International Journal of Engineering and Technology, Vol. 11, No. 3, June 2019

Influence Calculation Model of Microblog User Based on Content, Emotion and Users' Behavior

Fu Xie, Xueying Sun, and Fengming Liu

***Abstract*—Highly influential social users can guide public opinion and influence their emotional venting. Therefore, it is of great significance to identify high-impact users effectively. This paper starts with the users' text content, users' emotions, and fans' behaviors. It combines the amount of information in the content and sentiment tendency with the fans' forwarding, commenting, and Liking actions. And based on the principle of the three-degree influence, the users' influence calculation model is constructed. Finally, the experimental results show that the three-degree force calculation model is more accurate and effective than other similar models.**

***Index Terms*—Three-degree, microblog, user influence, social networks.**

I. INTRODUCTION

According to the statistics of the 40th Statistical Report on Internet Development in China, as of June 2017, the number of Chinese citizens reached 751 million, the Internet penetration rate reached 54.3%, and the scale of mobile Internet users reached 724 million. Social networking platforms based on mobile Internet have infiltrated into all aspects of people's lives.

Taking the Microblog social networking platform as an example, by September 2017, as shown in Fig. 1, the number of monthly active users of Microblog was 376 million, increased 27% compared with the same period in 2016, of which China' s mobile terminal accounted for 92% and the daily active users reached 165 million, an increase of 25% over the same period of last year. If we can find influential nodes in the network quickly and effectively, it will be better for the government to monitor public opinion and guide them in time. For the business community, it is possible to achieve targeted commercial promotion, advertising marketing, etc. according to the influence nodes in the network. Therefore, the research on the analysis, measurement, modeling, and dissemination of node influence in social networks has a very important theoretical and practical values.

According to the three-degree influence theory, this paper constructed a computing model based on users' blog content and fans' behavior. Firstly, calculate the content of the blog post and measure the amount of content information through the information entropy. Secondly, analyze and calculate the sentiment tendencies of the blog post. In addition, calculate the "forward-comment-like" influence of the user' s three-tier. Finally, verify it by crawling data. Results showed that the three-degree influence calculation model proposed in this paper is more accurate and effective than other similar models.

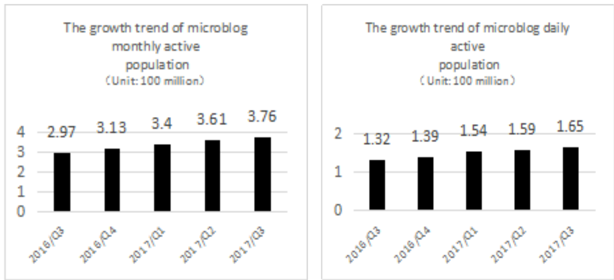


Fig. 1. The statistics of Microblog.

II. RELATED RESEARCH AT HOME AND ABROAD

Research on influence attracted the attention of scholars from all walks of life in the early 20th century. When Triplett [1] studied social promotion theory, he discovered that when a person received attention from others, his/her performance would become more prominent. By the 1950s, Kate *et al.* [2] found that influence played a very important role in daily life or in political elections. In recent years, with the rise of large-scale social networks such as Microblog, various theories have shown that the distance between people has become shorter and shorter, and the links have become closer and more influential.

Zhang Jing and Tang Jie [3] pointed out three aspects of influence, the first is the individual's influence on one user only, means the ability of affecting other users and the probability of being affected by others; the second one was the influence between the two users of A and B, means A influences B and B influences A; the third is the group influence, which is the most complicated, refers to the influence of a group of multiple users on one user. Shi Cong *et al.* [4] integrated user behaviors and blog post content and evaluated the predicted users and the influence of the issued microblog through the extraction of behavioral characteristics and content characteristics such as the users' blog posting time and "forward-like" behavior. Chen Zhenchun *et al.* [5] analyzed users' influence in terms of user behavior and community structure. First, it divided the network into groups. Then, the users' influence was measured from the number of fans of the user and the quality of the fans and the number of communities the user crosses. The authors in [6] evaluated users' influence from two aspects of user behavior and user relationship. First, analyze users' blog post content. Second, improved based on the PageRank algorithm, and calculates activeness for each node. Finally calculate the BRR (Behavior-Relationship Rank) of the node. and sort the nodes

Manuscript submitted December 6, 2018; revised April 12, 2019.
The authors are with th Shandong Normal University, China (e-mail: liufm69@163.com).

DOI: 10.7763/IJET.2019.V11.1149

based on their value i.e. influence. Xu Danqing *et al.* [7] proposed a new social influence model, the PTIM model, which was an iterative combination of the characteristics of the user' s fans and small-world attributes and has achieved relatively good results for all evaluation indicators. Fowler *et al.* [8] found that the happiness between people could reach a three-degree separation by studying the spread of happiness in social networks, that is the principle of the three-degree influence. The node affects the neighboring node, this is the first-degree effect; it can also affect the neighboring node's neighboring node, which is the second-degree effect; and it also affects the neighboring node' s neighboring node of the neighboring node, which is the third-degree effect. The study found that the effect diminished with the separation of time and geography, and the influence beyond three degrees was negligible. In the literature [9], [10], through real data experiments on Facebook and SNS, it pointed out that both the intensity of friendship between friends or the intimacy between users have a great impact on user behavior. Through the interaction intensity and intimacy analysis, a richer social graph can be obtained. The above studies on influence started from different aspects and can be roughly divided into three aspects: user relations, behavior, and content of blog post. The dissemination of information was interlocking and complex in the network. As above mentioned [5], attention was paid to the quality of the users' fans. However, the dissemination of information is not merely a forwarding level and can be spread by multiple users.

III. The Three-Degree Calculation Model of Users' Influence Math

A. The Calculation of First-Degree Influence Based on the Content of Information

The more content a microblog user publishes, the greater the amount of information they contain and they are more likely to be forwarded. This article used information entropy to represent how much information the user content contained. For the analysis of the content of the blog post, this paper used the NLPIR semantic analysis system. NLPIR (Natural Language Processing and Information Retrieval Sharing Platform) is a word segmentation system developed by Dr. Zhang Huaping. The more information in the post content, the higher the complexity, the greater the entropy value, and the greater the probability of being followed, browsed, and forwarded. The keyword collection function of the NLPIR semantic analysis system can get the keyword set K, and the keywords' weight in the content also can be obtained.

$$P_i = \frac{W_i}{\sum_{i\in k}^{n} W_i} \tag{1}$$

where W_i refers to the weight of the extracted keywords in the content; and P_i refers to the probability that the *i-th* keyword appears in the content. The entropy [11] is calculated as follows:

$$E = -\sum_{i\in k} P_i \cdot \log P_i \tag{2}$$

While analyzing the content, we need to consider the topic degree of the blog post, whether the content fits the hot topic. Calculate the similarity of blog post content and topics. For text similarity, we use the cosine similarity to calculate the similarity of blog content and hot topics in this paper.

$$Sim = (content, hottopic) = \cos(K, H) = \frac{\sum_{1}^{n}(K_i \times H_i)}{\sqrt{\sum_{1}^{n} K_i^2} \times \sqrt{\sum_{1}^{n} H_i^2}} \tag{3}$$

In the formula, K_i, H_i refer to the weights of features extracted from the blog post content and feature weights of the current hot topics.

Finally, the influence of the post content can be expressed as:

$$TI = \begin{cases} E(1+Sim(content, hottopic)), Sim>0 \\ E, Sim=0 \end{cases} \tag{4}$$

B. The Calculation of Second-Degree Influence Based on Bowen's Emotion

Microblog posts not only contain the views and opinions of users, but also embody their emotions. Therefore, the post's emotions also play a crucial role in the user influence. The study of text sentiment tendencies can be mainly divided into two aspects: the method based on dependency syntax analysis [12-14] and the method based on the term bag model [15-17]. Referring to the methods of these two aspects, the rules for calculating the sentiment orientation of sentences are given below.

Calculation rules 1. Emotional calculation of emotional words. The emotional word is recorded as "word", emotional intensity is "*EIntensity*", emotional positive and negative polarity is "*EPolarity*". the calculation is defined as follows:

$$Value(word) = \theta \times EIntersity(word) \times EPolarity(word), \theta \in R \tag{5}$$

EIntensity, the emotional intensity, is the original part of speech, and refers to the emotional tendency of the word itself such as "praise", "pampering" which has a clear emotional tendency; *EPolarity*, the positive and negative polarity of emotion is the positive and negative emotional tendency of the word. θ refers to the harmonic coefficient which is the modified polarity of the word, and it is expressed as an adverb of degree such as "100%" and "compared". For example, "100% good" and "better" are positive words, "100%" is greater than "more", and the emotional value of two words is different, so that the "good" emotional value is +1. The intensity adverbs of degree adverbs "100%" and "comparative" are +3 and +1. Then, the values of "100% good" and "good" affective values are +3 and +1.

Calculation rule 2. Sentence calculation rules. V is the tendency value of emotional words, and its definition is as follows:

$$Value(sentence) = \sum_{i=1}^{m} V(word) \tag{6}$$

Among them, m refers to the number of emotional words in a sentence, and the emotional tendency of the entire sentence can be obtained by adding up the calculated emotional tendencies of each emotional word.

Calculation rule 3. The emotional value of blog post content is defined as follows:

$$Value(text) = \frac{1}{n-m} \sum_{i=1}^{n} V_i \times W_i \tag{7}$$

Among them, *n* is the total number of emotional sentences in the blog, and m is the number of sentences in the blog that do not contain emotions. V_i refers to the emotional value of the *i-th* sentence, and W_i refers to the weight of the *i-th* emotional sentence in the blog post.

The above calculation can divide the blog into three levels.

$$\begin{cases} Value > posThreshold \text{ , Positive Emotion} \\ negThreshold < Value < posThreshold \text{ , neutral attitude} \\ Value < negThreshold \text{ , Negative Emotion} \end{cases} \tag{8}$$

PosThreshold and *negThreshold* are the propensity thresholds for positive emotions and negative emotions, respectively.

C. The Calculation of Third- degree Influence Based on Fan Behavior

Every user in the Microblog network is viewed as a node, and some kind of connection exists among users as a point-to-point connection. This paper used user behavior to hierarchically divide the network. As shown in Fig. 2, let V_1 be the central node and V_1 as the center to extend outward to simulate the spread of the messages. V_2, V_3, V_4 are neighbors of V_1 (one-degree); V_5, V_6, V_8 are neighbors of V_1 neighbor (second-degree); V_7, V_9 are the neighbors (third-degree) of neighbors of V_1 neighbors, which constitutes a simple three-degree social network structure diagram.

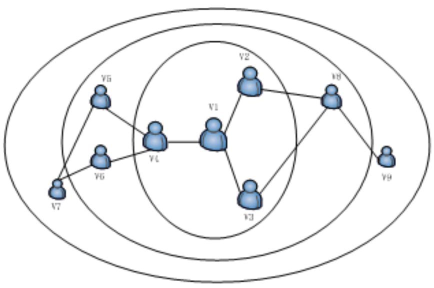


Fig. 2. Three-degree social network structure.

The behavior of fan users on a microblog mainly includes three types: forwarding, liking, and commenting. We use *BI* to represent behavior-influence, the expression is as follows:

$$BI = U\,(FI, CI, LI) \tag{9}$$

FI indicates the forwarding influence, which is how many people forwarded the blog post. *CI* indicates the influence of the comment, i.e., how many people have commented on the blog post. *LI* is the influence of praise, which is how many people like it.

User *v* is the fan of *u*, then *v*'s attention to *u*'s microblogs can be measured by the time difference between *u* publishing blog posts and *v*'s behavior. Assume that the blog post set posted by *u* is M_u and the time set is T_u, the blog set forwarded, commented and liked by *v* is M_v and the time set is T_v. Calculate the time difference of all the actions, and find the mean value. Where the smaller the mean value, the greater attention that *v* represents for *u*, the greater the likelihood of *v*'s forwarding. And calculate its variance S^2. The smaller the fluctuation of the variance, the higher the reliability of the mean, and the more stable the attention of *v* to *u*.

The specific calculation is as follows:

$$M = \frac{\sum (T_v - T_u)}{N_{vu}} \tag{10}$$

where *M* is the mean of all time differences; $(T_v - T_u)$ is the time difference for a blog post; $\sum (T_v - T_u)$ is the sum of all time differences; and N_{vu}, which is how much time difference there is.

$$S^2 = \frac{\sum (T_v - M)^2}{N_{vu}} \tag{11}$$

S^2 is the variance of the time difference, and the smaller the variance, the more stable the attention degree. Finally, the variance is translated into the probability that the fan produces behavior. Based on the variance, the frequency of the behavior that generates the behavior should also be taken into consideration.

$$F = \begin{cases} \frac{n}{N}(1 - 0.01S^2), S^2 > 0 \\ 1, S^2 = 0 \end{cases} \tag{12}$$

where *n* is the total number of fans' actions. *N* is the total number of users' microblogs.

Therefore, the influences of forwarding, commenting, and liking are calculated as follows, respectively:

$$FI = \sum_{v \in Fan(u)} F_{vu} \tag{13}$$

$$CI = \sum_{v \in Fan(u)} F_{vu} \tag{14}$$

$$LI = \sum_{v \in Fan(u)} F_{vu} \tag{15}$$

We introduced α, β, γ, three parameters that represent the weights of the three types of behaviors, *BI* was calculated as follows:

$$BI = \alpha FI + \beta CI + \gamma LI \quad (16)$$

The principle of the three-degree influence showed that information was attenuated during propagation, so we introduced an attenuation factor ω, and as the distance increased, the value of the factor will decrease. For content information, emotions, and users' behavior, weighted fusion was based on their relationship. Calculated as follows:

$$SBI = a \cdot (TI \times Value) + b \cdot (\omega_1 BI_1 + \omega_2 BI_2 + \omega_3 BI_3) \quad (17)$$

SBI represents the sum of users' three-degree influence, which is users' influence; *TI* is the content influence; *Value* refers to the emotional tendency value; BI_1, BI_2, BI_3 represent the influence of each layer, respectively; and a and b are the weights of the evaluation indicators.

IV. Experiments

A. Data Collection

Through the Python crawler, the real data set obtained from the API interface provided by the Weibo platform, was to verify the calculation model proposed in this paper. This article has crawled 946 microblogs from the official website of Shandong Normal University from November 2017 to the end of February 2018. There were 66,617 three-layer forwards, comments, and likes, including 5931 pieces of forwarding data, 12,065 pieces of review data, and 48,675 pieces of like data. In addition, we also got the relevant information from 500 hot topics including attributes as content, the number of fans, reading volume, and discussion volume. Before data is used, we has removed the inactive users.

B. Data Processing

For the analysis of the content of the blog, we calculated the content information of the blog based on the NLPIR semantic analysis system of the Chinese Academy of Sciences, and calculated the emotional tendency value according to the sentiment calculation rules given above. When calculating fans' attention, we first counted the nicknames of the fans under Microblog, the time when they acted, and the number of microblogs each person interacted with, selected the appropriate time unit, and then got the fans' attention.

Fig. 3 shows attention of some fans. The center point of the circle is 0, and it spreads out. The outermost layer is 1. We could see that most of the fans' attention was relatively low, and there were a few fans with high attention.

Fig. 3. Attention.

Three weights α, β, γ were introduced when we calculated each layer of users' influence. They represent the proportions of the three types of actions: forward, comment, and like. To calculate three weights, we used the analytic hierarchy process to get $\alpha = 0.539$, $\beta = 0.297$, $\gamma = 0.164$, as Table I. For the introduced attenuation factor, we obtained that ω_1=0.136, ω_2=0.098, ω_3=0.024 refers to literature [8], [18]. According to subjective evaluation method and reference [4], [5], a =0.3, b = 0.7.

TABLE I: The Value of Α, Β, Γ

	Like(γ)	Comment(β)	Forward(α)
Like(γ)	1	2	3
Comment(β)	1/2	1	2
Forward(α)	1/3	1/2	1

C. Algorithm Comparison Evaluation Index

The evaluation index adopted in this paper was the coverage rate [19], which was the ratio of the number of nodes affected by the selected user to the total number of nodes in the microblog. The greater the coverage rate, the wider the range of influence that the users spread, and the more people are affected, the greater the influence.

$$Coverage = \frac{The\ number\ of\ affected\ nodes}{Total\ number\ of\ nodes} \times 100\% \quad (18)$$

D. Compare

The purpose of the experiment was to prove that the algorithm proposed in this paper had a higher accuracy, so compared it with the PageRank algorithm and BRank algorithm [20], which only considered user behavior and the RBRank algorithm [21] which considered user relationship behavior. The result is shown in Table II.

From Table II, we can see that the algorithm proposed in this paper improved the accuracy, but the difference of results was not very obvious. Analyzing the reasons, we believed that Shandong Normal University was the school portal, and the content and emotions conveyed were mainly positive. The fans were mainly students. The composition and structure were unitary, and the influence of fans was limited. Therefore, we needed to select other users with rich attributes to conduct the data acquisition and analysis again.

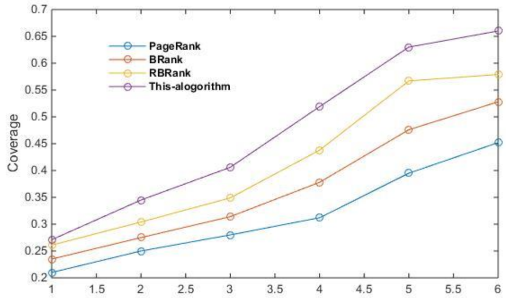


Fig. 4. Comparison between this algorithm and other algorithms.

In the next experiment, we selected Fan Bingbing as a keyword and crawled its data for the past three months. The crawled attributes still included fans' nicknames, profiles, ID, followers, fans, microblogs, blog posts, and release dates. The number of forwarding, comments, number of likes, forwarders nickname, profile, ID, content of forwarded blog, forwarding time, totaled 280,460 items. According to the results of this algorithm, the comparison results are shown in Table III.

TABLE II: COMPARISON BETWEEN THIS ALGORITHM AND OTHER ALGORITHMS

Compare	Algorithm			
	PageRank	BRank	RBRank	This Algorithm
Coverage	20.9%	23.5%	26.1%	27.1%

TABLE III: COMPARISON BETWEEN THIS ALGORITHM AND OTHER ALGORITHMS

Compare	Algorithm			
	PageRank	BRank	RBRank	This Algorithm
Coverage	39.5%	47.6%	56.7%	63%

From Fig. 4, we can see that the RBRank algorithm and the algorithm of this paper had larger influence values than the PageRank and BRank algorithms. It explains that from multiple angles, selecting multiple eigenvalues in combination to evaluate users' influence is much more advantageous than simply calculating users' influence from one aspect. In addition, through the comparison, we found that the coverage rate of the proposed algorithm was the largest, indicating the best effect. Therefore, we reached a conclusion the proposed algorithm considered the users' sentiment content and emotional tendencies as well as the characteristics of the user behavior, expanding the depth of calculations, and calculating the three-degree influence was better.

V. CONCLUSION

This paper proposed a calculation method based on the users' text content and users' behavior. Through the evaluation of content information and emotional tendency, the overall statistics on the users' forwarding, comments, likes, we constructed a users' influence computing model and improved accuracy. From the above experiment, we could see that selecting different experimental objects would have different effects, and the effect was more significant when the experimental objects were rich in attributes. Therefore, in future work, we will consider adding users' attributes to the calculation model to make the analysis of influence more perfect. Moreover, the cross relationship between fans is not considered in text, we should also be noted in the future work.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (No. 71701115), the National Social Science Foundation of China (No. 14BTQ049), the Shandong Natural Science Foundation(ZR2017MF058), and Special project for Internet development of social science planning special program of Shandong province(17CHLJ23) , Social Science Planning Program of Shandong Province(13BTQJ01).

REFERENCES

[1] N. Triplett, "The dynamogenic factors in pacemaking and competition," *American Journal of Psychology*, vol. 9, no. 4, pp. 507-533 ,1970.

[2] E. Katz and P. F. Lazarsfeld, "Personal influence: the part played by people in the flow of mass communications," *American Journal of Sociology*, vol. 21, no. 6, pp. 1583-1583, 1955.

[3] J. Zhang and J. Tang, "A review of social influence analysis," *Chinese Science: Information Science*, vol. 47, no. 8, pp. 967-979, 2017.

[4] C. Shi, J. Tang, and Y. Hu, "Predicting microblog user influence based on user behavior and blog content," *Journal of Chinese Computer Systems*, vol. 38, no. 7, pp. 1495-1500, 2017.

[5] Z. C. Chen, X. J. Liu, and B. Li, "Analysis of influence of microblog user communication based on behavior and community," *Application Research of Computers*, vol. 35, no. 7, pp.2075-2078, 2018.

[6] S. Kang, C. Zhang, Z. Lin *et al.*, "Complexity research of massively microblogging based on human behaviors," in *Proc. International Workshop on Database Technology and Applications*, Wuhan, China, pp. 27-28, 2010.

[7] D. Q. Xu, Y. Q. Liu, M. Zhang *et al*., "Research on user influence based on online social network," *Chinese Journal of Information*, vol. 30, no. 2, pp. 83-89, 2016.

[8] H. F. James and A. C. Nicholas, "Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the framingham heart study," *British Medical Journal*, vol. 338, no. 7685, pp. 23-27, Jan.3, 2009.

[9] H. J. Ilić, L. Pintar *et al*., "Proof of concept for comparison and classification of online social network friends based on tie strength calculation model," *ICIST,* 2016.

[10] W. Al-Ghaith, "Understanding social network usage: Impact of co-presence, intimacy, and immediacy," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 8, 2015.

[11] R. Li, H. P. Zhang, Y. P. Zhao, and J. Y. Shang, "Research on Chinese document automatic abstracting technology based on topic model and information entropy," *Computer Science*, vol. 41, no. 2, pp. 298-332, 2014.

[12] S. Feng *et al*., "Analysis of bowen affective tendency based on dependency syntax," *Computer Research and Development*, vol. 49, no. 11, pp. 2395-2406, 2012.

[13] F. Liu, L. Wang *et al.*, "A Web Service trust evaluation model based on small-world networks," *Knowledge-Based Systems*, vol. 57, no. 2, pp. 161-167, 2014.

[14] M. Hu and T. Yao, "Chinese micro blog view sentence recognition and evaluation object extraction method," *Journal of Shandong University*, vol. 51, no. 7, pp. 81-89, 2016.

[15] F. Liu, L. Wang *et al.*, "Analysis of network trust dynamics based on evolutionary game," *Scientia Iranica, Transaction E: Industrial Engineering*, vol. 22, no. 6, pp. 2548-2557, 2015.

[16] G. Paltoglou and M. Thelwall, "A study of information retrieval weighting schemes for sentiment analysis," in *Proc. Meeting of the Association for Computational Linguistics*, Uppsala, Sweden. DBLP, pp. 1386-1395, 2010.

[17] H. Zhang, H. Li, and Q. Li, "Research on automatic calculation algorithm of emotional word discovery and polarity weight," *Journal of Chinese Information Processing*, vol. 31, no. 3, pp. 48-54, 2017.

[18] J. Zhao *et al.*, "The evaluation method of node influence based on the three-degree theory," *Journal of Fuyang Teachers College*, vol. 33, no. 4, pp.78-82, 2016.

[19] W. Hou, Y. Huang, and K. Zhang, "Research of micro-blog diffusion effect based on analysis of retweet behavior," in *Proc. International Conference on Cognitive Informatics and Cognitive Computing*, IEEE, pp. 255-261, 2015.

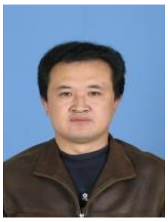
[20] X. Li *et al*., "Novel user influence measurement based on user interaction in blog," *proc. International Conference on Advances in Social Networks Analysis and Mining*, IEEE, pp. 615-619, 2013.

[21] Y. Huang and L. Li, "Analysis of user influence in social network based on behavior and relationship," in *Proc. International Conference on Measurement, Information and Control*, Harbin, pp. 682–686, August 2013.

Fu Xie was born on July 9, 1971 in Laixi, Shandong, China. He was graduated from Shandong Normal University in December 2014 with a Ph.D. in management science and engineering.

He is currently an associate professor at the School of Information Science and Engineering at Shandong Normal University. His research interests include information security, cloud computing, and data science.

Xueying Sun was born on November 28, 1993 in Liaocheng, Shandong. Now she is a master's student, studying at Shandong Normal University. Her research interests include social networks and data science.



Fengming Liu is the professor of Business school at Shandong Normal University in the China. He received a B.Sc. and a M.Sc. both from Shandong Normal University, China, and a Ph.D. from the Donghua University of China. His research interests include information management, social networks, and game theory.