Study on Influence Diffusion in Social Network

Hsiao-Wei Hu and Shao-Yu Lee

***Abstract*—** As the growth of online social network mining, more and more researchers focus on the diffusion effect in social network. Most of these studies concentrate on identifying the key nodes, detecting potential groups and the predicting the users' behaviors, but they seldom consider the impact of time. In fact, like the prime time in TV network, the influence power of time is an important factor to determine the diffusion effect in social network. For selecting the appropriate time to spread message, not only the number of online users, but the structure of relationship and attributes of users need to be taken into account. In this study, we define three types of users according to their profiles, and propose an approach to extract the influence power patterns of time, so we can predict the proper time to announce information from historical data.

***Keywords*—**Social Network Mining, Frequent Pattern.

I. Introduction

THE dependence of people on Internet results in the rapid expansion of Internet applications and services. With the growth of Internet service, more and more information sharing systems are developed. Especially, social networking service has got a lot of attention recently and becomes one of the most popular sites on the Web now.

With the growth of members/users and the large variety of population, these social networking service websites are not only social platforms, but state-of-art marketing channels of business today. That is, social networking service websites are a form as well as an online tool of electronic word-of-mouth platform. So, corporations can make marketing strategies through observing the opinions or the behaviors of customers on these online tools. The enterprises can promote their products effortlessly through the websites, and get a lot of feedback. Besides, according to [1], the social networking service websites can also become the Web-based consumer opinion platforms, and enable customers to share the opinions and experiences with other consumers.

Facebook, one of the most famous social network services and websites in whole world, already has more than 1 billion active users (Facebook Newsroom, 2012; Wikipedia, 2012). In March 2010, SocialMediaExaminer.com synthesized five studies to show why Facebook is a marketing powerhouse [2]: First, according to survey by the Nielson Company, they found out users spend 7 hours on using Facebook per month [3]. Second, the research [4] of TechCrunch.com proposes that 44% of social sharing items are distributed by Facebook. Third, Compete.com (2010) reported that Facebook has surpassed Yahoo, and becomes the number-two popular site in U.S.(Google is number one.) Fourth, [5] show that users are more likely to reveal their true selves online, instead of "faking it". Finally, [6] indicated that Facebook can boost sales and customer loyalty.

Sum up the above, we believed Facebook is a new battlefield for advertisers. And an increasing number of articles start to focus on successful factors for marketing [7][8][9]. The other advantages of marketing in Facebook are its powerful functions, like "the wall", "ticker", "fans club", and "group", etc. The above functions are parts of the key factors that make Facebook become the popular marketing channel. Facebook has a lot of usable functions, let users need not to just wait for updating. If business can operate the fans page effectively, it's possible to make marketing and advertising successful [10], so the study of social network becomes a hot topic recently.

Social network is a social structure which made up by nodes (such like individuals or organizations) and links (between these nodes). Individuals and organizations can be linked together by their relationships, such as colleagues, schoolmates, friends, business partners, etc. In order to help business understand their social network structure of fans, researchers usually use social network analysis tools to obtain the profiles of the network. But, if business want to acquire potential information, social network mining is required. For instance, community detection [11][12][13], intention analysis and prediction [14][15], etc.

Social network mining is the research of social relations between nodes. The study of social network mining technologies focus on the level of individuals, groups, organizations, and even whole network [16]. The main methodologies of social network mining are sociometric analysis and graph theory [17]. We provide the reasons why online social network mining is important and popular: Online social network mining is a popular research issue mainly because online social network mining can across the geographical and time constrains, so online social network mining is more malleable than the traditional geographic-based community mining. Because of the success of online social network and media-sharing sites, business pays more attention to online social network mining. In addition, social network mining can help business to identify potential customers, to find out the purchasing patterns from large amount of data, and so on [18]. For example, in [19], authors use the techniques of social network mining, such as group detecting, group evolution tracking and group life-cycle modeling, in telecom

H.-W. Hu is with the Department of Information Management, Fu-Jen Catholic University, Taipei, Taiwan. (Corresponding author, phone: +886-2-29054134; fax: +886-2-29054134; e-mail: camihu@gmail.com).

Shao-Yu Lee is with the Information Management Department, Fu Jen Catholic University, New Taipei City, Taiwan R.O.C. (e-mail: mark19890619@hotmail.com).

CRM domain. This can help business to understand the users' interaction through comprehending the structure of social network and get more opportunities of making profits.

Understanding the diagram structure of social network can help people to assess the existing systems, to design future social network based systems, and to Gain website traffic or information/advertisement attention through social network [20]. For example, researchers use online social network to improve Internet search. They proposed that leveraging online social networks to build new search system [21]. Furthermore, [22] proposes that using social network mining to filter the e-mail spam. They present a new system which has zero false positives of connected users. Besides, Social network marketing now is a popular advertisement type, and the goal of online advertisement is to drive traffic to brand websites [23].

There are a large number of studies about social network mining. Previous researches can be divided into only two categories, static-based methods and dynamic-based methods. Static-based method means to analyze the structure of social network and understand the characteristics of partition or whole network. In [24], authors used the snowball sampling method to crawl data and analyzed the topological characteristics of online social network service. Dynamic-based method is a process of monitoring social network and to detect when the change of social network structure occurs and what the reason of it. In order to find out what structural features will influence individuals to join communities and how the overlaps of communities change over time, the study used the snapshot to detect dynamic change of social network in [25].

Among these studies, we notice that impact of timing is an important factor that needs to be considered, especially when we are analyzing the structure of online social network, especially for marketing purpose. Because, information can only be spread to users who are actually online, we can hardly deliver the message to users who are offline, because the messages or the information posted on wall are likely fade out with time. For example, prime time is the daypart with the most viewers and is generally where television network and local stations reap much of their advertising revenues. Like the prime time of television, we believe there is the prime time in social network to spread information, so the timing is important for spreading.

In addition, even if the online users do not pay attention to the messages released by groups or fans clubs seriously, they may still have subconscious impression. This argument was very popular in the past, and it is called subliminal perception [26]. Subliminal perception means the impact of thoughts, feelings or actions when the consciousness stimulus is lower than threshold. That is, the brain can receive the messages, but the subjects do not realize the perception. We believe that there is the subliminal perception not only in Industrial and Organizational Psychology, but also in online social network. And, the subliminal perception in online social network is closely related to the timing of information announcing. So, in this study we take "timing" to distribute message into account and develop a hybrid method, combining static-based method and dynamic-based method, to help decision makers identify the "prime time" to post messages that can bring the most positive feedback.

The selection of timing to spread information is important, so we think it's necessary to find out the correct time. Identifying the most influential time can help organizations to select the correct time for delivering the information to enhance the efficiency of information diffusion, reduce the waste of resources (such like time), and avoid annoying users. We can cite the well-known saying of marketing and communication here, "Reaching the right people at the right time in the right way."

However, previous studies have not considered (discussed) the impact of timing, and therefore, have no methods designed for identifying appropriate time point for spreading information in social network. To remedy this research gap, in this paper, we propose a new social network mining approach which can help business to identify the most appropriate time point for spreading information more efficiently and more effectively.

In our study, we prefer to use frequent pattern mining algorithm because the target database is transaction database, which means the data is composed of items. The well-known method of frequent pattern mining is Apriori.

The rest of this paper is organized as follows: We describe related work in Section 2, and detail definition of the problem in Section 3. We show our methodology in Section 4, and experiment in Section 5.

II. Related Work

In this section, we describe the studies of social network mining, electronic word-on-mouth, subliminal perception, as well as Apriori algorithm.

A. Data Mining and Social Network Mining

[27] surveys the literatures of data mining techniques from 2000 to 2011. Authors classify the techniques into a few types, and discuss the applications and the recent developments of these techniques. This paper can help researchers to get the guideline of data mining techniques study easily and quickly. In [28], authors give a comprehensive definition and generalize the characteristics of social network sites. Besides, they use the historical point of view to talk about the evolution of social network sites, the studies about social network sites, and speculate the future research. It can give beginners a reference guide to understand the existing research and potential research in social network.

A clear definition of social network and social network mining (analysis) are proposed in [29], and it also provides detailed introduction of studies from pure mathematical analyses in graphs to analyzing the social networks in Semantic Web. Also, the authors classify the research topics and describe the features of them. Social network mining can be rough divide into two categories: static-based methods and dynamic-based methods. We've compiled a list below.

In [30], authors provide two methods, context-dependent and context-independent approaches, to identify the actors in a social network. They classify the actors into four types: leaders, lurkers, associates and spammers. It can help business to find out the key person and exclude the spammers. In addition, researchers can easily classify the nodes of a social network through this method and do further studies.

TABLE I. STATIC-BASED AND DYNAMIC BASED METHODS OF SOCIAL NETWORK MINING.

Static-Based Methods	Dynamic-Based Methods
Community detection in social networks with genetic algorithms. (Pizzuti, 2008)	***Characterizing user behavior in online social networks. (Benevenuto, et al.,2009)***
Fast algorithm for detecting community structure in networks. (Newman, 2004)	***Identifying User Behavior in Online Social Networks. (Maia, et al., 2008)***
Community detection in large-scale social networks.[31] (Nan Du, et al., 2007)	***Identification of leaders, lurkers, associates and spammers in a social network context-dependent and context-independent. (Fazeen, et al.,2011)***
Finding community structure in very large networks. (Clauset, et al., 2004)	

B. Electronic Word-of-Mouth

The interaction of users in social network can be regarded as an electronic word-of-mouth marketing for business. [32] study how the electronic word-of-mouth influences the purchase action of customers. They develop a customer purchase intention model, and simulate various situations. The researchers define every possible context and examine the purchase intention in all contexts. The results prove once again that the electronic word-of-mouth will strongly influence the purchase intention of customers. In [33], they examine micro-blogging as a form of electronic word-of-mouth branding. Researchers consider that micro-blogging service may be a main application in branding campaigns. Using micro-blogging as part of marketing strategy is the trend, and micro-blogging will becomes the favorable resources for organizations.

C. Subliminal Perception

[34] and [35] review the studies of subliminal perception in laboratory and provide the future works. In the treatise of [36], BBC sent a film to both the eastern and western region networks, and BBC presented a picture of subliminal stimulus on the film in the eastern region network, but not in the western region network. The researchers observe that there are differences between these audiences definitely.

D. The Ariori Algorithm

Agrawal and Srikant propose a new algorithm, called apriori, to find association rules from large scale database. Apriori is based on the minimum support threshold to obtain the outcome, and the results consist of more item.

III. PROBLEM DEFINITION

To ease the presentation, we use "social network" to represent "online social network". Before we jump into the detail of the algorithm, we first need to clarify the right person/node and the right time.

A right node means a node that has the ability to influence other nodes in the graph. That is, the right node can easily influence other nodes through direct or indirect connects. Influence is the capability of information diffusion in social networks. With the greater influence, the information can be spread to other users/nodes more quickly and widely.

We divide influence power of node into two dimensions, influence range and influence degree, as Table 2. Influence range means the spreading scope of message or behavior, and influence degree represents how deep the node can affect neighbors.

TABLE II. NINE TYPES OF INFLUENCE POWER.

		Influence Range		
		Low	***Medium***	***High***
Influence Degree	***Low***	Type 1		
	Medium		Type 2	
	High			Type 3

[a] *Type 1: low influence power node*

[b] *Type 2: medium influence power node*

[c] *Type 3: high influence power node*

Right time means the time which has the greatest influence. That is, a right time point is a specific time for a social network that has a large number of right people who are actually online.

.

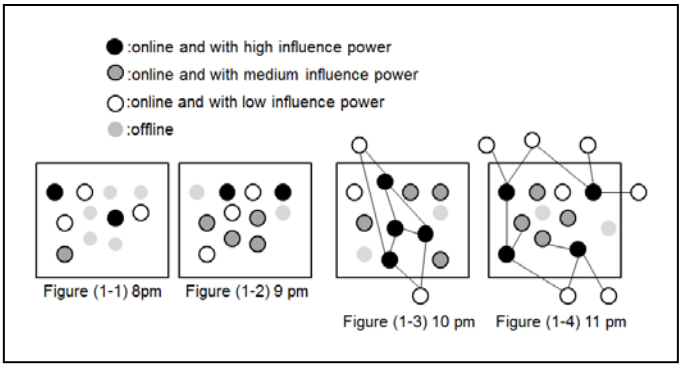


Fig. 1 Online status of nodes in social network

Figure 1 displays some possible online status of nodes in social network. Figure (1-1) and figure (1-2) have same number of online nodes which have high influence power and different number of online nodes which have medium and low influence power. According to the figure (1-1) and figure (1-2), we can infer that the capability of information diffusion may be more powerful in figure (1-2) than in figure (1-1) because the number of online nodes in figure (1-2) is more than figure (1-1) and the number of nodes which have medium influence power in figure (1-2) is more than in figure (1-1).

In figure (1-3), there is the same number of online nodes with figure (1-2), but there are more high influential nodes in figure

(1-3). So we can infer that the influence power of figure (1-3) is stronger than figure (1-2) because of the number of high influential nodes.

As figure (1-3), there is the same number of online nodes which have and have no great influence in figure (1-4), but the relation structure of nodes are different. We can see that there are two nodes which have been influenced and not in group in figure (1-3), but six in figure (1-4). So we can infer that the information diffusion may be more effective in figure (1-4) than figure (1-3). For selecting the right time, we need to consider the distribution of nodes and the relation structure of nodes.

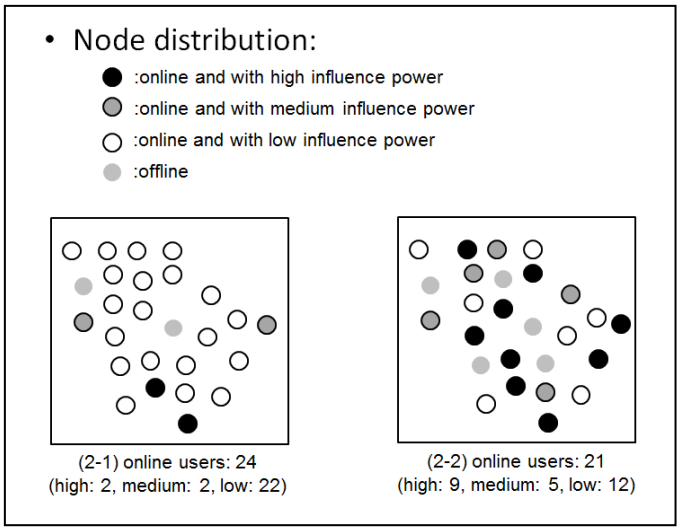


Fig. 2 Online status of nodes in social network

There are two different relation structures of nodes in figure 2. The amount of online users in figure (2-1) is more than figure (2-2), but the quantity of high and medium influence nodes in figure (2-2) is much more than in figure (2-1). If we only use the number of online users to evaluate the influence power, the results is difficult to convince people. So we need to consider the distribution of nodes. Besides, not only the amount of online nodes and their distribution, but the relation structure of nodes must be taken into account. Here come some examples in figure 3.

.

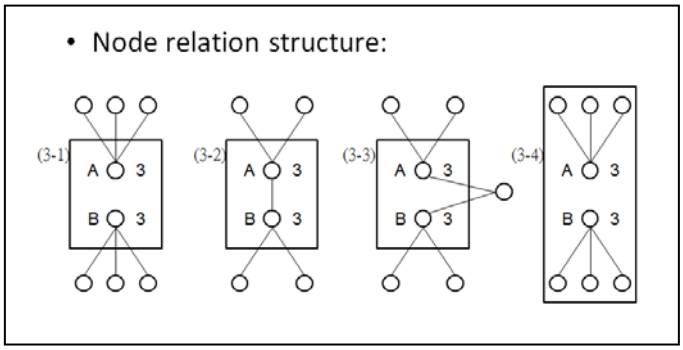


Fig. 3 Comparison between different relation structures of nodes

Figure (3-1) shows that there are two nodes, A and B, and they both have three neighbors which are not in the group. So in figure (3-1) the information can be spread to 6 nodes or more. The difference between figure (3-1) and (3-2) is that node A and B have relationship with each other in figure (3-2), so the information can only spread to four or more nodes which are not in the group.

The difference of figure (3-2) and figure (3-3) is that node A and B have relationship to each other in figure (3-2), but not in figure (3-3). However, A and B have mutual friend in figure (3-3), so the information can spread to five or more nodes which are not in the group. Like figure (3-1), the two nodes, A and B, have totally six neighbors, and has no relationship or mutual friend with each other in figure (3-4). But, in figure (3-4) their neighbors are all in the group, so the information cannot be spread to other node which is not in group.

A. Definition 1: Social Network

Time period $T = \{t_i \mid i = 1 \text{ to } k\}$ is a set of time periods defined by user. If the minimum time granularity p is set as an hour, then the t_1 could be a period started from 0:00 to 1:00, t_2 is a period started from 1:00 to 2:00, and t_i is a period started from $(0{:}00 + (i-1) \times p)$ to $(0{:}00 + i \times p)$.

Node $N = \{n_j \mid j=1 \text{ to } m\}$ is a set of registered node in a social network SN. $N_i^+ = \{n_j \mid j=1 \text{ to } m\}$ and $N_i^+ \subset N$. N_i^+ is a set of online nodes within a time period t_i. $|N|$ is the number of nodes in SN. $|N_i^+|$ is the number of online nodes in SN within t_i.

B. Definition 2: Influence Range

Let $|F(n_j)|$ be the number of friends of a node n_j. We believe that the more friends you have, the messages may spread wider from you. So the number of friends is the most simple and effective factor to evaluate the influence power of nodes.

The number of friends is not enough to evaluate the influence power of nodes because the connection structure of nodes' friends may impact the range of information spreading. The more mutual friends the nodes have, the smaller range the information spread. So we need to consider another factor, the structure of friends (SF).

$$SF(n_j) = \frac{\sum S(n_f)}{|F(n_j)|}, n_f \in F(n_j),$$

$$\begin{cases} \text{If node } n_f \text{ in the group, } S(n_f) = 0 \\ \text{If } n_f \text{ not in the group, } S(n_f) = 1/|FiG(n_f)| \end{cases}$$

$|FiG(n_f)|$ means the number of n_f's friends in the group.

Let $IR(n_j) = |F(n_j)| * SF(n_j)$, that means the influence range of node n_j.

C. Definition 3: Influence Degree

Influence degree depicts the characteristics of nodes, such like the frequency of pressing like, comment and sharing. $ID(n_j)$ represent the influence degree of node n_j.

Let the number of like buttons pressed by node n_j per week be $L(n_j)$, the number of responding the messages per week be $R(n_j)$, and the number of posting the messages per week be $P(n_j)$.

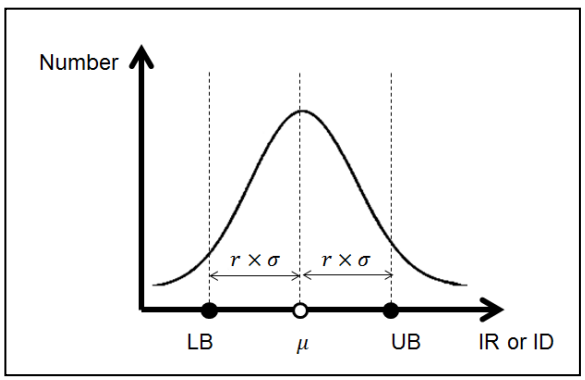
According to the research by [37], they used Clicks to evaluate the value of Likes, Comments and Impressions. The results are as follows:

- Avg. Clicks Per Like: 3.103
- Avg. Clicks Per Comment: 14.678
- Avg. Clicks Per Impression: 0.005
- Shares are much more important than the above.

And, a report of Pew Internet & American Life Project suggests that the average Facebook user has 229 friends and 26% people "Like" another user's contents [38].

Sum of above, we generalize the efficacy of Like, Comment and Share as follows:

- The influence power of per Like: 3.103 (Rounding: 3)
- The influence power of per Comment: 14.678 (Rounding: 15)
- The influence power of per Share: More than 59.54 (Rounding 60) = (avg. number of friends) × (ratio of pressing Like) = 229 × 0.26 = 59.54
- We let $ID(n_j) = a \times L(n_j) + b \times R(n_j) + c \times P(n_j)$, and $a + b + c = 1$.
- $a = 3/(3+15+60) \fallingdotseq 0.04$, $b = 15/(3+15+60) \fallingdotseq 0.19$, $c = 60/(3+15+60) \fallingdotseq 0.77$
- $ID(n_j) = 0.04 * L(n_j) + 0.19 * R(n_j) + 0.77 * P(n_j)$



Fig, 4 The distribution of IR or ID. (μ is mean and σ is standard deviation.)

After find out ID and IR of nodes, we can calculate two points, LB and UB, by statistical processing. As shown in figure 4, let $LB = \mu - (r \times \sigma)$ and $UB = \mu + (r + \sigma)$. r is a variable between 1 to 3. If r is bigger, the amount between LB and UB will becomes larger. The two points can divide IR and ID into three levels, so we can get the table 2.

D. Definition 4: Transaction Database

Transaction $TR_s = \{t_S, V_S, [(n_o, V_o) \ldots]\}$, stored in database D. t_S is the time of transaction TR_s. $n_o \in N^+$, the online nodes at time t_S. V_S/V_o means the value (influence diffusion power) of transaction TR_s/node n_o.

$$V_S = x * |L| + y * |M| + z * |H|$$

$|L|$, $|M|$, $|H|$ are the amounts of V_o's L, M, H.

$0.1 \le x \le 0.3$, $0.4 \le y \le 0.6$ and $0.7 \le z \le 0.9$,

and the variables x, y, z are random.

Like the influence power we mentioned before, through statistical process, V_S can be classified into three levels.

E. Definition 5: Period and the Selected Database

Let δ be the length of time we want to predict. If $V_S = H$, then we will pick the transactions $V_{S-\delta}$ into the selected database *d*. That means the pattern we find can predict the influence diffusion power after length of time δ is High or not.

F. Definition 6: Support and Frequent Pattern

Suppose *I* is a set of items, including one or more item *(n,V)*.The support value of *I* in *d* is the occurrence frequency of itemset *I* in the selected database *d*, represented as *support(I,d)*.

$$support(I,d) = \frac{count(I,d)}{|d|}$$

count(I,d) means the number of transactions which contain itemset *I* in the selected database *d*. |*d*| is the amount of transactions in database *d*.

If *support(I,d)* is greater than a given threshold *minsup_d*, itemset *I* is the frequent pattern in database *d*.

IV. Methodology

In this section, we will discuss the notion of our algorithm. First of all, we need to define some symbols. A set of candidate frequent patterns in database *d* is denoted as *Cd*. *FPd* represents the frequent patterns in database *d*.

A. Phase 1:Set peroid and create the selected database d.

Assume we have a complete transaction database *D*, and we want to find out what patterns will result in the high influence power time after a period δ. In order to save the scanning time, we need to take out the transactions which precede the high influence power transaction and their interval must be δ. The collection of selected transactions is called the selected database *d*.

B. Phase 2: Scan database d and compute the support value of length-1 itemset.

After we obtain the selected database *d* , it's necessary to scan all database *d* and count the occurrence frequency of every length- 1 itemset. If there is *count(I, d)* greater than a given threshold *minsup_d*, itemset *I* will become length-1 frequent pattern.

C. Phase 3: Scan all frequent pattern in database d.

Combine any two items which are frequent occurrence and get length-2 candidates *Cd*. Scan all database *d* and compute the support value of length-2 *Cd.* If there is *count(I, d)* greater than a given threshold *minsup_d*, itemset *I* will become length-2 frequent pattern. Repeat this step until there is no more frequent pattern. So that, we can acquire all frequent patterns *FPd* in the selected database.

D. Phase 4: Check all frequent pattern in transaction database D.

FPd is the frequent pattern in d, but it doesn't indicate that there is a high influence power time after *FPd* occurs. So we need to examine the validity of *FPd* by scanning the complete database *D*. We need to calculate the support value *support(D, I)* which means when itemset *I* occurs at some time, the influence power after δ period is high or not. If *support(D, I)* is greater than a given threshold *minsup_D*, we can confirm the *FPd* has ability to affect the influence power.

V. Experiment

We establish a film-related fans club, and collect the data of fans and their friends to build up our database.

Every Facebook user has a unique identity number, so the number will be the symbol to identify user in our database. According to the users' profile, we can easily compute their influence range and influence degree. Users will be classified users into three types (high, medium and low influence power). Aggregate these data above, and become our complete transaction database *D*.

Before mining the frequent patterns, we need to set the parameters δ, *minsup_d* and *minsup_D*. In order to predict the right time accurately, different parameters δ are required, but the more parameter δ you want to analysis, the more time need to spend. *minsup_d* and *minsup_D* also need to adjusted, so that we can obtain appropriate results.

VI. Conclusion

Online social network is a popular service recent years. It not only provides a social platform for users, but can make lots of visible and invisible profits for business by promotion, eWOM, advertisement, CRM. In the past studies of social network mining, business can easily find out the potential customers, information diffusion path, the behaviors of their consumers, etc., but they can't determine one of the most important things, the timing of posting messages. Therefore, we want to detect the appropriate time to announce messages for business. According to our analysis, we realize that there are many factors which will impact the influence power of nodes and time. Through our experiment, we discover that the conditions we assumed in section 3 exist in reality, and this shows the importance of mining the influence time. This study will help business to detect the prime time in their online social network, and predict the influence power of future.

References

[1] Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury, Twitter Power: Tweets as Electronic Word of Mouth, *Journal of the American Society for Information Science and Technology, 60(11)* , 2009, pp. 2169–2188.

[2] Amy Porterfield, 5 New Studies Show Facebook a Marketing Powerhouse, Social Media Examiner, Mar. 15, 2010. Retrieved Dec. 15, 2012.

[3] Facebook Users Average 7 hrs a Month in January as Digital Universe Expands, NeilsonWire, Feb. 16, 2010. Retrieved Dec. 15, 2012.

[4] Erick Schonfeld, Facebook Drives 44 Percent Of Social Sharing On The Web, TechCrunch, Feb. 16, 2010. Retrieved Dec. 15, 2012.

[5] Mitja D. Back, Juliane M. Stopfer, Simine Vazire, Sam Gaddis, Stefan C. Schmukle, Boris Egloff, and Samuel D. Gosling, Facebook Profiles Reflect Actual Personality, Not Self-Idealization, *Psychological Science, 21(3)*, 2010, pp. 372-374.

[6] Utpal M. Dholakia and Emily Durham, Harvard Business Review, Mar. 2010. Retrieved Dec. 15, 2012.

[7] Gina Sverdlov, The Facebook Factor, Forrester, Apr. 9, 2012. Retrieved Dec. 15, 2012.

[8] Jason Keath, Justin Kistner, Ellie Mirman, and Justin Levy, Facebook Page Marketing, HubSpot, unknown date. Retrieved Dec. 15, 2012.

[9] Jeremiah Owyang, The 8 Success Criteria For Facebook Page Marketing, Altimeter, July 27, 2010. Retrieved Dec. 15, 2012.

[10] Facebook Function in Business Advancement – The New Experience of Business Promotion, Everything Facebook, Feb 5,2012. Retrieved Dec. 15, 2012

[11] Clara Pizzuti, Community Detection in Social Networks with Genetic Algorithms, *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, Atlanta, GA, USA, Association for Computing Machinery, 2008, pp. 1137-1138.

[12] Aaron Clauset, M. E. J. Newman, and Cristopher Moore, Finding community structure in very large networks, *Physical Review E, 70, 066111*, 2004.

[13] M. E. J. Newman, Fast algorithm for detecting community structure in networks, *Physical Review E, 69, 066133*, 2004.

[14] Marcelo Maia, Jussara Almeida, and Virgílio Almeida, Identifying User Behavior in Online Social Networks, *Proceedings of the 1st Workshop on Social Network Systems*, 2008, pp. 1-6.

[15] Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida, Characterizing user behavior in online social networks, *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, Chicago, Association for Computing Machinery, 2009, pp. 49-62.

[16] Elizabeth F. Churchill, and Christine A. Halverson, Social Networks and Social Networking, *IEEE Internet Computing Magazine*, 2005, pp. 14-19.

[17] John Scott, *Social Network Analysis : a hand book*, California: SAGE Publications, 2000.

[18] Francesco Bonchi, Carlos Castillo, Aristides Gionis, and Alejandro Jaimes, Social Network Analysis and Mining for Business Applications, *ACM Transactions on Intelligent Systems and Technology (TIST), Volume 2 Issue 3 Article 22*, April 2011.

[19] Bin Wu, Qi Ye, Shengqi Yang, and Bai Wang, Group CRM: a New Telecom CRM Framework from Social Network Perspective, *Proceedings of the 1st ACM international workshop on Complex networks meet information & knowledge management*, Hong Kong, Association for Computing Machinery, 2009, pp. 3-10.

[20] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee, Measurement and analysis of online social networks, *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, San Diego, Association for Computing Machinery, 2007, pp. 29-42.

[21] Krishna P. Gummadi, Alan Mislove, and Peter Druschel, Exploiting Social Networks for Internet Search, *Proceedings of the 5th Workshop on Hot Topics in Networks*, Irvine, Association for Computing Machinery, 2006.

[22] Scott Garriss , Michael Kaminsky , Michael J. Freedman , Brad Karp , David Mazières , and Haifeng Yu, Re: Reliable Email, *Proceedings of the 3th USENIX Symposium on Networked Systems Design and Implementation*, San Jose, USENIX, 2006, pp. 297-310.

[23] Tracy L. Tuten, *Advertising 2.0: Social Media Marketing in a Web 2.0 World*, Westport, Connecticut: Greenwood Publishing Group, 2008.

[24] Yong-Yeol Ahn, Seungyeop Han, Haewoon Kwak, Young-Ho Eom, Sue Moon, and Hawoong Jeong, Analysis of Topological Characteristics of

Huge Online Social Networking Services, *Proceedings of the 16th International World Wide Web conference*, Alberta, 2007, pp. 835-844.

[25] Lars Backstrom, Dan Huttenlocher, Xiangyang Lan, and Jon Kleinberg, Group Formation in Large Social Networks: Membership, Growth, and Evolution, *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, Philadelphia, Association for Computing Machinery, 2006, pp. 45-54.

[26] Goldiamond, Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology, *Psychological Bulletin, Vol 55(6)*, 1958, pp. 373-411.

[27] Shu-Hsien Liao, Pei-Hui Chu, and Pei-Yuan Hsiao, Data mining techniques and applications – A decade review from 2000 to 2011, *Expert Systems with Applications*, *Volume 39, Issue 12*, 2012, pp. 11303-11311.

[28] danah m. Boyd & Nicole B. Ellison, Social Network Sites: Definition, History, and Scholarship, *Journal of Computer-Mediated Communication, 13(1), article 11*, 2008, pp. 210-230.

[29] Mohsen Jamali & Hassan Abolhassani, Different Aspects of Social Network Analysis, *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, Hong Kong, 2006, pp. 66-72.

[30] Mohamed Fazeen, Ram Dantu, and Parthasarathy Guturu, Identification of leaders, lurkers, associates and spammers in a social network: context-dependent and context-independent approaches, *Social Network Analysis and Mining, Volume 1, Issue 3*, 2011, pp. 241-254.

[31] Nan Du, Bin Wu, Xin Pei, Bai Wang, and Liutong Xu, Community detection in large-scale social networks, *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, San Jose, 2007, pp. 16-25.

[32] Jung Lee & Jae-Nam Lee, Understanding the product information inference process in electronic word-of-mouth: An objectivity–subjectivity dichotomy perspective, *Information & Management, Volume 46, Issue 5*, 2009, pp. 302-311.

[33] Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury, Micro-blogging as Online Word of Mouth Branding, *Proceeding CHI '09 Extended Abstracts on Human Factors in Computing Systems*, Boston, 2009, pp. 3859-3864.

[34] Adams, Joe K., Laboratory studies of behavior without awareness, *Psychological Bulletin, Vol 54(5)*, 1957, pp. 383-405.

[35] Philip M. Merikle & Jim Cheesman, Current Status of Research on Subliminal Perception, *Advances in Consumer Research, Volume 14*, 1987, pp. 298-302.

[36] Geoffrey Underwood, Subliminal perception on TV, *Nature, Vol 370(6485)*, 1994.

[37] Comments 4X More Valuable Than LIKES, EdgeRank Checker, Nov. 15, 2011. Retrieved Dec. 15, 2012

[38] Sara Kehaulani Goo, Facebook: A Profile of its 'Friends', Pew Research Center Publications, May 16, 2012, Retrieved Dec. 15, 2012.