User Departure Behavior Prediction in Social Group

Jiayi Liu, Qilong Han*, Yinlong Li
College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China
*Corresponding author: hanqilong@hrbeu.edu.cn

Abstract. We propose a departure behavior prediction algorithm to predict the user's departure behavior. The k-core decomposition is conducted on the graph. By considering the positive and negative affect users obtain, we calculate a score for each user. Finally, the user is predicted by the sliding window model. Experiments show that the DBP algorithm has a high prediction.

1. Introduction
With the rapid development and popularity of social networks, a variety of social applications has emerged and been widely used, such as Facebook, WeChat, Weibo, which allows people to share information anytime, anywhere. At the same time, it also brings relevant economic benefits and social influence to enterprises and individuals.

These social applications has a highly interactive feature: group. A group is a collection of users with common topics, interests, and needs. The members of the group can be your real friends or strangers you don't know. Facebook is currently the most popular social platform in the world, and its Facebook Group is one of its most popular applications. As the most popular social application of China, WeChat provides a convenient and fast way of group chatting without threshold. The social group represented by Facebook Group and WeChat group attracts a large number of users. Social group has become an important channel for frequent knowledge communication and dissemination.

We found that although the group provides a communication platform, the activity of members within the group will change over time. Some members may become inactive and leave the group because they are no longer interested in topics within the group. Or some members of the group such as classmates and work groups, leave the group due to graduation, resignation and other reasons.

For social apps, users are the most important resource, and they determine the social benefits and market values they can obtain. Nowadays, the common online marketing is the word-of-mouth effect. This kind of marketing means makes use of the intra-group communication effect more. If the user's departure behavior in the group can be effectively identified, some measures can be taken to effectively prevent users from leaving so as to obtain better economic and social benefits.

The existing research focuses on the dynamics of the formation and evolution of social networks, while there are few analyses on the decline of user activity in social networks and their departure from social networks. According to the current study, users' behavior of becoming inactive or even leaving their social circle is mainly influenced by their friends, and this effect will also have a cascading effect. On this basis, this paper proposes the prediction of users' leaving group behavior under the influence of friends and their groups.
2. Related Work
Numerous studies have found that activities among friends in the group are relevant (see the example in [1,2,3,4]). That is, if his friend does this, the user is more likely to take the same behavior [2,5]. A large number of users’ behaviors can be triggered by the behaviors of a small number of their friends [6,7]. Dasgupta et al. finds that the probability of user departure is correlated with the number of friends who have left by studying the departure behaviors of users in mobile phone networks [8]. Wu et al. has studied [9] the reasons why users leave social network. He found that users have the characteristics of “Advance and retreat together”, and users usually leave social network with their friends. The probability of a user leaving is related to the activity of his group, regardless of the size of the user’s group. The probability of a user leaving can be predicted by his friends and group.

Suppose user with k or more friends will stays in social network, while user with fewer friends ends up leaving the network. If a person leaves the network, it will affect his or her friends, which may cause their connections to drop below k. Therefore, based on the above assumptions, users with fewer friends thank will also leave the network. This effect spreads throughout the network and can lead to cascading exits, significantly reducing the number of users. Finally, there is a subgraph left in the network, where each node has at least k adjacent nodes, at this point, the unraveling process stops. This is a well-known concept in graph theory. This subgraph is called the k-core graph of the original network, which is unique because it does not depend on the order in which the nodes leave the original network. The k-core decomposition can capture each user’s social engagement [10]. Malliaros and Vazirgiannis [11] study the property of user leaving dynamics based on game theory, and use the core number of nodes to propose an easy-to-calculate metric for engagement. They find that nodes with larger cores in the graph show better engagement, so they are less likely to leave (or at least have less motivation to leave). Based on the above theory, in this paper, we propose a user departure behavior prediction algorithm based on topology structure.

3. Problem Statement
We present some basic definitions and formalize the problem in this section.

Social Network: A social network is defined as a graph \( G = (V,E) \). \( V \) is the set of nodes, where nodes represent users, \( E \) is a set of edges, where edges represent relationships between users. Moreover, each edge \( e(v_i,v_j) \in E \) is associated with a weight \( w \) indicating the degree of interactions between \( v_i \) and \( v_j \).

Social Group: A social group in a social network is defined as a set of users in that social network. Given a social network \( G = (V,E) \), a social group \( C \) is defined as a subgraph of \( G \), denoted by \( C \subseteq G \).

Dynamic Social Network: Let \( G = \{G_1,G_2,...,G_n,...\} \) be a dynamic social network, where \( G_1,G_2,...,G_n,... \) is a series of snapshot graphs. The social groups on the snapshot \( G_i \) are denoted as \( C_i = \{C_{i,1}^i,C_{i,2}^i,...,C_{i,n}^i\} \), where \( n \) is the number of social groups on \( G_i \). Note that \( C_{i,1}^i \) and \( C_{i,n}^i \) are the same community \( C_i \) at different timestamp.

In this paper, we utilize the k-core decomposition to capture each user's social engagement level, and more specifically the core number \( \text{core}(v_i) \) of each node \( v_i \in V \).

k-core: Given a graph \( G = (V,E) \), let \( H \) be a subgraph of \( G \), denoted by \( H \subseteq G \). Subgraph \( H \) is the maximal connected subgraph of \( G \), in which all nodes have degree at least \( k \).

Node's core number: The core number \( \text{core}(v_i) = k \) of a node \( v_i \) in \( G \) is the highest order of a core that \( v_i \) belongs to a k-core but not to any \( (k+1) \)-core.

The engagement level of each node \( v_i \in V \) is defined as its core number \( \text{core}(v_i) \).
4. Departure Behavior Prediction

In this section, we will present the definition we propose. And the social group benefit score used to measure the likelihood of a user leaving will be presented in this section. Finally, we introduce departure behavior prediction algorithm by using sliding window.

Malliaros and Vazirgiannis study shows [11] that the engagement of a user can be measured by the core number of its. Therefore, we proposal graph engagement as follows:

Definition 1(Graph Engagement): Let \( f(e) = \Pr(X \geq e) \) be the cumulative distribution function of the sizes of the node’s engagement. Then the total engagement level of a graph \( G \), denoted as \( \varepsilon_c \), is defined as the expectation of \( f(e) \).

The calculation formula of graph engagement is as follows:
\[
GE(G) = \sum_{i \in C} k_i p_i
\] (1)

where \( k_i \) is the core number of the nodes, \( p_i \) is the distribution probability of the core number \( k_i \), \( k_{\max} \) represents the maximum value of all the core numbers.

Wu et al. studies [9] that the fraction of inactive friends has the better predictive power on the departure probability. And the intensity of interaction between friends will also affect each user [2,13]. Malliaros and Vazirgiannis study shows [11] user with high engagement are less likely to leave. For this reason, we assume that all neighbors less than user’s engagement are users who leave the group.

Definition 2(Departure Effect): For any user \( v_i \), if \( v_i \) has neighbors that the core is lower than \( v_i \), the effect of these neighbors on it is called Departure Effect, denoted as \( DE(v_i) \).

The departure effect is calculate as follow:
\[
DE(v_i) = \frac{\sum \delta_i}{\sum w_i}
\] (2)

where \( \delta_i \) is the weight of \( v_i \), and the neighbors that the core is lower than \( v_i \), \( w_i \) is the weight of \( v_i \), and all neighbors, the weight is calculated by Jaccard similarity.

Malliaros and Vazirgiannis [11] believe that the user's benefits in the social network depends on their own decisions and its neighbor's, and that the user will choose to stay only if its benefits in the network increase. Garcia et al., through the research on the topological structure and k-core, believe that two important reasons for the loss of users are that friends leave and the network cost exceeds the benefit [12]. Based on the above theory, we proposal social group benefit. First we're assuming that all neighbors less than user’s engagement are users who leave the group. The remaining nodes compose the subgraph \( \overline{G} \) induced by \( G \).

Definition 3(Social Group Benefit): For any user \( v_i \), the effect of graph engagement and departure effect on \( v_i \) is called social group benefit, denoted as \( SDB(v_i) \).

The Social group benefit is calculate as follow:
\[
SGB(v_i) = GE(\overline{G}) + DE(v_i)
\] (3)

where \( GE(\overline{G}) \) represents the benefit of user \( v_i \) remaining in the graph \( \overline{G} \) even after his friends leave, and \( DE(v_i) \) is the probability that user \( v_i \) is affected by the departure of its friends.

---

**Algorithm**

**Departure Behavior Prediction Algorithm**

**Input:** A dynamic network \( G = \{G_1, G_2, \ldots, G_n\} \), the set of communities \( C_i \) for each \( G_i \), a sliding window with window size \( \tau \) and slide size \( \mu \), the parameter \( \eta \)

**Output:** The list of departure nodes
1: Set Departure $\leftarrow \{\}$;
2: Set $t \leftarrow 1$, $W_{start} \leftarrow t$, $W_{end} \leftarrow W_{start} + \tau - 1$;
3: Calculate the core of each node by k-core decomposition;
4: for each snapshot $G_t$ do
5: for each $v_i \in C_t$ do
6: Remove all neighbors that the core are smaller than $v_i$ to generate a subgraph $\overline{G}$;
7: Calculate the Graph Engagement of $v_i$ according to Eq. (1);
8: Calculate the departure affect of $v_i$ according to Eq. (2);
9: $SGB(v_i) \leftarrow$ Calculate affect score of $v_i$ according to Eq.(3)
10: if Departure has less than n element then
11: Add $v_i$ to Departure;
12: else
13: max_SGB $\leftarrow$ the largest benefit score in Departure;
14: if $SGB(v_i) <$ max_SGB then
15: pop the element with the largest benefit score;
16: Add $v_i$ to Departure;
17: end if
18: end if
19: end for
20: if $t = W_{end}$ then
21: output Departure;
22: $W_{start} \leftarrow W_{start} + \mu$, $W_{end} \leftarrow W_{start} + \tau - 1$;
23: else
24: $t++$;
25: end if
26: end for

Algorithm 1 presents the process of predicting departure user under sliding window. We set the initial values of the sliding window and the current network snapshot (Line 2). Then, we calculate the benefit score of the nodes and rank them according to the benefit score (Line 4-line 19). If the current snapshot is the last snapshot of the window, we output the prediction result of departure node, and the window slides (Line 20-line 26).

5. Experiment
The datasets used in the experiment are the Patent citation network. A data set of 4 million nodes contains publications for multiple discipline patents. Since there are few cross-disciplines in various disciplines, we take computer and chemistry discipline separately as the experimental datasets to verify the precision of the DBP algorithm. We compare DBP with two baseline algorithms based on node degrees and kernels. The window size is 4 and the sliding size is 2. All predictions are compared to the graphs after two years. We will use the precision and recall to verify the algorithm.
Figure 1 to Figure 4 show that as the number of predicted nodes decreases, the precision of the prediction increases. On both discipline datasets, our proposed algorithm is more accurate than the two baseline algorithms based on node degree and core number. In the case of high precision, the recall of the DBP algorithm prediction results is still very close to the two baseline algorithms. This shows that it is inaccurate to simply use the node degree and the number of cores to perform node departure prediction. Our proposed algorithm effectively captures the characteristics of nodes and can achieve higher prediction precision.

6. Conclusion
A departure behavior prediction algorithm is proposes in this paper to predict the user's departure behavior. We perform k-core decomposition on the graph. By considering the positive and negative affect users are experiencing, we calculate a score for each user. Finally, the user is predicted by the sliding window model. Experiments show that the DBP algorithm has a high prediction.

Acknowledgements
This work was supported by the National Natural Science Foundation of China under Grant No.61370084 and No.61872105, and the Fundamental Research Funds for the Central Universities under Grant No.3072019CF0602, and the Project for Innovative Talents of Science and Technology of Harbin under Grant No.2016RAXXJ013, and the China Numerical Tank Project.

References
[1] Anagnostopoulos A., Kumar, R. and Mahdian M. 2008 Influence and correlation in social networks In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 7–15.ACM
[2] Backstrom L., Huttenlocher D. Kleinberg, J. and Lan X. 2006 Group formation in large social networks: membership, growth, and evolution In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 44–54.ACM
[3] Crandall D., Cosley D., Huttenlocher D., Kleinberg, J. and Suri S. 2008 Feedback effects between similarity and social influence in online communities. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. pp.
160–168. ACM

[4] Romero D.M., Meeder B. and Kleinberg J. 2011 Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter In: Proceedings of the 20th international conference on World wide web. pp. 695–704.

[5] Leskovec J., Adamic L.A. and Huberman B.A. 2007 he dynamics of viral marketing ACM Transactions on the Web (TWEB) 1(1), 5

[6] Gladwell, M. 2006 The tipping point: How little things can make a big difference. Little Brown

[7] Rogers, E.M. 2010 Diffusion of innovations. Simon and Schuster

[8] Dasgupta, K. Singh, R. Viswanathan, B. Chakraborty, D. Mukherjea, S. Nanavati A.A. and Joshi A. 2008 Social ties and their relevance to churn in mobile telecom networks. In: Proceedings of the 11th international conference on Extending database technology: Advances in database technology. pp. 668–677. ACM

[9] Wu S., Das Sarma A., Fabrikant A., Lattanzi S. and Tomkins A. 2013 Arrival and departure dynamics in social networks In: Proceedings of the sixth ACM international conference on Web search and data mining. pp. 233–242. ACM

[10] Seidman S.B. 1983 Network structure and minimum degree Social networks 5(3), 269–287

[11] Malliaros F.D. and Vazirgiannis M. 2013 To stay or not to stay: modeling engagement dynamics in social graphs. In: Proceedings of the 22nd ACM international conference on Information & Knowledge Management. pp. 469–478. ACM

[12] Garcia D., Mavrodiev P. and Schweitzer F. 2013 Social resilience in online communities: The autopsy of friendster In: Proceedings of the first ACM conference on Online social networks. pp. 39–50. ACM

[13] Ugander J., Backstrom L., Marlow C. and Kleinberg J. 2012 Structural diversity in social contagion Proceedings of the National Academy of Sciences 109(16), 5962–5966