Modeling Social Influence in Mobile Messaging Apps

Songmei Yu[✉] and Sofya Poger

Felician University, Lodi, NJ 07644, USA
{yus,pogers}@felician.edu

Abstract. Social influence is the behavioral change of a person because of the perceived relationship with other people, organizations and society in general. With the exponential growth of online social network services especially mobile messaging apps, users around the world are logging in to messaging apps to not only chat with friends but also to connect with brands, browse merchandise, and watch content. Mobile chat apps boast a number of distinct characteristics that make their audiences particularly appealing to businesses and marketers, including their size, retention and usage rates, and user demographics. The combined user base of the top four chat apps is larger than the combined user base of the top four social networks. Therefore, it makes great sense to analyze user behavior and social influence in mobile messaging apps. In this paper, we focus on computational aspects of measuring social influence of groups formed in mobile messaging apps. We describe the special features of mobile messaging apps and present challenges. We address the challenges by proposing a temporal weighted data model to measure the group influence in messaging apps by considering their special features, with implementation and evaluation in the end.

Keywords: Social influence · Mobile messaging apps · Influence modeling

1 Introduction

Social influence refers to the behavioral change of individuals affected by others in a network. Social influence is an intuitive and well-accepted phenomenon in online social networks. The strength of social influence depends on many factors such as the strength of relationship between users in the networks (nodes), the network distance between users (edges), temporal effects (time), characteristics of networks and individuals in the network. In this paper, we build a temporal weighted data model that focuses on influence analysis in mobile messaging apps, using groups (communities) randomly formed in mobile messaging environment and individual behaviors in the groups, and describe the measures and algorithms related to group and individual characteristics analysis over a certain time period. Furthermore, we implement this model by applying it into real dataset, and evaluate its application within a given dataset.

First we survey the related work in traditional social networks. We point out the limitations of current existing models that did not give enough weight to special characteristics of mobile messaging apps, which leads to our weighted model to

integrate these factors into the existing techniques. Secondly, we propose the temporal weighted data model for social influence evaluation in mobile messaging apps by integrating special features. We use a bipartite graph to represent users and groups, add weights to different features of groups and users, and compute the value of influence at certain time period, as well as identify the change of this influence value over time periods. Thirdly, we discuss the applications of maximizing influence in mobile messaging apps by utilizing our proposed data model, on both group level and individual level. We implement and evaluate the model by using real data from WeChat. The paper is concluded with future work.

2   Related Work and Challenges

A social network is modeled as a graph G = (V, E), where V is the set of nodes, and E is set of edges which is a subset of V × V. The nodes correspond to users (people) and the edges correspond to social relationship. At the local level, social influence is a directional effect from node A to node B, and is related to the edge strength from A to B. On the global level, some nodes can have intrinsically higher influence than others due to network structures. These global influence measures are often associated with nodes and edges, which are presented as (a) Edge Measures, (b) Node Measures, and (c) Social Similarity, Influence and Actions. All of these factors were studied thoroughly [1, 2, 5, 6, 8, 10]. In this section, we will briefly discuss their limitations when applied to mobile messaging applications.

2.1   Special Features of Mobile Messaging Apps

The rise of mobile social messaging apps will increasingly threaten the influence and stickiness of conventional social networks. It is a potentially transformative trend that will see increasing numbers of app developers and business holders looking to re-orientate their social integration strategies to facilitate messaging and content sharing with communities outside of Facebook, Twitter and other more established social networks.

The leading service of this kind, WhatsApp, has in excess of 350 million active users around the world, with services such as WeChat and Line also boasting hundreds of millions of users. Traditional social networks and mobile messaging apps both provide platforms for people to connect and share content, so what distinguishes the two?

The first difference is regarding audience size, duration, and intent. A messaging app acts primarily as a one-to-one (or one-to-few) communication mechanism, and can be temporary or long-lasting. Content is intended to be private, or at least directed towards a specific group, not to the public. Whereas a social network consists of "many to many" connections, is durable, and is capable of producing network effects. Content is essentially public. Second, a user can create as many accounts as he/she wants in a traditional social network, but only one account that is based on cell phone number can be created in mobile messaging apps. Thirdly, groups/communities in traditional social network are predefined by the site administrator, and group admin account(s) are also

designated by the site administrator. Those groups and admin members could last for months or years. However, in mobile messaging apps, any account can form a group at any time that attract people who are interested to join and therefore assign itself as an administrator, or dissolve a group at any time. Time is essential in mobile groups. Finally, in traditional social network, an account can join any group by just posting or replying to other posts. There is no formal way to join or quit a group. In mobile messaging apps, an account needs to be invited to join a group by the group admin, or an existing group member. On the other hand, an account needs to formally quit a group by signing out or being removed by the group admin. In this way, groups in mobile messaging apps have more control on information diffusion than traditional social networks.

2.2   Existing Work and Challenges

Social network analysis often focus on macro-level models such as degree distributions, diameter, clustering coefficient, communities, small world effect, preferential attachment, etc.; work in this area includes [1, 11]. Recently, social influence study has started to attract more attention due to the popularity of many important applications such as Twitter and Facebook. Many of the works on this area present qualitative findings about social influences [14]. Some researchers studied the influence model quantitatively, focusing on measuring the strength of topic-level social influence. However, few research work have been done on dynamic formed group level in mobile messaging systems, by analyzing behavior characteristics of group and users, which posted challenges in answering certain key questions such as (1) Which group(s) has dramatic influence change during certain time period? (2) For certain topics, which group has more influence than others among all groups relate to the same topic? (3) How to identify and connect to a particular influential user through member invitation mechanism in mobile messaging apps? Above all, the key question is how to quantify the influence in mobile messaging apps of both groups and users.

2.3   Our Contributions

Based on the unique features of mobile messaging systems, we propose a temporal weighted data model in a quantitative way that tackles the special features of mobile messaging apps. In particular, our model integrates the following factors into a mathematical model: (1) the average number of posts within a group per day during a certain period T, (2) the average number of members in a group giving its limited member size during T, (3) the ratio of number of posts to the number of members within a group during T, (4) the diversity of posting members within a group during T. *i.e.* whether the posts are coming from only one or two members or from majority members in the group. We can use the total number of members who have at least one post to measure this attribute, and (5) the interactive ratio, i.e., the ratio of original posts to replying posts from other members. This is to measure interactivity within the group. More formally, given a mobile social network for certain topic(s), we construct a social influence graph that quantitatively computes the influence level for each group during a specified time period.

3   A Temporal Weighted Data Model

3.1   Problem Formulation

In social network, bipartite networks or affiliation networks are bipartite graphs that are used to represent the members and the groups/communities they belong to. In our problem, we define the relationship between members and member-created groups as a special bipartite graph: there is an edge between a user u and a group g, if the edge is a directional edge, it means the user is a creator/admin of this associated group, else this user is just a regular member of this group. Figure 1 shows the change of group creation and member association of each group from time T_1 to time T_2. We can see the time is essential as the status of groups and associated members of each group is dynamically changing, new groups are constantly created (group D is created in T_2), some old groups could be dissolved, and members are constantly added or dismissed from each group.
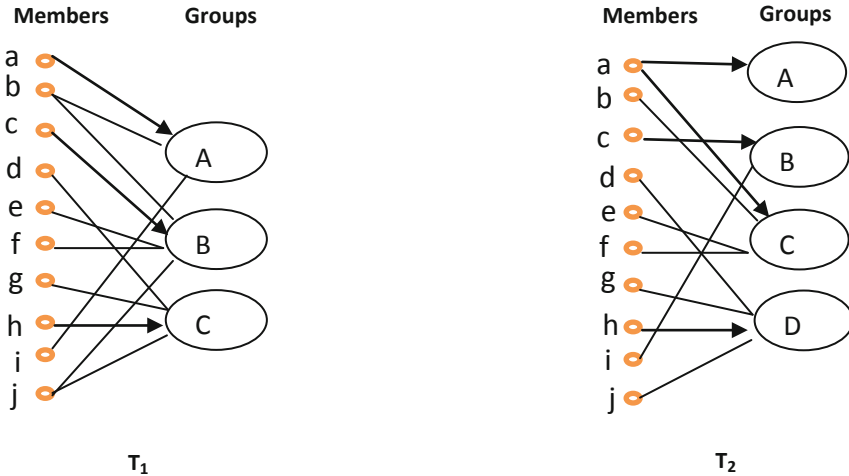

Fig. 1. The change of users/groups in the bipartite networks

The goal of social influence analysis in mobile messaging apps is to derive the *group level* social influence based on the features of topic distribution during a certain time period *t,* as well as the *individual level* of influence by constructing sub-graphs within a group. Weights will be considered for each factor. Here we assume each group is a topic related group, which means the topic is given for each group and all posts are related to the given topic. In this paper, we focus on the group level social influence computation.

We consider the following factors as special features associated with each group. Social influence in a mobile messaging group is primarily determined by these factors.

1. Number of posts
2. Number of members
3. Ratio of posts to members
4. Ratio of original post to replying posts from other members
5. Total number of members having a least one post

Table 1 shows the notations we use in our model:

Table 1. Notation

Symbol	Description
G	A single group
T	A time period a group G exists
N	Number of posts in G during T
M	Number of members in G during T
L	The upper limit of member size in G
O_i	Number of original posts from member i in G during T, $i \in Q$
R_{ij}	Number of replying posts from member j to member i in G during T, $i \in Q$, $j \in Q$
P	The set of posts in G during T
Q	The set of members in G
S_i	Status of a member i, $i \in Q$: $S_i = 1$ for admin, $S_i = 0$ for regular members
μ_i	The social influence of member i in G, $i \in Q$
μ_G	The social influence of G

Based on the above concepts, we can formalize the problem as follows:

Problem. Given a network $G = (P, Q, T)$, where P is set of relations (original or replying posts) within Q, Q is a set of nodes (members), and T is a specified time period that G is alive, how to (1) identify the influence level of this network G during T, (2) measure the influence level of each node in Q during T, and (3) evaluate the influence change of G and Q as T changes and therefore achieve the influence maximization at certain point.

3.2   Our Approach

The social influence analysis problem poses challenges in setting the weight for each factor associated with a mobile messaging group. For some factor, the value is dynamically changing during the time interval T. Hence setting weight for each factor to capture how it affects the social influence level of a group or a member within the group is challenging, as we are monitoring a dynamic changing entity with some factors' value are changing constantly.

Determining the Time Period T. Based on tour observation of datasets we collected from WeChat, an average life expectancy of a group is 9–12 months, from the time it is created to the time it is dissolved or it becomes quiet with no posts or very few posts posted. Of course there are exceptions that some groups are lasting for years and some

groups are only alive less than 1 month. Therefore, it is essential to set up the time period based on the datasets collected.

Identifying Influence Factors. For certain factors, the value are constantly changing. For example, if we set up T as one year, it is easy to count the number of posts within a group as we can just add all posts starting from the beginning to the end of the year. However, counting number of members is a little different as some members may leave or join the group during the middle of this specific month. Since we count all posts during this year, we need to count all members accordingly during this year even some of them leave or join in the middle. Similarly, we need to count the total original posts or replying posts from each member even this member leave or join in the middle.

We identify the following factors as influence factors of a group G based on the special features of mobile messaging apps:

- Post Size **N**: the total number of posts in G during T
- Group Popularity **M/L**: the current member size in G during T over its member size upper-limit L
- Post Liveness **N/M**: the total number of posts over the current member size in G during T
- Post Diversity **D**: the diversity of G which indicates the total number of members in G during T that has at least one post
- Post Interactivity **I**: the total number of replying posts from member j for each original post from member i ($i \neq j$, $i \in Q$, $j \in Q$)

Determining Influence Weight Vector. Each factor is considered as one dimension/attribute of social influence for a group. Compared to others in the same dataset, a high influential group can normally attract more people to join, has more posts during a specific time period, a member will be more active to post comments and talks to each other, and the interactivity ratio will be higher too. However, these dimensions may have covariant relationship with each other, for example, a higher value of member size will produce a lower ratio of posts to members, which will affect the group influence level negatively.

We use an *Influence Weight Vector* to determine how each factor contributes to the overall influence of a group within a given dataset that composes of several similar groups. It quantifies how much positive or negative influence a factor affects the total influence level of a group. The bigger the value, the higher influence the factor contributes to the group influence level.

Definition: an *Influence Weight Vector* is a row vector V_T for a reference dataset during a specified time period T, $V_T = [W_N, W_{M/L}, W_{N/M}, W_D, W_I]$, where W_N is the weight of the factor that is the total number of posts in G during T, W_M is the weight of factor that is the number of members given its limited size (M/L_M), $W_{N/M}$ is the weight of factor that shows the ratio of total posts to total members, W_D is the weight of factor that shows the diversity of posts in G, and W_I is the weight of factor that describes the ratio of original posts from one member to replied posts from other members.

Now we describe how to compute each weight in V_T. Assume we collect a dataset which has n mobile messaging groups (G_1, G_2, ..., G_n) during T, and G_k is one of them

that we are interested to measure its influence level where k = 1, 2, …, n. We call this dataset as a reference dataset. Generally all groups in this dataset should have similar topic(s), hence it makes sense to compare their social influence levels.

$$W_N = 1 - \frac{\sum Nk/n}{Nmax} \quad \text{where } k = 1, 2, \ldots, n. \tag{1}$$

N_k is the total number of posts within group k, $\sum N_k/n$ is the average number of group posts in reference dataset, N_{max} is the maximal number of posts that a group has in the reference dataset. We use 1- (the ratio of average number of posts to the maximal number of posts) to show the trust or importance of this factor on group influence. The rational is that if the average is close to maximum, W_N will decrease. We would think the variance of each group's post numbers is small, hence the total number of posts of each group would not be able to significantly indicate the influence level difference of each group.

$$W_{M/L} = \frac{\sum Mk/n}{L} \quad \text{where } k = 1, 2, \ldots, n. \tag{2}$$

Given the group size *L* that is set by the group admin (there could be default size for some mobile messaging apps), the ratio of current average member size in the reference dataset to the size limit shows how populated the group is. The bigger the value is, the higher the impact is. When it is 1, this factor reaches the maximum impact on the influence level.

$$W_{N/M} = \frac{\sum Nk / \sum Mk}{Max\left(\frac{Nk}{Mk}\right)} \quad \text{where } k = 1, 2, \ldots, n. \tag{3}$$

The ratio of total number of posts to the total numbers in the reference dataset shows on average how many posts each number. By dividing by the maximum value of posts per member of all groups in the dataset, this weight shows relatively how members contribute the total posts. The bigger the value is, the higher the impact is, as it shows each member on average has more posts to influence others.

$$W_D = \frac{\sum Mik}{\sum Mk} \quad \text{where } i \in Q_k, k = 1, 2, \ldots n. \tag{4}$$

W_D shows how diversified the posts are in a reference dataset. $\sum M_i$ computes the total members in the dataset who has at least one post, whereas $\sum M_k$ just computes the total members in the dataset. The ratio calculates the percentage of members who do the posts. For some group(s), only one or two persons are posting, others are silent, even there are a large number of posts, it is not influential due to low diversity. The higher the value of W_D is, the more people involved in discussion, then we would think the higher the influence of this factor could be.

$$W_I = \frac{\sum \text{Oik}}{\sum \text{Rijk}} \quad \text{where } k = 1, 2, \dots n, i \in Q_k, j \in Q_k, i \neq j \tag{5}$$

The ratio of original posts to the replying posts shows the interactivity of a chatting group, which plays an important role to determine the influence level. O_{ik} calculates the number of original posts from certain user i in group k, R_{ijk} calculates the number of replied posts from users other than i in group k. When we add all original and replying posts in the reference dataset and find out the ratio, we set up a threshold that measures the weight of interactivity. There could be a case where a lot of users post, but very few reply, which shows low response on the topic posted, therefore there is low influence in the group. The higher the threshold is, the more important the interactivity is in determining the group influence level.

Compute the Influence Level of a Group (μ_G). After we compute the influence vector V_T, we come up an algorithm to calculate the group social influence level μ_G. The bigger the value of μ_G is, the higher the social influence that this group has within a specific mobile messaging community.

Algorithm. Computing the Social Influence Level of all groups in a given reference dataset during period T. We assume the dataset has k groups which are focusing on similar topics.

Input: $G_k = (P, Q, T)$, $k = 1, 2, \dots, n$
Output: μ_G for G_k, $k = 1, 2, \dots, n$

1. Calculate the influence weight vector of the selected reference dataset that includes groups G_k ($k = 1, 2, \dots, n$): $V_T = [W_N, W_{M/L}, W_{N/M}, W_D, W_I]$

2. For ($k = 1, k < n, k\,{+}{+}$):
　　　Collect resource data of each group G_k: [N, M/L, N/M, D, I]
　　μ_G　= $[N, M/L, N/M, D, I]^T$ x V_{GT}
　　　　= N x W_N + M/L x $W_{M/L}$ + N/M x $W_{N/M}$ + D x W_D + I x W_I

4　Model Implementation and Evaluation

We collected several groups from WeChat which have similar discussion topics, and built statistics for this bipartite network accordingly. This is our reference dataset. By analyzing these groups, we expect to find which group has higher social influence during each period, and reach a clear picture to view the influence level change of each group over time periods. Ideally, we could identify the maximized influence level at certain point for each group and use the pattern to predict the future trend if possible. This result makes great business sense to attract sponsorship, donation and advertisement for the groups.

Specifically, the dataset we collected are three groups which focus on elementary education discussion in WeChat in 2018, all of which have 500 upper limit of member

size (a default WeChat group size assigned by the app). We want to compare their influence levels on the group level. We collected all needed data of each month for each group within three months, including post counts by differentiating original posts and comments, member counts, posts per each member, etc. Then we calculated the influence weight vector of this dataset, and finally calculated their social influence levels of each month accordingly.

We implemented this process using Python and reached the result as follows (Tables 2 and 3):

Table 2. Source data

Source Data	Jan 2018					Feb 2018					March 2018				
	N	M	L	D	I *	N	M	L	D	I*	N	M	L	D	I*
Group 1	356	112	500	42	120/236	702	228	500	120	122/600	1005	234	500	124	350/655
Group 2	1242	201	500	86	301/941	980	178	500	98	560/420	789	145	500	120	203/586
Group 3	1003	450	500	359	240/763	1345	456	500	400	506/839	1809	500	500	324	402/1407

*We use the number of original posts/the number of replying posts to represent group interactivity I.

Table 3. Calculated results

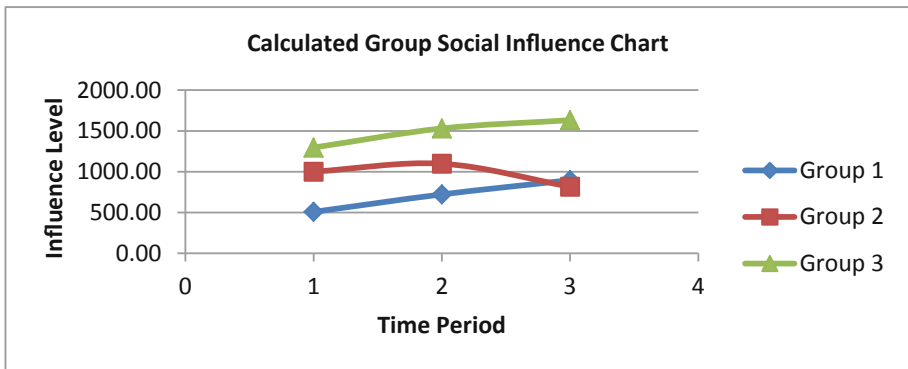|		Jan 2018	Feb 2018	March 2018
Dataset	Influence weight vector ** [W_N, $W_{M/L}$, $W_{N/M}$, W_D, W_I]	[0.14, 0.51, 0.55, 0.64, 0.34]	[0.25, 0.57, 0.64, 0.72, 0.65]	[0.34, 0.59, 0.75, 0.65, 0.36]
Group 1	Influence level μ_G. **	507.99	721.93	896.71
Group 2	Influence level μ_G. **	1000.79	1098.04	816.51
Group 3	Influence level μ_G. **	1295.99	1530.61	1631.99

Chart 1. Group influence levels

Chart 1 shows the influence level change for each group over three periods. We can see that group 1 is getting more influential over the time periods, group 2's influential level started declining from the second period and is below group 1's level at the third period although it has much more members in the beginning, and group 3 is gradually increasing the influential level and is always on the top among three groups. We could not identify the trend for each group's social influential level as more time periods needed, but at certain point we can point out which group has more influential level than others. Also, within each group, we can identify the maximized influential point within a certain time periods.

5   Conclusions and Future Work

Social influence plays an important role in social media information dissemination. Social behavior and how consumers think have conventionally been disseminated by media such as television, radio, newspapers and magazines, but in the 21st century, social media especially mobile messaging apps have begun to replace traditional media's enduring and influential role on consumers' behavior. In mobile messaging apps, people are interested in social influence on two levels, group level and individual level. In this paper, we focus on the group level. We propose a Temporal Weighted Data Model to measure the social influence level for a special group within a set of similar groups, and find out the maximized influence level for each group during a certain period. This model is implemented and evaluated in the end.

Our future work is to build a model to measure individual social influence in each group by constructing a sub-graph within the group. Moreover, based the proposed data model, we will discuss strategies to increase the influence in mobile messaging applications for both group level and individual level.

References

1. Albert, R., Barabasi, A.L.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74** (1), 47 (2002)
2. Anagnostopoulos, A., Kumar, R., Mahdian, M.: Influence and correlation in social networks. In: Proceeding of the 14th SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), pp. 7–15 (2008)
3. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
4. Buckley, C., Voorhees, E.M.: Retrieval evaluation with incomplete information. In: SIGIR 2004, pp. 25–32 (2004)
5. Chu, C.-T., et al.: Map-reduce for machine learning on multicore. In: Proceedings of the 18th Neural Information Processing Systems (NIPS 2006) (2006)
6. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S.: Feedback effects between similarity and social influence in online communities. In: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), pp. 160–168 (2008)

7. Craswell, N., de Vries, A.P., Soboroff, I.: Overview of the TREC-2005 enterprise track. In: TREC 2005 Conference Notebook, pp. 199–205 (2005)
8. Das, A., Datar, M., Garg, A., Rajaram, S.: Google news personalization: scalable online collaborative filtering. In: Proceeding of the 16th International Conference on World Wide Web (WWW 2007) (2007)
9. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. In: Proceedings of the 6th Conference on Symposium on Operating Systems Design & Implementation (OSDI 2004), p. 10 (2004)
10. Dourisboure, Y., Geraci, F., Pellegrini, M.: Extraction and classification of dense communities in the web. In: WWW 2007, pp. 461–470 (2007)
11. Flake, G.W., Lawrence, S., Giles, C.L.: Efficient identification of web communities. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2000), pp. 150–160 (2000)
12. Frey, B.J., Dueck, D.: Mixture modeling by affinity propagation. In: Proceedings of the 18th Neural Information Processing Systems (NIPS 2006), pp. 379–386 (2006)
13. Mei, Q., Cai, D., Zhang, D., Zhai, C.: Topic modeling with network regularization. In: Proceedings of the 17th International World Wide Web Conference (WWW 2008), pp. 101–110 (2008)
14. Papadimitriou, S., Disco, S.J.: Distributed co-clustering with map-reduce. In: Proceedings of IEEE International Conference on Data Mining (ICDM 2008) (2008)