Discovering Leaders from Social Network by Action Cascade

Ming-Feng Tsai Chih-Wei Tzeng Arbee L.P. Chen

Department of Computer Science
National Chengchi University
Taipei 11605, Taiwan (ROC)
{mftsai, 98001, alpchen}@cs.nccu.edu.tw

Abstract

This paper proposes an approach to discovering community leaders in social network by means of a probabilistic time-based graph propagation model. To conduct the approach, we define an exponential decay function for influence as a function of time, and build action-specific influence chains by multiplying path propagation values; then, we create general influence chains by normalizing over all possible actions. In the study, our approach identifies community leaders as those people whose initiated influence chains are relatively more than the chains they involved. In our experiments, a small Facebook network dataset with 134 nodes and 517 edges is employed to assess the performance of the proposed method. In addition, several baselines are also carried out for comparison, including three naive and one user-involved approaches. The experimental results show that, compared with the baselines, the proposed method can effectively identify community leaders within the social network, achieving 0.8 in terms of F-measure.

Categories and Subject Descriptors H.3.5 [*On-line Information Services*]: Web-based services; J.4 [*Social and Behavioral Sciences*]: Sociology

Keywords Social Network Analysis, Path Mining, Opinion Leader Discovery

1. Introduction

The rising of social websites, such as Facebook and Twitter, has allowed people to connect and re-connect with friends, colleagues, and family from across the world. Nowadays, such websites are among the most popular sites on the Web and still growing rapidly. Social networking data, such as user interactions, is becoming tremendous and easily available. In this paper we attempt to utilize users' influence chains to identify community leaders from social network.

The proposed approach contains two components: The first one is based on a probabilistic graph propagation model, in which we use an exponential decay function to model the influence between users. The second one is a leader discovering algorithm, in which a hierarchy concept is employed to identify community leaders from the influence chains obtained in the first component. Because there may be some irrelevant influence paths, we present a variant of Apriori Probabilistic Path Mining (APPM) method to prune the irrelevant paths. After finding users' influence chains, we then use the hierarchy concept to locate community leaders.

In our experiments, a small-scale social network dataset including 134 nodes and 517 edges is used to verify the performance of the proposed approach. Compared with three heuristic and one user-involved baselines, the proposed approach not only improves the accuracy of leader identification, but also enables us to get more insight into a social network via the mined influence paths.

2. Methodology

We describe the key concept of the proposed approach. Given a social graph in Figure 1(a) and action logs in Table 1, we first use exponential function as shown in Equation (1) to transform the data into probabilistic propagation graphs in Figure 1(b) and 1(c).

$$P(\Delta) \propto \exp^{-\frac{\Delta}{\alpha}}; \qquad (1)$$

To quantify a propagation path $(e_1, e_2, \ldots, e_n)_a$, we use probability to calculate the path, and denote it as $P_a(e_0, e_1, \ldots, e_n)$:

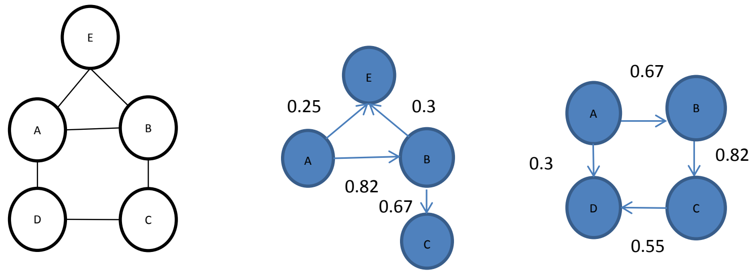
$$P_a(e_0, e_1, \ldots, e_n) = \prod_{0}^{n-1} P_a(e_i). \qquad (2)$$

By averaging all actions in a social graph in Equation (3), where A is the set of all actions, we can identify important propagation paths in a social network.

$$P(e_0, e_1, \ldots, e_n) = \frac{\Sigma_A P_a(e_1, e_2, \ldots, e_n)}{|A|} \qquad (3)$$

Since there may some irrelevant paths, a variant of Apriori Probabilistic Path Mining (APPM) [1] is applied to prune the paths that have no sufficient support; after the pruning, pairs

WP6 SNS'12, April 10, 2012, Bern, Switzerland.
ACM 978-1-4503-1164-9/12/04.



(a) Social Graph (b) G_{a1} (c) G_{a2}

Figure 1. Social Graph and Probabilistic Propagation Graphs

User	Action	Time
Alice (A)	Black Swan trailer (a1)	1
Bob (B)	Black Swan trailer (a1)	2
Alice (A)	Friend Martrix photo (a2)	2
Cathy (C)	Black Swan trailer (a1)	4
Bob (B)	Friend Martrix photo (a2)	4
Cathy (C)	Friend Martrix photo (a2)	5
David (D)	Friend Martrix photo (a2)	8
Eric (E)	Black Swan trailer (a1)	8

Table 1. Examples of Action Log

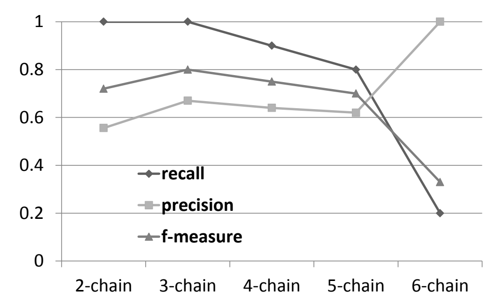


Figure 2. Performance with Different Number of Chains

of top frequent $(k-1)$ paths are merged to form a k-path, and then the merged paths are transformed as influence chains.

After finding influence chains from a social graph, we then use the chains to discover community leaders. In the proposed approach, we record the number of chains that each user leads and gets involved. Our algorithm identifies leaders according to the number of influence chains they initiated and the number of influence chains they involved.

3. Experiments

In our experiments, a small-scale dataset collected from Facebook is used to examine the performance of our proposed method. The social graph contains 134 nodes and 517 edges, and actions are defined as the postings containing links like music video, movie trailer, news, and articles. The possible response for an action/posting can be sharing the posting, leaving comments for the action, or pressing "like" button of the posting. For the ground truth, we ask three annotators to label the dataset. The three annotators are all in the social network of the dataset, and asked to give a list of top 15 community leaders in their mind. We use the intersection of the three lists as our ground truth, in which there are 10 leaders remained. For comparison, according to [2], three heuristic methods are carried out in our experiments as baselines, including using the number of shares, comments, and likes. In addition, we also conduct a more competitive baseline than the three heuristic approaches, in which the selection of action is user-involved. Three metrics are used in our experiments for evaluation, including precision, recall and F-measure. Below we report the experimental results in terms of the three criteria, and provide some discussions on the results.

Figure 2 shows the experimental results of using different number of influence chains for community leader detection. As shown in the figure, our method improves the detection precision as the number of influence chains increases; however, in the other hand, recall and overall performance drop. When the number of influence chains is set to be 3, we obtain the best performance of 0.8 in terms of F-measure.

4. Conclusions

The contribution of this work includes the proposition of the probabilistic time-based graph propagation model for community leader discovery. In addition, we also present a variant of the APPM algorithm to mine influence paths, and then utilize the mined paths to discover community leaders. For future work, we would like to study the effect of the proposed method on ranking community leaders; in addition, we also attempt to use machine-learning techniques to obtain the optimal parameters within our approach.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *VLDB '94*, pages 487–499.

[2] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Discovering leaders from community actions. In *CIKM '08*, pages 499–508.