Identifing Influential Users in an Online Healthcare Social Network

Xuning Tang
College of Information Science and Technology
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104
xt24@drexel.edu

Christopher C. Yang
College of Information Science and Technology
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104
Chris.yang@drexel.edu

Abstract— **As an important information portal, online healthcare forum are playing an increasingly crucial role in disseminating information and offering support to people. It connects people with the leading medical experts and others who have similar experiences. During an epidemic outbreak, such as H1N1, it is critical for the health department to understand how the public is responding to the ongoing pandemic, which has a great impact on the social stability. In this case, identifying influential users in the online healthcare forum and tracking the information spreading in such online community can be an effective way to understand the public reaction toward the disease. In this paper, we propose a framework to monitor and identify influential users from online healthcare forum. We first develop a mechanism to identify and construct social networks from the discussion board of an online healthcare forum. We propose the UserRank algorithm which combines link analysis and content analysis techniques to identify influential users. We have also conducted an experiment to evaluate our approach on the Swine Flu forum which is a sub-community of a popular online healthcare community, MedHelp (www.medhelp.org). Experimental results show that our technique outperforms PageRank, in-degree and out-degree centrality in identifying influential user from an online healthcare forum.**

I. INTRODUCTION

Since the H1N1 flu (Swine Flu) pandemic began in late April of 2009, millions of people in the United States have been infected. At least 20,000 people were hospitalized and more than 1,000 people died. The flu continued to evolve and spreads rapidly across the nation. On October 24th 2009, U.S. President Barack Obama signed a proclamation declaring a national emergency, which indicated that the disaster of H1NI flu had already been a nationwide security issue threatening the national stability. An investigation of the public reaction toward this pandemic became increasingly important. A better understanding of the communication network of people during this critical time is useful. It helps to identify those people who are creating rumors and causing fear in the society. On the other hand, we can also find out those people with certain influence in the community and take advantage of their influence to propagate accurate and positive health information in order to ensure the social stability. Given the advance of online social networking technologies, we may track the discussions of the online healthcare forum to understand the influence and flow of information related to the epidemic outbreak.

Due to the rapid advancement of social media technology, online social networking communities such as Facebook, MySpace, and Twitter is playing an important role in our daily life. These social media platforms have the most prevalent channel for people to express their opinions and emotions and interact with million other users socially around the world. In other words, the emergence of online social networking community provides a virtual environment for people to interact with each other through the internet, which is also a digital footprint of the social conversations and activities. Nowadays, with the information stored in the online social networking community, we can extract the data to identify those people who are making use of their influence to spread inaccurate and unhealthy information. We can also identify some other influential users who provide positive and correct information to the community. In the healthcare domain, it is particularly important to ensure that the online healthcare social networking is utilized to promote the positive health information and improve the health outcome of the general public rather than spreading out inaccurate information that may threatening the public safety. Therefore, there is a desire for developing methods and techniques to analyze and quantify the user influence in the online social networking environment.

In this work, we purpose a novel hybrid model to quantify user influence within an online social networking community. Our test bed is the Swine Flu online forum which is a sub-community of the online healthcare social networking community, MedHelp (www.medhelp.org). MedHelp is a healthcare information site that connects regular people with medical experts, and other patients to provide them with aid and support. Founded in February 1994, MedHelp has become one of most important online healthcare social networking community over the world. As an important information portal, users of MedHelp actively post and follow threads which makes a significant impact on their knowledge and attitude toward the H1N1 influenza. Our goal is to quantify and compute the user influence within the Swine Flu forum automatically. We develop a mechanism to extract the social network from the online community and propose a hybrid model, which take link and content information into account, to quantify user influence within the forum automatically. Experimental result shows that our proposed technique is promising to identify influential users and it also outperforms some other techniques such as degree centrality and PageRank.

The paper is organized as follows. In the next section, we discuss the relevant work. We provide an overview of the MedHelp community in Section 3. In Section 4, we introduce the mechanism of constructing social network according to data features and purpose our UserRank algorithm. Experimental results are presented in Section 5.

II. RELATED WORK

The study of influence spreading through a social network has been conducted in the field of social sciences for many decades. The main purpose of these studies is trying to understand how the dynamics of adoption are likely to unfold within the underlying social network. Some earliest research focused on the adoption of medical [3] and agricultural innovations [14]. Some other later research investigated the diffusion process of "word-of-mouth" effect and its application to the viral marketing [1, 7, 8, 11]. The above works focus on studying the network properties and their impact on the interaction among social actors, but did not identify influential users within a social network.

Using a data mining approach, Domingos and Richardson [5, 6] purposed an algorithm for solving the influence maximization problem. In their work, they considered the problem as a probabilistic model. Given a social network, assuming that the seller is going to give away n copies of an item, the research problem is identifying a subset of n nodes in the network such that the subsequent adoption of the good is maximized. Heuristic algorithms were used to choose customers with large influence on the network. Kempe, Kleinberg, and Tardos [10] formulated the problem as a discrete optimization problem. They proved that the optimization problem is NP-hard, and presented a greedy approximation algorithm applicable to different models, which guarantees that the influence spread is within (1-1/e) of the optimal influence spread. In a recent work, Chen, Wang and Yang [2] proposed new heuristics that have influence spreads close to the greedy algorithm while running at more than six orders of magnitude faster than the greedy algorithm. The above work proposed techniques to retrieve seed users who can maximize the information propagation within a social network. However, the social network topology is the only factor taking into account by these techniques. The content feature is not considered in computing user influence.

Other than modeling the problem as an optimization model, link analysis techniques have also been applied to identify reputational users in a community. Zhang, Ackerman and Adamic [15] evaluated link analysis algorithms such as PageRank or HITS for expert finding in a closed domain. In their work, they assumed that the importance of web pages was similar to the user influence in social network. However, their result showed that a complex algorithm, such as PageRank, was not performing as good as relatively simple measures such centrality measure for finding experts. Tang, Sun, Wang and Yang [13] proposed a Topical Affinity Propagation (TAP) approach and modeled the problem of identifying experts by a graphical probabilistic model. Goyal, Bonchi and Lakshmanan [9] considered an alternative definition of influential users based on frequent patterns. They claimed that an action propagated from one user to another if and only if one user conducted this action ahead of another user. They developed an algorithm to extract the leaders who frequently conducted some actions ahead of a group of users based on frequent pattern discovery. Although their algorithm is efficient and effective to extract group leaders, their definition of group leader may not be applicable to many online social networking platforms. A user in a social network may conduct an action for many different reasons but not necessarily following an action made by another user. As a result, it is not always appropriate to claim that one user is leading another user only because one conducts an action ahead of another. In our approach, we take both link analysis and content analysis into consideration. We consider both message similarity and response immediacy which yield higher accuracy of ranking influential users in our proposed UserRank algorithm.

III. OVERVIEW OF MEDHELP

A. Online Healthcare Community

MedHelp is one of the largest online healthcare communities in the world. MedHelp community consists of 189 Healthcare Support Forums and 101 "Ask a Doctor" Forums covering different diseases. The total number of forum in MedHelp is continuously increasing. Since its opening in 1994, nearly 3 million threads are posted in the community. It also attracts over 8 million visitors every month. There are several notable characteristics of MedHelp. For examples, users can share information and seek help on discussion board, publish journals in their personal pages and posting notes to each other's personal pages. As a result, MedHelp is an informative healthcare forum as well as an interactive social networking site.

Once a user registers to the MedHelp community, s/he is enabled to edit a personalized page called "My MedHelp". The user can publish journal on his/her own page or write comment to any journal published by other users in MedHelp community. The user can join and follow any forum within the MedHelp community. Each forum focuses on a specific theme, which is a disease such as Swine Flu or Cancer. Once a user joins a forum, such as Swine Flu forum, s/he becomes a member of this forum. There is no limit on the number of forum users. No matter if a user is a member of a specific forum, as long as s/he has already registered to the MedHelp Community, s/he can post a new thread or reply to any thread on the discussion board of the forum of interest. The user can invite as many friends as s/he likes. The friendship is an asymmetric relationship in this community, which means when B adds C to his/her friend list, B will not exist in C's friend list unless C explicitly adds B to his/her friend list. Users can also post notes to each other's personal pages regardless of the friendship relationship.

B. Data Crawling

To identify the influential users within the Swine Flu forum, we developed a robot to crawl the forum data from MedHelp and saved the formatted data in our database. We narrowed down our research focus on threads and their replies in the discussion board of Swine Flu forum. Figure 1 shows the structure of a thread and its corresponding replies. For each thread within the Swine Flu forum, we record the user identity who publishes the thread, the timestamp and content of the thread, the user identities who have replied to the thread and the user identities to whom they have replied to, and the timestamp and content of the replying messages.

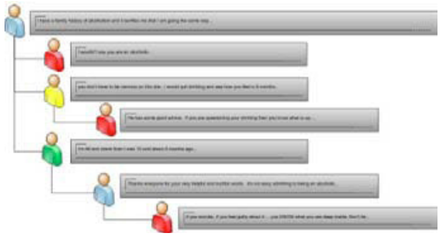

Figure 1. Structure of Thread and Replies

IV. METHODOLOGY

In this section, we first introduce the framework for identifying influential users in a given online healthcare support forum. Secondly, we describe how to construct social networks based on collected data. At last, we introduce the UserRank algorithm to compute user influence within the online social network.

A. Framework for Identifying Influential Users in Online Forum

Figure 2 depicts the framework for identifying influential users in an online healthcare support forum. The framework consists of two phases: forum data collection and social network analysis.

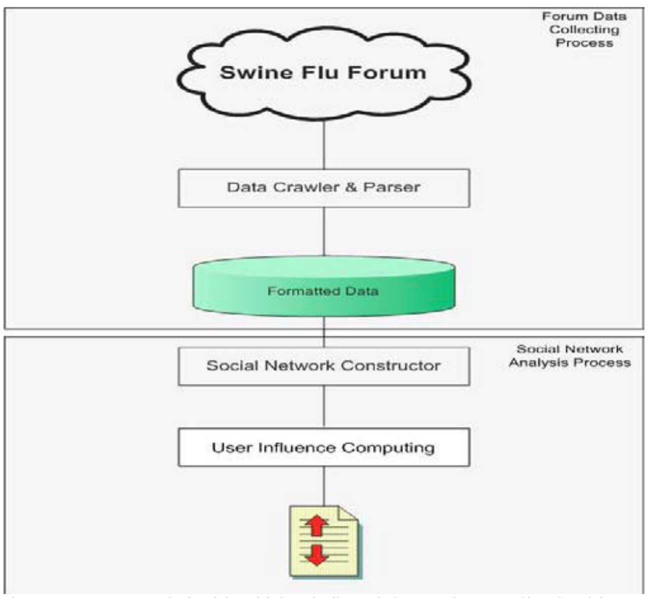

Figure 2. Framework for identifying influential users in an online healthcare support forum

In the forum data collection phases, we built a crawler which collect all the posts and replies on the discussion board of Swine Flu forum of MedHelp community. To analyze the users, we had also collected the information of all users who registered as members of Swine Flu forum. Parsers were also built to parse and filter the collected data and the formatted data were stored in a database for the social network analysis process.

In the social network analysis phase, a social network is constructed based on the selected features of threads and replies. Once a social network is constructed, the user influence computing process quantifies user influence by our proposed technique. In this paper, we focus on how to construct social network and how to quantify user influence, which are the most important components of this model.

B. Constructing Social Network Based On Thread

Based on the formatted data, we propose the method of constructing social network. Let a node of the social network be a user who posts or replies to a thread in a discussion board. An edge in the network represents the action of replying a message. In MedHelp community, the main purpose of posting a thread on discussion board is asking questions or seeking for help. For this reason, within a thread, we assume that every time when A replies to B, A will make some impact on B because A answers B's question. In this case, A has some influence on B. An edge from B to A is constructed in the social network to confer the authority from B to A. In addition, as shown in Figure 1, only considering direct reply is not sufficient. Within a thread, users can reply to other user's reply which means that they are supplementing additional information or purposing different solution. For example, if B replies to C and then A replies to B's reply, there will be an edge from C to A indicating that C confers some authorities to A because A replies to C indirectly.

To have a better understanding of these direct and indirect replies, we use a tree structure to represent a thread. For example, Figure 3 illustrates a thread tree representing the thread in Figure 1. A node in a thread tree does not represent a user but represents a message posted by a user in this thread. For instance, the red user has posted 3 different messages in a thread. As a result, there are 3 nodes colored in red in the thread tree. Each node corresponds to a message posted by the red user. There are three branches rooted from the node corresponding to the initial post in this thread tree. Any thread can be represented by a similar structure by exploiting the post and reply relationship in a thread.

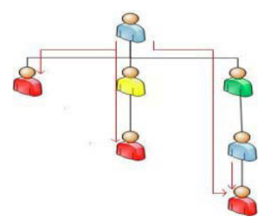

Figure 3. Tree Structure of Thread and Replies and Influence Blue User (B) getting from Red User (A)

In Figure 3, there are 4 paths from the red user (A) to the blue user (B), two of them are direct replies and two of them are indirect replies. To compute the weight for the edge (B,A), we introduce two basic principles of measuring influence in a thread tree:

- The shorter the distance between A and B, the larger influence A makes on B. This is considered as the immediacy effect. On a discussion board, once a user posts a message, it is likely that he checks whether his/her message receives any replies within an immediate period of time. Those who reply earlier will usually draw more attention and be more influential in the social network. The user who posts the questions will receive more influence from those immediate replies. The replies that are received later have a lower immediate impact. In some cases, the later replies may also be carried away by other less relevant discussions.

- The higher content similarity between the post and reply, the larger influence A makes on B. The similarity between post and reply is measured by cosine similarity typically used in the vector space model of many information retrieval tasks. Content similarity is particularly important when there are increasing spamming and random messages for advertising and other malicious purposes in online social networking.

Suppose user P posts several threads and user R replies to n of them. Within each thread of P, let M^P_j be the j^th^ message posted by P, which can either be a question or a reply, $0 < j < k1$. $k1$ is the total number of messages posted by user P within this thread. Let C^P_j be the content of this message and T^P_j be the timestamp of this message. Similarly, let R^P_i to be the i^th^ message within this thread which is posted by R, $0 < i < k2$. $k2$ is the total number of messages posted by user R within this thread. C^R_i corresponds to the content of this message and T^R_i corresponds to the timestamp of this message. $Similarity(C^P_j, C^R_i)$ denotes the cosine similarity between C^P_j and C^R_i.

Let $D(M^R_i, M^P_j)$ be the distance between message M^P_j and M^R_i in the thread tree. The distance between two messages in a thread tree can be defined in two possible ways:

Definition 1 (path length): The distance between two messages equals to the length of the downward path from one message to another message in the tree.

Definition 2 (time difference): The distance between two messages equals to the time difference of these two messages, $T^R_i - T^P_j$.

In our experiment presented in Section V, we compare the impact of these distance functions on the proposed ranking algorithm.

Based on the definitions above, we compute the weight of edge (P, R) by:

$$Weight(P,R) = \frac{\displaystyle\sum_{\substack{\forall \text{threads} \\ \text{that } P \text{ posts} \\ \text{and } R \text{ replies}}} \sum_{\substack{\forall j,i:\, M_j^P \text{ is an} \\ \text{ansector of } M_i^R \\ \text{in the thread tree}}} \beta\alpha^{|D(M_j^P, M_i^R)|} + (1-\beta) Similarity(C_j^P, C_i^R)}{n}$$

where $0 < \alpha < 1$ and $0 < \beta < 1$

In computing the weight from P to R, we consider all n threads that R replies to those P posts and every message in these threads where R responds to P. The weight on edge (P,R) is computed as the average flow in $\boldsymbol{n}$ threads. Within each thread, the larger number of paths from P to R, the larger the flow from P to R is. On each respond R makes to P, we incorporate both the distance between P and R's messages and the content similarity between these messages. Since α is between 0 and 1, a shorter distance yields a larger flow from P to R. A larger content similarity between two messages also leads to a larger flow from P to R. We combine the distance and content similarity measurement by parameter β, which is between 0 and 1.

For example, as shown in the thread in Figure 3, suppose it is the only thread in which user B and user A have participated and we employ the path length (Definition 1) to compute message distance, then $Weight(B,A)$ equals to:

$$\begin{aligned}&\{(\beta\alpha^1 + (1-\beta)Similarity(C_1^B, C_1^A)) + \\ &(\beta\alpha^2 + (1-\beta)Similarity(C_1^B, C_2^A)) + \\ &(\beta\alpha^1 + (1-\beta)Similarity(C_2^B, C_3^A)) + \\ &(\beta\alpha^3 + (1-\beta)Similarity(C_1^B, C_3^A))\} / 1\end{aligned}$$

C. Calculating User Influence in Social Network

Although PageRank [12] is developed to rank web pages according to their importance, it can also be applied on social network analysis to identify influential participants [15]. In social networks, the basic idea of PageRank is that if a participant $\boldsymbol{i}$ has a directed edge to another participant $\boldsymbol{j}$, then $\boldsymbol{i}$ is implicitly conferring some authority to $\boldsymbol{j}$ [15]. To compute the amount of authorities $\boldsymbol{i}$ confers to $\boldsymbol{j}$, we need to compute transition probability from $\boldsymbol{i}$ to $\boldsymbol{j}$, $P(\boldsymbol{j}|\boldsymbol{i})$. Given that the number of outlinks of $\boldsymbol{i}$ is N_i, the transition probability from $\boldsymbol{i}$ to $\boldsymbol{j}$ is $P(j|i) = 1 / N_i$. The amount of authority contributing from $\boldsymbol{i}$ to $\boldsymbol{j}$ is computed as $Authority(i) \times ((1-\alpha) P(j|i) + \alpha/|V|)$, where $0 < \alpha < 1$ and $|V|$ is the total number of nodes in the social network.

In this work, we propose the UserRank algorithm by incorporating the response immediacy and content similarity. The transition probability is computed by the weights of edges in the social network instead of a normalization of out-degrees of a node. Thus, the transition probability is computed as:

$$P(j \mid i) = \frac{Weight(i->j)}{\sum_{k:i->k} Weight(i->k)}$$

Where $Weight(i \rightarrow j)$ is the weight of edge (i,j) in social network. As a result, we measure the influence that $\boldsymbol{i}$ make on $\boldsymbol{j}$ as:

$$Influence(i) \times \left((1-\alpha) \frac{Weight(i \rightarrow j)}{\sum_{k:i\rightarrow k} Weight(i \rightarrow k)} + \frac{\alpha}{|V|} \right)$$

To compute the influence score of all users, we initialize the influence scores of all users to 1, and then we compute the influence score for each participant iteratively until they converge. Finally, we get the influence score vector corresponding to the dominant eigenvalue of the thread network.

We use the proposed UserRank algorithm to compute users' influence and the most influential users in thread network are identified by those who have the highest influence scores.

V. EXPERIMENT AND DISCUSSION

In this section, we present experiments that evaluate the effectiveness of our approach.

A. Data Sets

We tested our techniques on Swine Flu forum. Swine Flu forum was a small size forum which had 72 registered members and each of them had 15.6 friends on average. On the discussion board of swine flu forum, there were 90 threads in total.

B. Evaluation Measures

The quality of a ranking method requires human evaluation. To evaluate a ranking algorithm quantitatively, user studies are conducted to obtain the standard user influence ranking of a thread network by human judges first. In addition, two performance metrics are adopted in our experiments:

- **Precision:** Precision measures the number of influential users extracted by an algorithm that are also influential users obtained in the standard ranking divided by the number of influential users extracted by an algorithm. Suppose the standard ranking ξ contains N users who are labeled as Top-N influential users in the user studies. To evaluate the precision of an algorithm, a ranking ψ which contains top N influential users are extracted. The precision of the algorithm is defined as:

$$\frac{|\xi \cap \psi|}{N}$$

- **Rank Distance:** Rank Distance [4] is used to measure the similarity between two rank lists. A ranking of a set of n objects can be represented as a permutation of the integers 1, 2, … , n. S is a set of ranking results. $\sigma_i \in S$. $\sigma_i(p)$ represents the rank of object p in the ranking result i. The rank distance is computed as:

$$D(\sigma_i, \sigma_j) = \sum_{p=1}^{n} |\sigma_i(p) - \sigma_j(p)|$$

C. Experiemnt One

In this experiment, based on the techniques introduced in Session IV.B, two social networks S_1 and S_2 were constructed from the Swine Flu forum data, where the weights on edges of S_1 were computed according to the path length (Definition 1) and the weights on edges of S_2 were computed according to the time difference (Definition 2). We used UserRank algorithm to identify influential users within these two networks. Table I shows that UserRank algorithm does not obtain any substantial difference on these two social networks.

TABLE I. PRECISION AND RANK DISTANCE OF USERRANK ALGORITHM ON DIFFERENT SWINE FLU SOCIAL NETWORKS

	UserRank on Social Network S_1	*UserRank on Social Network S_2*
Precision	100%	100%
Rank Distance	2	2

D. Experiment Two

In this experiment, following the techniques introduced in Session IV.B and using time difference (definition 2) to compute message distance, we constructed the social network based on formatted data of the Swine Flu forum. We ran our proposed UserRank algorithm to identify influential users. At the same time, we benchmarked our proposed techniques with several other techniques including: (i) **PageRank**: The PageRank algorithm proposed by Page and Brin [12] and introduced in Section IV.C was used to compute user influence. The PageRank algorithm did not take content similarity and thread structure into account, (ii) **In-Degree:** The ranking of users was computed by the number of in-links. The more in-links a user had in the social network, the higher the position of this user in the ranking was. (iii) **Out-Degree:** The ranking of users was computed by the number of out-links. The higher out-degree corresponded to the larger influence this user had in the social network.

Figure 4 shows the precision of PageRank, UserRank, In-Degree and Out-Degree respectively to identify top 5 influential users in Swine Flu forum. Only top 5 influential users are considered because Swine Flu forum is a relatively new forum with fewer numbers of users.

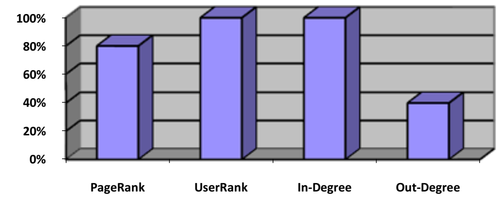

Figure 4. Precision of Different Technique in Swine Flu Forum

As shown in Figure 4, UserRank and In-Degree achieves the highest precision, which is higher than PageRank and Out-degree. In particular, Out-Degree obtains substantially lower precision than the other three techniques.

Table II presents the rank distance between the ranking obtained by each of the four algorithms and the standard ranking.

TABLE II. RANK DISTANCES BETWEEN STANDARD RANKING AND THE RANKINGS GENERATED BY DIFFERENT TECHNIQUES FOR TOP 5 INFLUENTIAL USERS OF SWINE FLU FORUM

	PageRank	*UserRank*	*In-Degree*	*Out-Degree*
Rank Distance	5	2	4	54

As shown in Table 1, UserRank algorithm is better than PageRank algorithm, In-Degree and Out-Degree measurement. In-Degree is slightly better than PageRank. Out-Degree measurement is the worst in ranking influential users.

E. Experiment Three

To test the scalability of our proposed technique, we conducted another experiment on a larger dataset, alcoholism forum. Alcoholism forum has 446 registered members and each of them has 9.2 friends on average. On the discussion board of alcoholism forum, there are 737 threads in total and each thread received 7.2 replies on average. We employed the same evaluation measures as introduced in Section V.B.

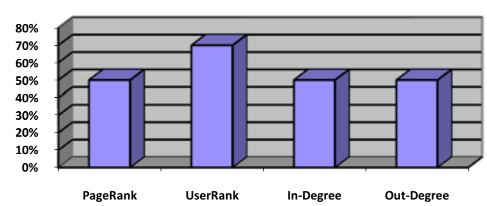

Figure 5. Precision of Different Technique in Alcoholsim Forum

TABLE III. BETWEEN STANDARD RANKING AND THE RANKINGS GENERATED BY DIFFERENT TECHNIQUES FOR TOP 10 INFLUENTIAL USERS OF ALCOHOLISM FORUM

	PageRank	***UserRank***	***In-Degree***	***Out-Degree***
Rank Distance	51	36	29	89

Figure 5 shows that UserRank achieves substantially higher precision (40% higher) than PageRank, In-Degree, and out-Degree. However, in-Degree achieves lower rank distance than UserRank. In this case, UserRank obtains a higher rank distance mainly due to an outlier. In general, UserRank achieves a better performance than PageRank and Out-degree but relatively weak in rank distance when comparing with In-Degree.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we investigate the problem of identifying influential users in online communities. The previous work focuses only on the link structure of social networks to identify influential users. We propose the UserRank algorithm that incorporates the link structure, content similarity and responding order and time of repliers. Experimental results show that our technique outperforms PageRank, in-degree and out-degree on identifying influential user from online healthcare support social network. Based on the framework and the techniques proposed in this paper, researchers can build a real-time system to monitor those sensitive or important forums to identify influential forum users.

Network influence analysis represents a new and interesting research direction in social network mining. There are many potential future directions of this work. One interesting issue is incorporating language model into our work so that we may have better understanding of thread and journal content when calculating user influence. Another issue is studying the relationship between influential users and social network evolution.

REFERENCES

[1] J. Brown and P. Reingen, Social ties and word-of-mouth referral behavior, The Journal of Consumer Research, vol. 14, no. 3, 1987, pp. 350-362.

[2] W. Chen, Y. Wang and S. Yang, *Efficient influence maximization in social networks*, *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2009, pp. 199-208.

[3] J. Coleman, E. Katz and H. Menzel, *Medical innovation: A diffusion study*, Bobbs-Merrill Co., 1966.

[4] L. Dinu, *On the Classification and Aggregation of Hierarchies with different Constitutive Elements*, Fundamenta Informaticae, vol. 55, no.1, 2003, pp. 39-50.

[5] P. Domingos and M. Richardson, *Mining knowledge-sharing sites for viral marketing*, *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* FL, USA, 2002, pp. 61-70.

[6] P. Domingos and M. Richardson, *Mining the network value of customers*, *Proceedings of the 7th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2001, pp. 57-66.

[7] J. Goldenberg, B. Libai and E. Muller, *Talk of the network: A complex systems look at the underlying process of word-of-mouth*, Marketing Letters, vol. 12, no. 3, 2001, pp. 211-223.

[8] J. Goldenberg, B. Libai and E. Muller, *Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata*, Academy of Marketing Science Review, vol. 9, no. 3, 2001, pp. 1-18.

[9] A. Goyal, F. Bonchi and L. V. S. Lakshmanan, *Discovering leaders from community actions*, *Proceeding of the 17th ACM conference on Information and knowledge management*, Napa Valley, California, USA, 2008, pp. 499-508.

[10] D. Kempe, J. Kleinberg and Tardos, *Maximizing the spread of influence through a social network*, *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2003, pp. 137-146.

[11] V. Mahajan, E. Muller and F. Bass, *New product diffusion models in marketing: A review and directions for research*, The Journal of Marketing, vol. 54, no.1, 1990, pp. 1-26.

[12] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank citation ranking: Bringing order to the web* (Technical Report), Computer Science Department, Stanford University, 1999.

[13] J. Tang, J. Sun, C. Wang and Z. Yang, *Social influence analysis in large-scale networks*, *15th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2009, pp. 807-816.

[14] T. Valente, *Network models of the diffusion of innovations*, Computational & Mathematical Organization Theory, vol. 2, no. 2, 1996, pp. 163-164.

[15] J. Zhang, M. Ackerman and L. Adamic, *Expertise networks in online communities: structure and algorithms*, Proceedings of the 16th international conference on World Wide Web, Banff, Alberta, Canada, 2007, pp. 221-230.