

Research on Social Network Mining and Its Future Development

Albert Ching-man Au Yeung and Tomoharu Iwata[†]

Abstract

In this article, we explore some core research problems in social network mining and discuss our latest research results. With the world becoming increasingly connected in this age of globalization, communications is no longer bounded by geographical location or languages; in particular, online social networking has become an important area of research.

1. Introduction

Social networks have been studied for many years by social scientists [1], who are particularly interested in understanding the roles of the people in a social network, how they are connected, and how information spreads among them. Social networks greatly influence how people interact and communicate with one another.

In recent years, online social networking has experienced explosive growth, transforming the World Wide Web into a platform for social interactions. Users share their opinions, photographs, music, and videos on social networking services (SNSs). Micro-blogging services like Twitter have also become an important tool for disseminating realtime information [2]. This phenomenon has motivated the development of social network analysis using computers and algorithms.

2. Social network mining

In social network mining, we apply data mining algorithms to study large-scale social networks. Social network mining has attracted a lot of attention for many reasons. For example, studying large social networks allows us to understand social behaviors in different contexts. In addition, by analyzing the roles of the people involved in the network, we can under-

stand how information and opinions spread within the network, and who are the most influential people (**Fig. 1**). In addition, since social network users may receive too much information from time to time, social network mining can be used to support them by providing recommendations and filtering information on their behalf.

In social network mining, we generally ask three broad questions:

- (1) What are the characteristics of the social network?
- (2) How can we model the network?
- (3) How can we support its users?

When trying to answer the first question, we aim to identify different properties of a given social network. For example, what do people do in this social network? Do they exchange messages, or do they share items among themselves? We can also ask, for any two persons, what is the probability distribution of the distance between them? Are there any clusters or communities within the network? Answering these questions enables us to understand how information flows and how social relations in the network evolve. For example, some research on trying to understand the social networks of Twitter [2] and Flickr [3] has been done.

After understanding the characteristics of a particular social network, we may want to construct a mathematical model that explains the processes in the network. A mathematical model lets us predict future changes in the network. For example, what is the probability of a new edge between two given persons?

[†] NTT Communication Science Laboratories
Soraku-gun, 619-0237 Japan

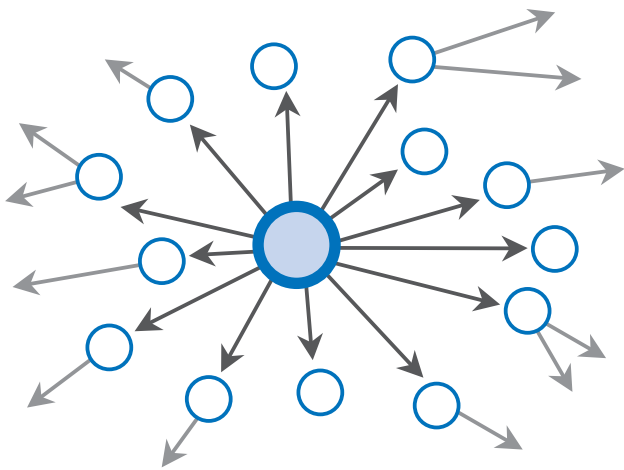


Fig. 1. Identifying the most influential people in a social network.

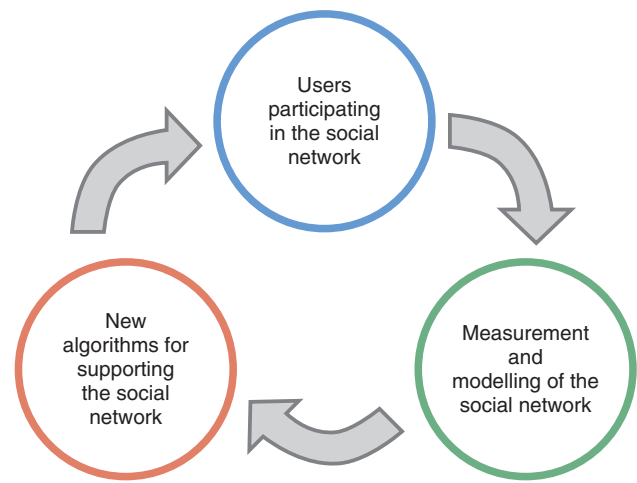


Fig. 3. Life cycle of social network mining.

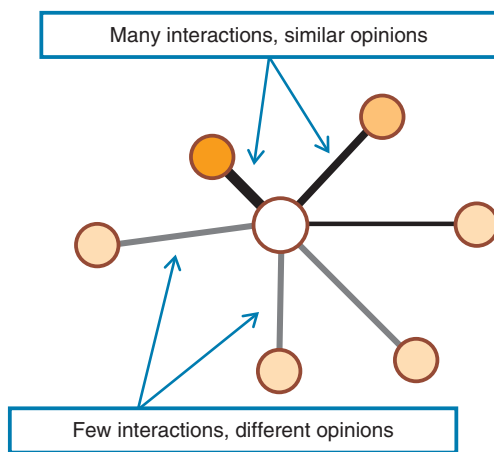


Fig. 2. Using social network mining to estimate the strengths of social relations

When a new person joins the network, who will he or she connect to? We may also want to model the behavior of the people in the network in order to explain when and why two persons interact.

Once we have some knowledge about a social network and its underlying mechanism, it would be good if we could make use of it to support communication among the people in the network. For example, on the basis of their past activities, can we predict who is most likely to become a friend of a given person? Can we estimate and infer the strengths of social relations among different people (Fig. 2)? And can we make better recommendations to the users on the basis of

their social circles?

The above three steps form a cycle (Fig. 3) that one can travel along in order to continuously gain more and more insight into how people interact and then improve the experience of the social network’s users, thus attracting more people to participate in it.

3. Our research on trust networks

We recently proposed a new method for analyzing trust networks on the web and generating more accurate predictions [4]. Trust networks are social networks in which a person is connected to others because the person believes that they are trustworthy.

To study trust networks, we collected data from a popular product review site called Epinions [5]. It lets users write comments and rate any product that they have brought, such as digital cameras, vacuum cleaners, and books. In addition, it lets users create a trust network. For example, if a user thinks that another user’s comments and ratings are reliable, he/she can add that user to his/her own trust network (Fig. 4).

Using data collected from Epinions, our first step was to investigate how trust relations shape user opinions. To do this, we calculated the similarity between pairs of users who had established trust relations between them at some point. Similarity was calculated from which products they had rated and the ratings they had given to them. The results are shown in Fig. 5.

The graph shows that the similarity between users increases over time both before and after a trust

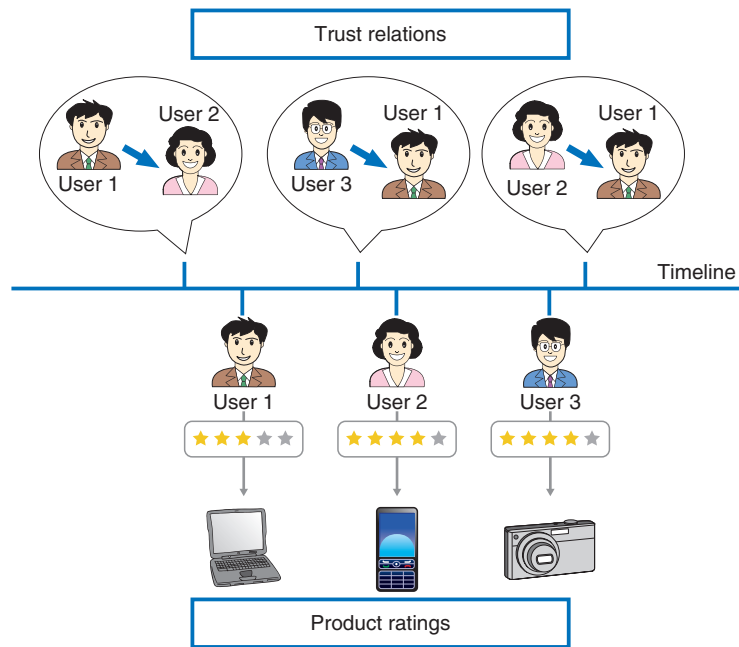


Fig. 4. Epinions user activities.

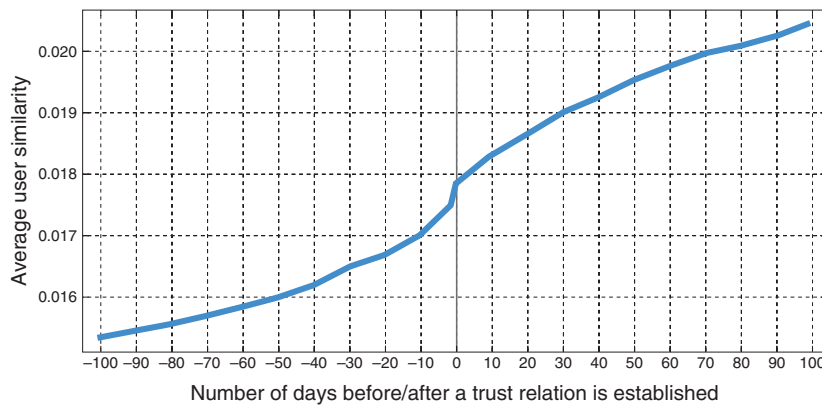


Fig. 5. Change in similarity over time between pairs of users connected by trust relations.

relation is established. This can be explained by two factors. First, the increase before trust is established can be explained by the theory of homophily, which states that people who are similar to each other tend to congregate (birds of a feather flock together). Therefore, when similarity increases to a certain level, it triggers one user to trust another. Second, the increase after trust is established can be explained by influence. That is, when a user trusts another, the former will be influenced by the latter and his or her opinions will become more similar to the latter's.

On the basis of this finding, we further proposed an algorithm based on matrix factorization to predict the rating that a given user will give to a given product. More accurate predictions will let us generate better recommendations for users. In our method, we assume that a user's rating is determined by two factors: whether the user and the product are compatible with each other, which is determined using the basic matrix factorization technique, and the influence from other users in the trust network. If other users tended to give this product high ratings, then this user

is likely to be influenced by them to give a higher rating too. Our algorithm first estimates the strengths of influence between different users on the basis of past activities and then uses these strengths to make predictions.

Our experiments showed that this new algorithm can produce predictions with lower errors than standard matrix factorization techniques. This project has demonstrated that our understanding of a social network can be used to develop new algorithms for supporting the network's users.

4. Summary

Since more and more people are likely to interact and communicate with one another on the web, social network mining will become even more important. It not only holds the key to understanding social behav-

ior and group dynamics on a huge scale, but also is crucial in developing new tools and functions to support communication in social networks.

References

- [1] S. Wasserman and K. Faust, "Social Network Analysis: Methods and Applications," Cambridge University Press, 1994.
- [2] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" Proc. of the 19th International World Wide Web Conference (WWW2010), pp. 591–600, Raleigh, NC, USA, 2010.
- [3] M. Cha, A. Mislove, and K. Gummadi, "A Measurement-driven Analysis of Information Propagation in the Flickr Social Network," Proc. of the 18th International World Wide Web Conference (WWW2009), pp. 721–730, Madrid, Spain, 2009.
- [4] C. M. Au Yeung and T. Iwata, "Strength of Social Influence in Trust Networks in Product Review Sites," Proc. of the 4th ACM International Conference on Web Search and Data Mining (WSDM2011), pp. 495–504, Hong Kong, China, 2011.
- [5] Epinions. <http://www.epinions.com/>



Albert Ching-man Au Yeung

Research Associate, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received the M.Phil. degree in computer science and the B.E. degree in information engineering from the Chinese University of Hong Kong and the Ph.D. degree in computer science from the University of Southampton, UK, in 2001, 2004, and 2009, respectively. He joined NTT Communication Science Laboratories in 2009. His research interests include social media, social network analysis, user modeling, and recommendation systems.



Tomoharu Iwata

Research Scientist, Learning and Intelligent Systems Research Group, NTT Communication Science Laboratories.

He received the B.S. degree in environmental information from Keio University, Tokyo, the M.S. degree in arts and sciences from the University of Tokyo, and the Ph.D. degree in informatics from Kyoto University in 2001, 2003, and 2008, respectively. His research interests include data mining, machine learning, information visualization, and recommender systems. He received the IPSJ (Information Processing Society of Japan) Best Paper Award, FIT (Forum on Information Technology) Young Researcher's Award, and Funai Best Paper Award. He is a member of the Institute of Electronics, Information and Communication Engineers and IPSJ.