Social Network Analysis and Its Developments

DENG Xiaoxiao[1] MAO Guojun[2]

[1] Macau University of Science and Technology
[2] Central University of Finance and Economics

Abstract

A social network is made up of a set of social actors and their ties between these actors, which have become an important research filed of sociology and its related studies. Together with the complex network concepts, it has formed many new methods and models. Based on information technology such as data mining, it is now having the trend of development to automatically deal with a large scale of social data. This paper focuses on the social network analysis by using both the complex network theory and the data mining techniques. While introducing their basic methods, it presents their research progresses.

Keywords: Social network, complex network, data mining.

1. Overview of Social Networks

The social network is a developing theory in the social sciences to study relationships between individuals or organizations [1]. A base of the social network approach to understand social interactions is that social phenomena should be investigated through the properties of relations between units, instead of the properties of these units themselves.

Basic concepts of the social network were formed in the 1930s, but it just blossomed out gently before 1970s. The appearance of "New Harvard", a group consisted of sociologist Harrison White and his students at the Harvard University in 1970s, made the social network researches into an important period. The main feature during this period is the organic link between microcosmic social actions and macrocosmic structures of society by a social network. The active researchers include at the time included Harrison White, Samuel Leinhardt, Ronald Burt and Peter Marsden.

Harrison White put forward the model called the Opportunity Chain (1970); Samuel Leinhardt gave a series of systematic standards to express a social network (1980); Ronald Burt presented a structure model for social network (1982); and Peter Marsden described a general analytic methods for social network(1992).

From 1990s, the research on social networks has reached another new alp, which has attracted a great deal of experts and scholars working in different research fields such as sociology, anthropology, psychology and mathematics. Thus, interdisciplinary cooperation has become a new development trend in the social network research field.

For example, Ronald Burt gave the Structural Hole Theory (1992), i.e. the role of a structural hole would take on the competition advantage comparing with the other behaviors in the social network [2]; Duncan Wattts created the Small-Word model (1998), which shown two arbitrary people were often connected by only limited number of intermediators (known as six degrees of separation) [3];

Reca Albert found the Scale-Free feature (1999), which expressed that many social networks had the power-law degree rather than formal scale distribution [4]. In addition, Lin Nan introduced social capital concept into social network (2001) which insisted the individuals' actions should be reflections of social capitals [5].

Stepping into the 21 century, social network has been put into further research along with the development of other techniques, where complex network theory and data mining technology are two important supports to the modern social network analysis.

Owing to the real world is a complex social system, so a social network that describes social actions and their relation should be seen a special complex network. Meanwhile, the population of computers and the Internet can help compute a large scale of social data and so it is possible to analyze the social networks in the automatic way.

In this paper, we consider the problems how to use complex network and data mining techniques in order to analyze social networks, and discuss their developments in such studies.

2. Social Network Analysis Based on Complex Networks

A complex network is a graph which can be used to describe the objects and their relations in the real word. Many social, biological, and technological networks often display the features of complex networks. Such features include non-random degree distributions, high clustering coefficient and hierarchical organizations. For example, starts in galaxy, things in food chains and routers of Internet can all be related with some obvious attributes of complex networks.

The research of complex networks is a young and active area of scientific studies. In the context of network theory, a complex network often has some non-trivial topological features. That is, unlike a regular lattice or a random graph, the patterns of connection between their elements in complex networks can be neither purely regular nor purely random. Therefore, it is difficult to describe and analyze a complex network by a simple method, and so it has brought together researchers from many areas including mathematics, physics, computer science, and others.

In general, mathematics is always as an important way to describe the topological structure of a complex network; Physics can help discover useful dynamic features with changing over time in a complex network; Computer science will provide technical support in automatically and intelligently computing and analyzing data in a complex network.

Two well-known complex networks are the Small-World networks and the Scale-Free networks. Both are characterized by specific structural features: short path lengths for the former and power-law degree distributions for the latter.

The first network was put forward and widely put into practice is the small-word network model, which laid the research foundation for analyzing society networks by breaking the limit in stochastic network.

Two main properties make the small-word networks different from others:

- First, a small-word network always has a shorter average path length so that most nodes in the network could be reached by a small number of steps.
- Second, a small-word network has a higher average clustering coefficient so that many dense overlapping groups arise in the network.

In 1998, Duncan Watts and Steven Strogatz published the first small-world network model. As Fig. 1 shows, a small-

word network can be generated by smoothly interpolating a few of connections in a random graph.

Comparing with Fig.1(a), Fig.1(b) only does a small quantity of changes, but its small-world attributes are enhanced in a large level. That is, the number of edges that connect most vertices has made less in Fig.1(b) than in Fig.1(a), and some nodes form obvious clustering property like the vertices linked by the red lines.

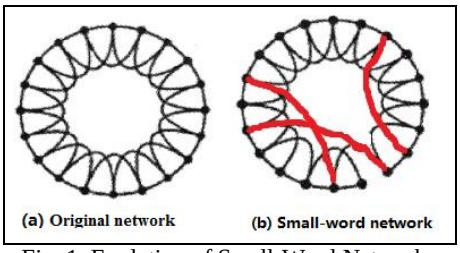

Fig. 1: Evolution of Small-Word Networks

The small-world phenomenon has been manifested to exist in many social networks, including friend or acquaintance relations, and business or talking partners. As far as the social network is concerned, the persons in a group can be described as a social network, but some ones can construct a closed aggregation under a certain interpersonal relationship. Typical examples of them are the friend circles in social networking tools like Facebook, Micro-blogs and QQ.

Scale-free networks are another type of complex networks that widely used, derived from observation and analysis on web pages links on the Internet. It was found that "there are a few of web pages in an Internet website that are connected to many pages, but most pages in this website are just linked into others in a small number" [4].

A scale-free network always follows the particular mathematical distribution function called a power law. The power law implies that the degree distribution of a network has no characteristic scale. As Fig.2 states, unlike a random network, there are often some vertices that have the magnitude larger degrees than the average ones in a scale-free network.

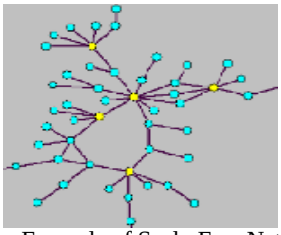

Fig. 2: An Example of Scale-Free Networks

The scale-free network model can easily tell the reason why so many people have routines to follow in social network, and implicates that like natural disease, information spread would be done through some important pickup points. As far as some new communication tools like micro-blogs is concerned, scale-free networks can be used as a base of expression and analysis to find out interpersonal authoritativeness and important talking roles on them.

Many studies have proven the availability of small-word and the scale-free networks to many social problems, but they cannot solve all the complex social issues yet. In fact, the study of complex networks has been progressing, and has attracted attention to more and more researchers.

Its developments lead to two main directions. One is looking for new complex network models or features through social phenomena. The other is searching for attributes and support theories for complex networks faced with more complicated social relations like competitions, rights and reputations.

In a social network, there are a number of strong lies like relations between relatives or friends, but weak lies like nodding acquaintances also exist. As the strong lies tend to be easily taken advantage of, but weak ones can be ignored that they often can help people under-

stand so many social phenomena. For example, there is a famous social observation that said: “The most way for finding a job through personal recommendations is from once-met or few-met acquaintances rather than the best friends”. Therefore, we can take it for granted that plenty of social phenomena may be explained by weak lies-based networks, where information share like cultural disseminations or opinion spreads in social communication groups is just some examples.

Once more, stimulated by ubiquitous competitions in social communications, structural holes theory come out. Structural holes express the values of rare links in a group, especially in high density ones. From view of competition, the more scarce a social resource, the higher its competitive power. Thus, discovering structural hole-based models is ongoing to solve complex social problems.

Because many different types of relations in the social system could form complex social network configurations, so the complex network-based social network analysis is useful to a broad range of social fields. They include, but are not limited to social communication laws, organizational studies, and others.

3. Data Mining-based Social Network Analysis

Data mining is the technology of analyzing data from a large scale of data. Recently there has been a rapid increase in interest regarding social network data mining. The main motivation is from the demand to exploit knowledge in large datasets that cannot be handled by traditional methods, collected from online communion environments on Internet. Vast amounts of user-generated data are created on social media sites every day, which present an opportunity for data mining applications to develop new algorithms to deal with data from social Internet-based media.

In 2010, Jennifer Jie Xu published the book “Data Mining for Social Network Data”, which roundly discussed the rresearch questions, main techniques and effectiveness about applying data mining into social networks [8]. As far as mining effectiveness is concerned, for example, data mining techniques can help identify the influential or important people in a social group, detect implicit and valuable clusters in a social network, and recommend good friends or valuable products for talking or business.

In short, social network analysis is a multidisciplinary field dedicated to the analysis and modeling of relations among various objects in the society, and so data mining is as an important technique that can help understand how the behavior of individuals interact in a social network and what patterns their interactions would have.

Specially, the pervasive use of social media has generated vast amounts of social data, so mining social media has been becoming new research focus. However, data generated on social media sites are different from conventional attribute-value data, and they are often vast, noisy, unstructured and dynamic. These characteristics give a new challenge to data mining techniques.

Due to exploding popularity of online social networks, huge amount of user-generated data is available which makes a new problem called big data mining. Of course, some other challenges in data mining-based social network analysis exist. One of the challenges is the dynamic nature of the real-world social networks that tend to change with time, so new methods need to be developed in efficiently dealing with the dynamic link problem. In addition, online social networks are a rich source of both structural and nonstructural data, and their data

mining-based models are different, thus the new mining frameworks and models are necessary to form an efficient platform in collaboration with each other to yield more meaningful and realistic results.

Social network analysis of using data mining techniques has been paid loads of attention to researchers or business, and many challenging tasks including mining theory, architecture, models and detail techniques, will be done in the future.

4. Conclusion

Social Network has being play a vital role in both sociological and other research fields. A basic social network structuralizes social behaviors and their relations in an organization or community. Also, the Internet is filled with millions of individuals who are looking to meet other people for sharing first-hand information and experiences, and so finding useful individual links to perceive social structure on Internet-based communication media is becoming a challenging task.

Two aspects are calling big attentions. One is employing complex network concepts to build available models for better understanding the complicated social system. And the other is embedding data mining techniques into social networks for intelligently analyzing a large scale of social data. Therefore, more complex social problems will be involved while intelligent methods are employed for the social network analysis in the future.

5. Acknowledgment

We wish thank China National Science Foundation for supporting this research (Project No. 61273293).

6. References

[1] W. Stanley, "Social Network Analysis," *Cambridge University Press*, 1996.

[2] J. Lee and S. Kim. "Exploring the Role of Social Networks in Affective Organizational Commitment : Network Centrality, Strength of Ties, and Structural Holes," *The American Review of Public Administration*, 41(2), 2011.

[3] M. Newman and D. Watts, "Scaling and Percolation in the Small-Word Network Model," *Physical Review E*, 1999.

[4] R. Albert R, "Statistical Mechanics of Complex Networks," *Ph D Dissertation, Cambridge University Press*, 2002.

[5] N. Lin, "Social Capital: A Theory of Social Structure and Action," *Cambridge University Press*, 2001.

[6] M. Kilduff and W. Tsai, "Social Networks and Organizations," *Sage Publications*, 2003.

[7] C. Jones and E. Volpe, "Organizational Identification: Extending Our Understanding of Social Identities through Social Networks," *Journal of Organizational Behavior*, 32, 2011.

[8] J. Xu and D. Hicks, "Data Mining for Social Network Data," *Springer*, 2010.

[9] A. Matthew, "Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites," *O'Reilly Media*, 2011.