logistic Regression (LRC): it trains a logistic regression classification model and then predicts users' labels in the test set.

Support Vector Machine (SVM) [5]: it is a binary classifier that proved to be effective on a huge category of classification problems. we use SVM with RBF kernel.

Random Forest (RF) [4]: it is an ensemble learning method for decision trees by building a set of decision trees with random subsets of attributes and bag-ging them for classification results.

Gradient Boosted Decision Tree (GBDT): it trains a gradient boosted decision tree model with features associated with user. Deep Neural Network (DNN) for user-level stress detection: it is proposed to deal with the problem of user-level stress detection problem with a convolutional neural network (CNN) with cross auto en-coders. This is the real baseline method that we can compare the proposed model. We employ scikit-learn[10] for the above methods for a fully investigation of the proposed methods, we consider the Effectiveness: We evaluate the performance of our model and comparison methods in terms of Accuracy (Acc.), Recall (Rec.), Precision (Prec.)

and F1-Measure. Efficiency. We can evaluate efficiency of the methods by comparing the CPU time of training. Experiments are performed on an x64 machine with 2.9 GHz Intel Core i7 CPU and 8 GB RAM.

7. Strong/Weak Tie:

Strong/Weak Tie [1] is one of the most basic principles in social network theories. We classify the constructed social relationships into strong or weak ties by the number of times that two users interact with each other via comment, @-mention, retweet, or like in a week. In our work, we tried different values for the threshold and finally chose three by cross-validation. If two users interact with each other more than three times, we call the relationship a strong tie, and otherwise a weak tie. This definition of user ties is adopted as the standard treatment in the research of social network analysis [7], so as to capture the most recent user relation-ships in a shifting environment. Fig. 8b illustrates the results. We can see that strong ties indeed have strong influence on users' stress states, and the influence of weak ties is relatively weak. For example, when a user has three stressed strong-tie connections, the probability that the user will become stressed increases to 13 percent, more than twice as high as for a user with three stressed weak-tie connections.

1) Summary: Based on the experimental results and analyses we know that: users' stress states are not only revealed in their own tweets, but also affected by the contents of their social interactions, including commenting on and re-tweeting others' tweets; and 2) users' stress states are revealed by the structure of their social interactions, including structural diversity, social influence, and strong/weak ties. These insights

quantitatively prove the necessity and effectiveness of combining social interactions for stress detection.

8 CONCLUSION

In this paper, we presented a framework for detecting users' psychological stress states from users' weekly social media data, leveraging tweets' content as well as users' social interactions. Employing real-world social media data as the basis, we studied the correlation between user' psychological stress states and their social interaction behaviors. To fully leverage both content and social interaction information of users' tweets, we proposed a hybrid model which combines the factor graph model (FGM) with a convolutional neural network (CNN).

In this work, we also discovered several intriguing phenomena of stress. We found that the number of social structures of sparse connection (i.e., with no delta connections) of stressed users is around 14 percent higher than that of non-stressed users, indicating that the social structure of stressed users' friends tend to be less connected and less complicated than that of non-stressed users. These phenomena could be useful references for future related studies.

ACKNOWLEDGMENT

This project is supported by Technical Research and Development Program Sponsored by karamveer dadasaheb kannamwar college of Engineering and Technology, Nagpur. Where the project is developed under the guidance of DR S.P. KHANDAIT.

REFERENCES

[1] A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi, and A. Pentland, "Daily stress recognition from mobile phone data, weather condi-tions and individual traits," in Proc. ACM Int. Conf. Multimedia, 2014, pp. 477–486.

[2] C. Buckley and E. M. Voorhees, "Retrieval evaluation with incom-plete information," in Proc. 27th Annu. Int. ACM SIGIR Conf. Res. Development Inf. Retrieval, 2004, pp. 25–32.

[3] X. Chang, Y. Yang, A. G. Hauptmann, E. P. Xing, and Y.-L. Yu, "Semantic concept discovery for large-scale zero-shot event detection," in Proc. Int. Joint Conf. Artif. Intell., 2015, pp. 2234–2240.

[4] W. Che, Z. Li, and T. Liu, "Ltp: A chinese language technology platform," inProc. Int. Conf. Comput. Linguistics, 2010, pp. 13–16.

[5] C. C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines, "ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 389–396, 2001.

[6] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in Proc. Int. Joint Conf. Artif. Intell., 2010, pp. 1237–1242.

[7] S. Cohen and A. W. Thomas, "Stress, social support, and the buffering hypothesis," Psychological Bulletin, vol. 98, no. 2, pp. 310–357, 1985.

[8] G. Coppersmith, C. Harman, and M. Dredze, "Measuring post traumatic stress disorder in twitter," in Proc. Int. Conf. Weblogs Soc. Media, 2014, pp. 579–582.

[9] R. Fan, J. Zhao, Y. Chen, and K. Xu, "Anger is more influential than joy: Sentiment correlation in weibo," PLoS One, vol. 9, 2014, Art. no. e100184.

[10] Z. Fang, et al., "Modeling paying behavior in game social networks," in Proc. 23rd Conf. Inform Knowl. Manag., 2014, pp. 410–420.