Social Network Analysis Using Big Data

O. Liu, K.L. Man, W. Chong, and C.O. Chan

Abstract— **The increasing use of social networks, such as Facebook, Twitter, and Weibo, has produced and is producing huge volume of data. Business firms and other organizations are interested in discovering new business insight to increase business performance. By using advanced analytics, enterprises can analyze big data to learn about relationships underlying social networks that characterize the social behavior of individuals and groups. Using data describing the relationships, we are able to identify social leaders who influence the behavior of others in the network, and on the other hand, to determine which people are most affected by other network participants. This study focuses on modeling the knowledge diffusion in social networks. We will present a new evolving model of a directed, scale-free network. We will test the effectiveness of our model by a simulation using data of a real-world social network.**

Index Terms—**Social Network, Big Data Technologies, Directed Network**

I. INTRODUCTION

DATA challenges today are often categorized as "Big" because they deal with one or more of the following: "big" volume, velocity or variety [1], [2], [3]. While the challenges of analyzing such "big data" are most often discussed, growing volume, velocity and variety of data are produced in social media. The increasing use of social networks, such as Facebook, Twitter, and Weibo (http://www.weibo.com/) , has produced and is producing huge volume of data. Twitter posts more than 500 million tweets every day. Weibo is reported to have over 766 million active users per day in 2014. Business firms and other organizations are interested in discovering new business insight to increase business performance.

The Big Data produced by social networks can be analyzed by current computer technologies [4], [5], [6] [7]. MapReduce, Hadoop and NoSQL techniques have supported distributed data storage, parallel data retrieval and processing. Many analytical methods and algorithms are designed for business analytics, such as K-means clustering, Association rules, Linear/logistic regression, and Time series. Many software companies have developed their BDA products. For example, IBM has a series of software to support BDA including InfoSphere, PureData, Cognos and SPSS modeler.

By using advanced analytics, enterprises can analyze big data to learn about relationships underlying social networks that characterize the social behavior of individuals and groups. Using data describing the relationships, we are able to identify social leaders who influence the behavior of others in the network, and on the other hand, to determine which people are most affected by other network participants. We can also use diffusion analysis to identify the individuals most affected by the group leaders and target the marketing to them.

Many models for social networks were established based on undirected networks. In this study, we will construct an evolving model of a new directed, scale-free network on the basis of the BA model [8]. Our proposed model adopts the mechanism of preferential attachment during network evolution, which is considered one of the key factors in the formation of scale-free networks. We will test the effectiveness of our model by a simulation using data of a real-world Chinese social network.

II. THEORETICAL FOUNDATION

A social network is a set composed of nodes and links between each two nodes [8]. The nodes also represent social actors, and the links are relationships. People or organizations as well as their social relations are connected by a social network. A social network structure is the actual or potential relational schema existing between actors. A social network structure can not only help us understand the kinds of gathering ways among a set of actors in a special space, but also help us understand the significance of one's behavior.

In a virtual community, a social structure is some mode of stable relationships and is always expressed as a network formed by a series of nodes (actors) and links that represent relationships among nodes. There are three basic substances of social network structures: the actor, relation, and network, which are introduced below.

A. Actor

Nodes in network are actors. Everything such as a user, a book, or a movie can be an actor in the virtual community. When doing research on a virtual study community based on a network environment, we can focus on some particular actors who have a very important influence on the majority of others on one side, and pay attention to the entire network on the other side.

B. Relation

A relation is a link of nodes. It is one of the core parts in basic research on virtual learning communities. Nodes (actors) share, transfer, and acquire information through direct or potential links (relations). There are three features of relation: content, direction, and strength.

Direction could be divided into directed and undirected. In some websites, A can follow B without B's acceptation. Then, we can say their relation has a clear direction. However, in other websites, A and B can be friends only if they follow each other. Therefore, their relation has no direction.

There are weak as well as strong ties between individuals. In light of the trust in a society, the clan trusts exist in strong connections, and the foundation of weak connections is impersonal trust in society. Strong connections can become weak or dissolve due to various reasons. In contrast, weak connections may also become strong with an increase in confidence. In a social network, we distinguish relations' strength by the frequency of their interaction.

C. Network

A network appears to be a set of relations, describing its relational schema or connection mode. There are two typically types of network: ego-centered network and whole network.

When one analyzes an ego-centered network, one generally pays attention to special actors, named "Internet stars," and the networks they (actors) establish and relations (links) with their neighbors.

The whole network method aims to study relations of all the actors within a certain range. Therefore, a large amount of relational data is of huge importance. This leads to the problem of big calculation. To solve this problem, it is necessary to get help from computer assistance computation.

Sociogram as well as Social Matrix are powerful tools as we seek to describe networks. The former makes the network more intuitive than the latter. However, when there is a large amount of actors, it is difficult to analyze the relational structure within Sociogram. Thus, Social Matrix seems like the best option.

III. Summary

The information era has witnessed an unprecedentedly high speed of data creation and knowledge diffusion. Many models for social networks were established based on undirected networks. In this study, we will construct an evolving model of a new directed, scale-free network. Our proposed model adopts the mechanism of preferential attachment during network evolution, which is considered one of the key factors in the formation of scale-free networks. We will test the effectiveness of our model by a simulation using data of a real-world Chinese social network.

References

[1] P. Setia, V. Venkatesh, and S. Joglekar, "Leveraging Digital Technologies: How Information Quality Leads To Localized Capabilities And Customer Service Performance," *MIS Quarterly*, vol. 37, 2013, pp. 565-A4.

[2] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, and P. Tufano. *Analytics: the real-world use of big data: how innovative enterprises extract value from uncertain data*, Executive Report, IBM Institute for Business Value and Said Business School at the University of Oxford, 2012.

[3] P. B. Goes, "Big Data and IS Research," *MIS Quarterly*, vol. 38, 2014, pp. iii-viii.

[4] S. Fan, R. Y. K.Lau, and J. L. Zhao, "Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix." *Big Data Research*, vol. 2, 2015, pp. 28-32.

[5] S. Mithas, M. R., Lee, S., Earley, Murugesan, S. and R. Djavanshir, "Leveraging Big Data and Business Analytics [Guest editors' introduction]." *IT Professional*, vol. 15, 2013, pp. 18-20.

[6] Y. Zhao, D. Li, and L. Pan, "Cooperation or Competition: An Evolutionary Game Study between Commercial Banks and Big Data-Based E-Commerce Financial Institutions in China." *Discrete Dynamics in Nature and Society*, vol. 8, 2015.

[7] C. K. Velu, S. E. Madnick, and M. W. Van Alstyne, "Centralizing Data Management with Considerations of Uncertainty and Information-Based Flexibility." *Journal of Management Information Systems*, vol. 30, 2013. pp.179-212.

[8] R. Albert, A. L. Barabási, Statistical mechanics of complex networks. *Reviews of Modern Physics*, vol. 74, 2002, pp. 47-97.