

Frequent Pattern Mining Method to Influence Social Network Users for Marketing in an Efficient Way

^[1] Yashaswini P R ^[2] Sanjay Kumar N V

^[1]M. Tech. IV sem, Dept. of CS&E, ^[2] Asst. Prof., Dept. of CS&E,

^{[1][2]}Kalpataru Institute of Technology

Tiptur, Karnataka

yashaswinipr.yash@gmail.com sanjaynv@gmail.com

Abstract - Social networks have become an important marketing tools for business to build brand pages to prompt their new products. If this medium is explored intelligently then the social network medium has a potential to provide many new ways to market the audience with the help of registered users indirectly, without knowing them. Social Network (Face book) has provided many tools for marketing purpose like groups, events, social ads. Fans' user-experience diffusion results in great marketing power that people never seen. Typically, some of the fans in group are influence users. They are market movers which mean they can influence others buying decisions. Businesses can affect online influence users by giving them extra benefits to turn them into spokesmen. When analyzing the users' scope of social networking, it can be concluded that the modern social communities influence in individual's private life. However, who is the influential user? What period of time is appropriate for information to spread? In this study, a framework based on frequent pattern mining is proposed to find the influence users as well as the proper time to spread information. It also presents a survey based research from users and organizations for finding their views on the tagged based marketing on the social networking website The one day 24-hour period can be divided into successive time segments. An influence transaction that contains fans' influence power will be defined in each time segment. After transactions being collected several days, helps in understanding the reason what are the barriers and what steps could be taken to make it effective for organizations, the frequent patterns can be found to deduce the proper time for influence users to spread information. The theoretical experiment is given to show how the proposed framework works.

Keywords-Social network mining; frequent item set; Social network influence; Marketing on social Network; Tagged Based Marketing.

I. INTRODUCTION

Social network websites, such as Face book, have attracted enormous users to perform various activities through the services websites provide. Basically, users in social networks will interact with each other by sharing opinions or exchanging messages [9]. There are two important characteristics can be found in social networks. The first characteristic is that people in social network can build lots of trust relations with others regardless of geographical locations [4]. Another important characteristic is that the trust relationship built with someone who we may have never encountered in person. These two characteristics spread of information or ideas [14]. Social network is a new information propagation model which the users are both information consumers and information producers [1]. People nowadays prefer spending more time in gossiping, playing games, browsing friend's profiles and chatting with their friends on social networking sites rather than

physically meeting. Success of the online social networks with growing trends has allowed organizations to penetrate this new domain for marketing their products.

Many social networks allow companies to market their product through different tools provided by them. Businesses realize the benefit of social networks and build their own fans group to post advertisement to prompt their products. In fans group, some of the fans post articles to share user experiences, and some fans add comments to reply those articles. When people browse these articles, the opinions or debates in posts may affect their buying decision. Now, not just a small number of people but more and more people consult the posts before they making products buying decisions.

Influence propagation is always an interesting subject in business [1] and the relative topics about influence have been studied by social scientists a long time [3, 5-6, 9]. One fundamental problem is how to identify the right influential users in social network. Not every fan's

posts have influence power. Businesses may want to know who the influence fans are in social networks [1]. The influence fans play the role as market-movers who can really affect buying decisions of social network users. If the influence fans could be identified, the business can target them and sell the product to them in a special price or even give them some free samples to make them to have more understanding about products. The influence fans will share their user experiences or comments that may affect other fans buying decisions making. That is, the business could plan their marketing strategies in social networks by turning influence fans as unofficial spokesmen and triggering a cascade of influence by which fans affect other fans or non-fans.[1] proposed a framework to find influential bloggers. Their method to identify influential bloggers is according to a set of collectable statistics. Some other mining methods are also proposed to identify influence users or do the research of information propagation problem [2, 16-18], however, none of them consider when the proper time is in the day to spread information. The information posted on fans group is easily faded out with time. This circumstance is often seen in those Face book groups that own enormous number of fans. Because too many fans want to share their opinions, a post will be flooded out in a very quick time. If a post is shared by an influence user but flooded out before it can be seen, the post won't have any influence power.

In this paper, we propose a framework based on frequent item sets mining method to find potential influence users as well as the proper time for influence users to spread information. The one day 24-hour period can be divided into successive time segments. An influence transaction that contains the fans' influence power will be defined in each time segment. After influence transactions being collected several days, the frequent item sets will be found to deduce the proper time in the day for the influence users to spread information.

This presents a survey based research from users and organizations for finding their views on the tagged based marketing on the social networking website The objective is to help the businesses to target right fans and encourage them to share helpful promotion comments on fans group at the right time. That is, even the posts will be flooded out in a short time, but we hope the right influence posts could be seen by the most online users at the right time.

II. RELATED WORKS

This section describes works related to data mining in social networks.

A. Social Networks

A social network can be seen as a graph structure that is composed by nodes and links which represent relationships between nodes. Through the social networking web sites, social network users share their opinions or post their products reviews or ratings. One important characteristic is these messages can be exchanged or spread very quickly. These social networking sites can be seen as new type of social media [9]. Since social network users increase a lot in these years, social networks have got businesses attention.

Businesses utilize this new type of media to build virtual community to initial activities to promote products, to communicate with their fans, and to collect fans data to analyze users' behavior.

B. Data Mining in Social Networks

Because businesses can analyze the data to derive implicit information to help the marketing strategies adjustments. However, to analyze the social network data is a very challenge work since the social media data are vast, noisy, unstructured, and dynamic [9]. Data mining methods are also good choices to analyze social network data because Data mining is the concept of aiming to analyze large amount data for discovering implicit but useful knowledge. is tics of social network data, several literatures have invented new data mining methods for social network data mining. Those new data mining methods have applied on community analysis, users' sentiments sensing, recommendation systems improvement, predict social links, and influence propagation exploring, etc.

C. Social Networks Influence Propagation

The aforementioned paragraph has stated the social networks are the platforms for information spread. Since some of the users' post have influence power that could affect others buying decisions, the influence will also be spread to everywhere in the networks. With the growth of social network users, businesses have realized that social network sites are also marketing channels and adjust their marketing strategies toward social networks. Businesses never want their resources to waste on ineffective marketing activities no matter these activities held in physical market or virtual community. Hence, a fundamental problem is whether there exist the influential users, and how to find them. If businesses can identify the influential users and make special marketing strategy for them, businesses will benefit for the influence propagation..

D. Frequent Item sets (Patterns) Mining

Assume that D is a transaction database. The set of all items is denoted as I . A transaction in D is denoted as t . A set $X \subseteq I$ is called an item set.

The set X 's support can be calculated as: $\text{support}(X) = |X \subseteq t|/|D|$. An item set is frequent if its support is greater than or equal to a pre-defined threshold min_sup . That is, a frequent item set must be contained in a certain number of transactions. The Apriori-based algorithm is the first work to find frequent item sets. However, the Apriori-based algorithm is inefficient because the frequent item sets are found by repeatedly scanning the database. Several non-Apriori methods have been proposed to solve the inefficient problem on mining frequent item sets [10, 13, 19]. FP-growth [10] is a well-known and efficient non-Apriori algorithm. It scans the database twice to build an FP-tree structure to store frequent item sets. Finding a frequent item set in FP-tree is easy since each branch in the FP-tree is a frequent item set. Because the FP-tree is a compact structure, the performance to track frequent item set is better than those Apriori-based algorithms [8, 11]. In this study, a FP-tree will be constructed in order to have better performance in tracking frequent patterns.

E. Influential in Buying

Influence of any person close to an individual has some impact on his life and this influence can also be seen in decision making, either it is of choosing any house hold item, selecting a career or anything else. In [19], the researcher has discussed that purchasing decisions are strongly influenced due to trust among people and also if people know each other. Many online shoppers wait for opinions about the product before purchasing anything. Ecommerce companies convinces people to give feedback of the products by writing reviews, rating etc. some e-commerce companies have also started to use online social marketing after analyzing the effect on the sales due to this activity. Further examined the impact of this social influence in terms of increase in the sales and revenue. Analyzing the user behavior on the internet is one of the most critical factors that help in making the website successful. In [12] the researcher has emphasized organization to understand the user behavior when they login to the social networking website.

Understanding user behavior would help organization in evaluating the design and performance of the website; this could help in placement of the advertising on the website for better visibility with no annoyance to users.

F. Marketing Strategies

Every company either its small, medium or large enterprise have their objectives, goals and plans. For achieving their goals, these companies do have to adopt

some strategies or techniques. Most successful companies strictly follow their objectives, goals for their survival. In new era internet has become a basic element in individual's life and many companies are focusing the internet for marketing their products. Some of the optimal strategies are described in the paper [16]. In this paper the researcher has described about the intelligent selling strategies that could be applied on social networks for influential marketing and also emphasizes that for increase sales, seller should provide benefits to their potential buyers.

G. Privacy Issues

Privacy concern among internet users is the major barrier in the success of the e-commerce and social networking websites. Though large companies like Amazon, eBay, Face books have taken measures in making the users information private but somehow they cannot guarantee 100% privacy. In [14] researcher has described about the privacy issues in the online social networking websites. In social networking sites, privacy is the main concern among users. Current social network platforms don't provide any mechanism to control the information leakage of user's personal information. The researcher also described that the most needed privacy on the social networking sites are profile privacy, application privacy and newsfeed privacy. The researcher has also proposed the modifications in the existing privacy platforms that help user sharing their information with the third party applications

III. TAGGED BASED MARKETING INSTEAD OF OTHER MARKETING TOOLS

Face book has provided many tools for the small and medium sized enterprise organizations for marketing their products and increase clientele using Face book as a domain. Just like creating groups, events, Profile pages etc. yet they all have certain limitations.

i) Profile Page

For penetrating any domain, the basic thing is to establish your existence. Profile pages are the entry point for any organization to show their presence on Face book. The profile page is just a page containing your information. On profile page the organization can share the promo picture, videos, introduce new brands to target the audience

ii) Groups

Groups are the most simplest and the oldest way for marketing your company or any brand on Face book. Groups allow the organization to create a place for customers and friends so that they can participate in the discussions on the Face book. Groups are mostly used for viral marketing. There are some limitations in using groups

- Groups are just like private chat rooms, only the persons added to the group can participate in the activities.
- Group does not support large number of members.
- Groups doesn't allow the integration with other Face book

iii) Events

Face book events is the application mostly used by organization to promote their events, product launches etc. when you create an event that means you are creating a page which is similar to the groups in which you can start discussions, share photos, videos and links.

iv) Face book Pages

Face book pages are a way for the many organizations to create their presence on Face book. Face book pages are like groups with some differences

- ❖ Face book pages allow organizations to integrate their page with other applications to extend their functionality and analyze the user presence.
- ❖ Face book pages also support the News Feed, which indirectly invites Friends of your page fan, to become the member.

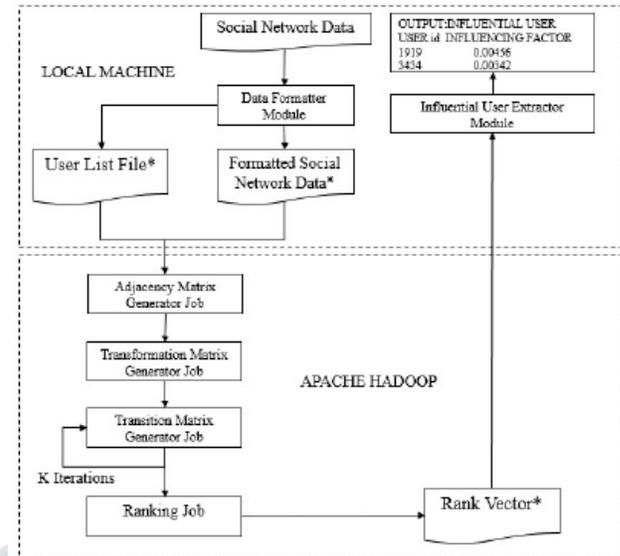
IV. TAGGED BASED MARKETING

Tagged based marketing is the extension of the Face book pages. This strategy will be helpful for the company to market their product, brand. Tagged based marketing allows the organization to tag the feedback of the product to their customer. Once the tag has been done then this tag will then be visible to the Friend List of the customer in the News Feed.

The benefit of the tag based marketing is: Suppose an organization take feedback of any product from the consumer and with his/her permission the feedback is tagged on the consumer profile. This tag will now be visible in the News

Feed of the Friend List of the consumer, which helps in the influential marketing; trust of the product will be increased in the consumer circle by consumer's influence. This is the indirect marketing that the organization would be doing with low cost as compared to the other marketing domain. The major benefit of this marketing is the trust of the product which is very important in today competitive market.

V. DESIGN AND ALGORITHM TO FIND INFLUENCE USERS



Introduced how to find influence users in social networks as well as the proper time for influence users to spread information. Here not invented new data mining method but developed a framework based on frequent patterns mining method. Following is the description of our proposed framework.

An influence user is thought as his messages or opinions posted on group wall will affect other users buying decisions.

Intuitively, could think that if the more influence power a user has, the more quickly and widely he can spread his opinions to other users. Before we can evaluate a user's influence power, we calculate his influence range and influence strength first.

1. Divide one day 24-hours period into successive time Segments $T = \{ t_i \mid i=1 \text{ to } K \}$ is the successive time segments set. The length of the time segment depends on the application needs.
2. Get online fans sets N_i for each t_i .
3. Calculate influence power p_j for each n_j in N_i . If a data pair (n_j, p_j) is seen as an item, the influence transaction TR_i can be defined in this step for each t_i . D is the transaction dataset for all TR_i .
4. Remove some items that their influence powers are less than a pre-defined power level, if necessary. TR_1 is denoted as a reduced influence transaction. D_1 is the reduced dataset.

5. Find L-frequent patterns from D1. L is a pre-defined value and $L \geq 2$. 6. Check the L-frequent patterns to identify who the influence fans are. 7. Compare each TR1 with L-frequent patterns to find when the proper time to spread messages is.

VI. EXPERIMENT RESULTS AND DISCUSSION

One experiment is given to evaluate our works on influence fans finding. The test is to prove that our proposed method can find the right fans to spread messages to affect users in social network at the right time of the day. In this test, a single time segment length is set to four hours. Hence, t_1 is the time period started from 0:00 to 4:00, t_2 is started from 4:00 to 8:00 and t_3 is started from 8:00 to 12:00. That is, T will be the set of $\{t_1, t_2, t_3, \dots, t_6\}$.

We assume the following conditions for this test.

1. N is $\{n_1, n_2, n_3, \dots, n_{10}\}$ and this test will collect transactions for three days.

2. Assume that we have gotten the number of each fan's outer friends and inner friends. The number of each post's thumbs up, comments, and sharing are gotten from fan's insight data.

3. **Equation** $R(n_j) = wr1 * |OGF| + |IGF| / wr2$ is used to get a fan's influence range. The weight number $Wr1$ is set to 1. The weight number $Wr2$ is set to 2.

4. **Equation:** $S(n_j) = ws1 * tup(n_j) + ws2 * cmt(n_j) + ws3 * shr(n_j)$ is used to get a fan's influence strength. The weight numbers for the factors of thumbs-up received, comments received and sharing received are all set to 1 since we think those factors are equally important.

5. The Table I is adopted to derive influence power of all fans.

6. The item with low influence power needs to be removed from each transaction. Table II shows the final collected online fans set. For each online fan, his influence range can be calculated according to (2) and influence strength can be calculated according to (3). After all data pre-processing work, Table III is the reduced dataset IH. Let the minimum support is 6, the 3-frequent pattern $\{(n_2, ph), (n_6, ph), (n_8, pm)\}$ can be found in the reduced dataset IH. That is, the fan $n_2, n_6,$ and n_8 are influential users. Next, comparing the 3-frequent pattern with each transaction, one can find the proper time for influence fans to spread their comments or opinions are t_1, t_4, t_5 and t_6 .

The experimental result show us if the comments posted on right time, the business could get the better influence propagation effect.

One experiment is given to evaluate our works on influence fans finding. The test is to prove that our proposed method can find the right fans to spread messages to affect users in social network at the right time of the day.

Table I. Influence Power Evaluation Table

Influence Strength	Influence Range		
	small	medium	big
low	P_l		
medium		P_m	
high			P_h

Table II. Online Fans Set Collected For Three Days

time period	Online Fans Set
t_1	$\{n_1, n_2, n_6, n_8, n_9\}$ $\{n_2, n_4, n_8, n_9\}$ $\{n_2, n_6, n_7, n_8, n_9, n_{10}\}$
t_2	$\{n_2, n_3, n_7, n_8, n_9\}$ $\{n_3, n_4, n_8, n_9\}$ $\{n_3, n_4, n_5, n_9\}$
t_3	$\{n_1, n_5, n_7, n_8\}$ $\{n_1, n_5, n_7, n_8\}$ $\{n_1, n_3, n_5, n_7, n_8\}$
t_4	$\{n_2, n_6, n_8, n_9\}$ $\{n_2, n_8, n_9\}$ $\{n_2, n_6, n_8, n_9\}$
t_5	$\{n_1, n_3, n_5, n_7\}$ $\{n_1, n_2, n_5, n_6, n_7, n_9\}$ $\{n_1, n_2, n_5, n_6, n_7, n_8\}$
t_6	$\{n_1, n_2, n_5, n_7\}$ $\{n_1, n_2, n_4, n_6, n_8, n_9\}$ $\{n_2, n_3, n_7, n_6, n_8, n_{10}\}$

VII. CONCLUSION

Social network is an information propagation model Information or ideas spread in social networks more easily than in physical world. Some information can influence social network users' buying decision making. Businesses realize they must use social networks as marketing channels to boost their sales volume. That is why so many companies build their Face book fans group for marketing purpose. In order to have efficient marketing activities, business always want to know who the right influential users are in their fans group. If the influence fans could be identified, the business can target them and turn those influence fans as unofficial spokesmen and triggering a cascade of influence by which fans affect other fans or nonfans. In this paper, we propose a framework based on frequent item sets mining method to find potential influence fans. We even consider that the proper time in the day for influence fans to spread is a very important subject because the post is very easily flooded out and if the message is posted at wrong time, it may not

be seen by anyone. It focuses on the product review of customers on social networking website to promote marketing and uses quantitative research methodology. To help understand the users and organizations behavior, a survey is being conducted from social network users and organizations. The survey consists of separate questionnaires for users and the organization. The proposed framework divides one day 24-hour period into successive time segments. Influence transactions will be defined in each time segment. An influence transaction dataset is used to get frequent pattern which will be used to deduce the proper time in the day for the influence users to spread information. The objective is to help the businesses to target right fans and encourage them to share helpful promotion comments on fans group at the right time. The study reflects that with the increasing growth of Social Networking websites, people opt to share more and more information and this growth has open a way for the to penetrate the new domain. The research finding consists of two phases: user and organization. While conducting survey from users, we found that most of the people support marketing on social website In the future, we will build an experimental Face book fan group page to collect fans' data and apply our framework to identify influence fans. We hope our proposed framework can really help the businesses in fan group page engagement.

REFERENCES

- [1] N. Agarwal, H. Liu, L. Tang, and P.S. Yu, (2008, February). "Identifying the influential bloggers in a community," in Proc. of the Int. Conf. on Web Search and Data Mining, pp.207-218, February,2008.
- [2] S. Aral, L. Muchnik, and A. Sundararajan, "Distinguishing influencebased contagion from homophily-driven diusion in dynamic networks," in Proc. of the National Academy of Sciences of the United States of America, vol. 106, no. 51, pp.21544- 21549, October, 2009.
- [3]] J. Brown and P. Reinegen, "Social ties and word-of mouth referral behavior," Journal of Consumer Research, vol. 14, no. 3, pp.350-362, 1987.
- [4] K. Burke, "Network to drive revenue," Target Marketing Magazine, vol. 29, no. 2, pp.25-26, February 2006.
- [5] P. Domingos and M. Richardson, "Mining the network value of customers," in Proc. of the 7th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, pp.57-66, August, 2001.
- [6] J. Goldenberg, B. Libai, E. Muller, "Talk of the network: a complex systems look at the underlying process of word-of-mouth," Marketing Letters, vol. 12, no. 3, pp.211-223, 2001.
- [7] A. Goyal, F. Bonchi, and L.V. Lakshmanan, "Discovering leaders from community actions," in Proc. of the 17th ACM Conf. on Information and Knowledge Management, pp.499-508, October, 2008.
- [8] G. Grahne, and J.F. Zhu, "Fast algorithms for frequent item set mining using FP-Trees," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 10, pp.1347-1362, October 2005.
- [9] P. Gundecha, and H. Liu, "Mining social media: A brief introduction," in Proc. INFORMS TutORials in Operations Research, Phoenix, Arizona, USA, pp.1-17, October 14-17, 2012.
- [10] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in Proc. of ACM SIGMOD Int. Conf. on Management of Data, Dallas, Texas, USA, pp.1-12, May 2000.
- [11] T. Hu, S.Y. Sung, H. Xiong, and Q. Fu, "Discovery of maximum length frequent itemsets," Information Sciences, vol. 178, no. 1, pp.69-87, January 2008.
- [12] H.-W. Hu, and S.-Y. Lee, "Study on influence diffusion in social network," Int. Journal of Computer Science and Electronics Engineering, vol. 1, no. 2, pp.198-204, 2013.
- [13] Y.-P. Huang, L.-J. Kao, and F.E. Sandnes, "Efficient mining of salinity and temperature association rules from ARGO data," Expert Systems with Applications, vol. 35, no. 1-2, pp.59-68, Aug. 2008.
- [14] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, pp.37-146, August, 2003.
- [15] A. Porterfield, "5 new studies show Facebook a marketing powerhouse," Social Media Examiner, March, 2010.
- [16] D. Qiu, H. Li, & Y. Li, "Identification of active valuable nodes in temporal online social network with attributes," International Journal of Information Technology & Decision Making, vol. 13, no. 4, pp.839-864, April 2014.
- [17] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in Proc. of

the 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data mining, pp.61-70, August, 2002.

[18] M.F. Tsai, C.W. Tzeng, Z.L. Lin, & A.L. Chen, "Discovering leaders from social network by action cascade," *Social Network Analysis and Mining*, vol. 4, no. 1, pp.1-10, January 2014.

[19] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. McLachlan, A. Ng, B. Liu, P. Yu, Z.-H Zhou, M. Steinbach, D. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp.1-37, January 2008.

