2017 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2017, Cancun, Mexico, December 4-7, 2017

Analysis of Social Networks and its Evaluation using Big Data Analyzing Technique

Atsushi Tanaka†, Seungil Cho†, Muneki Yasuda†, Tomochika Harada† and Michio Yokoyama†

† Graduate School of Science and Engineering, Yamagata University, 4–3–16 Jonan, Yonezawa 992-8510, Japan
Email: tanaka@yamagata-u.ac.jp

Abstract—In order to analyze social networks, an improved method for Clique Percolation Method(CPM) is proposed. Using this method, which is called pseudo ACPM, network analysis of friendship networks on SNS sites for college students is carried out. As the number of lack of nodes to fuse two cliques decreases, it is confirmed that small communities inside large communities can be detected. Our results are also evaluated using the method for big data analysis.

1. Introduction

It has been almost 20 years since the field of network science became very popular all over the world, and we study the universal characters regarding all connections among all sorts of things as network. The research field extend to Internet, ecology, physics, chemistry, sociology, epidemiology etc[1]. Many sociologists have been interested in human related networks, and clarified its structure by some experimental studies[2][3][4]. However since the methods and resources for analyzing huge network were insufficient, their analyses were limited. In recent years, the rapid progress of computers and the establishment of scientific method have enabled us detailed analysis of networks. Moreover recent rapid spread of social media has also enabled us to obtain easily and cheaply connection data of human networks, streaming information in there, and whole temporal change of them. In Facebook and Twitter, more than several hundreds million people are exchanging information over the boundary of nations and the power of their influence are spreading to several kinds of directions.

In this paper, we analyze data of SNS sites using our community detection method and evaluate it with the method for big data analysis.

2. Social Networks

Recent progress of social media is amazing. Facebook and Twitter especially have obtained more than several hundred million users all over the world, and also have achieved big success commercially. Many reports said many people obtained some necessary information over Twitter in the East Japan Earthquake in 2011, and it has been recognized that social media is important for the information acquisition method. Mixi, an original Japanese social media, succeeded in obtaining huge users at one time. Yuta et al. revealed its structure using network analysis[5]. Toriumi et al. also analyzed a lot of SNS sites widely and categorized them into several groups by their characteristics[6][7]. As above a lot of researches on structural analysis and modeling of inherent networks inside social media have been carried out.

Friendship network has had an interest widely since early days in the field of social science though, they could not obtain correct and huge scale data easily because of several problem like privacy. However the progress of social media enabled us to obtain human relation network data easily and minutely. You might say data on social media is virtual on the Internet though, these considerable parts are formed autonomously in the real space and their growth is speculated to be fast relatively. From this feature, analyzing these data is meaningful for the analysis of distributed autonomous system.

3. Community Analysis Methods

3.1. Existing Methods

The research of social media is suitable for studying the network dynamics. The networks in social media especially SNS strongly depend on human relationship, thus the community analysis is indispensable. The community analysis has been studied for a long time in sociology, and several kinds of methods have been proposed so far. The popular methods are the one using betweenness centrality by Girvan et al.[8], the one using Q value by Newman[9], and its developed version for larger-scale networks with more speed by Clauset et al.[10]. However, above methods are all for partitioning a large community into some distinct smaller communities. If we consider human relationship in a real world, there often happens that some people gets deeply involved in several communities at the same time. Therefore, if we consider the network in social media, it is rather natural to think that some communities overlap each other, and hence we use overlapping community analysis. Among several kinds of these techniques that have already proposed, the Clique Percolation Method (CPM)[11] is the most popular. So they have analyzed the network mainly with this method so far. For more information of community analysis, see the review article by Fortunato[12].

3.2. Improvement of CPM

In CPM, one of the cumulative algorithm, when two k degree cliques are fused to one community, $k-1$ nodes must be shared among them. It seems to be very rigorous especially for large k, so communities are difficult to become large. Here we permit the lack σ in overlapping to fuse two communities. It means we alleviate the connecting condition of cliques. Since if we set $\sigma = 1$ in this method, we can obtain exactly the same method of original CPM, so we call it Alternative Clique Percolation Method(ACPM)[13].

Since ACPM is based on the connection of k degree cliques, many nodes which do not form any degree cliques are all removed, thus they never become members of any community. This is a very rigorous condition. Thinking of real networks, it seems to be valid we regard a cluster with a little lack of edges as a clique, now we define the allowance parameter α as the rate of edges comparing with complete graph. We call this method pseudo ACPM.

4. Analytical Results

4.1. Target SNS Sites

In human networks, we can often observe community structures, but real relationships seldom become open, so it is often difficult to assemble the data. On the contrary, speaking of networks on social media relationships can be observed, and assembling the data is relatively easy. Thus many analyses of network structures and communities have been proceeded. In this paper, we analyze two SNS systems, tomocom.jp and FLECCS.

In order to investigate the network dynamics in SNS, we have constructed SNS site a.k.a. “tomocom.jp”. This SNS is only for college students and completely invitation-based system. There seminar students are invited by college teachers in several areas in Japan. It possesses higher reliability thanks to its guardian system. This site was opened in 2009, and captured more than 400 users by the end of 2010. Though the interaction within their own seminars is basic, the connections between different seminars have been created by writing and browsing their blogs. Since activation event of the site over several times were held, the connections became denser. We confirmed the network structure between communities has also smallworld property, so confirmed nested structure of smallworld[14].

As another object,we analyze the SNS called FLECCS. It is a product of the cooperation of 8 universities in Fukui prefecture in Japan and it provides not only SNS platform but also the environment of virtual university. Same as tomocom.jp, it is SNS for university students, however it adopts registration system, not invitation. Therefore the whole network is not one connected graph and there are some isolated small subnetworks.

4.2. Analysis using pseudo ACPM

We carried out community analysis on two described SNS sites using our pseudo ACPM. Result on tomocom.jp is shown in Fig.1. Yellow, red, blue and gray circles in those figures mean the largest, the second largest, the third largest community nodes and other ones respectively. Edges are restricted only on those between extracted nodes. We fix pseudo parameter $\alpha = 0.92$, and vary the parameter σ from 3 to 5. Though $\sigma = 1$ corresponds to the original CPM, They are not the same because $\alpha \neq 1$. In the case of $\sigma = 5$, only two communities are detected. On the other hand, in the case of $\sigma = 3$, a giant community is divided into middle-scale communities. Six month later, we can detect more large and dense communities in a similar way. We remark two communities do not always have to share overlapping nodes in our method.

In FLECCS, we can also detect only two communities in the case of $\sigma = 5$ though, we can not obtain any clear and dense communities inside largest one(Fig.2). That is mainly because FLECCS does not have clear communities from the beginning and it is based on registration system, while tomocom.jp adopts invitation system. In Fig.2, you can find overlapping nodes between two communities shown with black circles, while you can not find them in tomocom.jp. Six month later, such inclination does not change so much comparing with that of tomocom.jp.

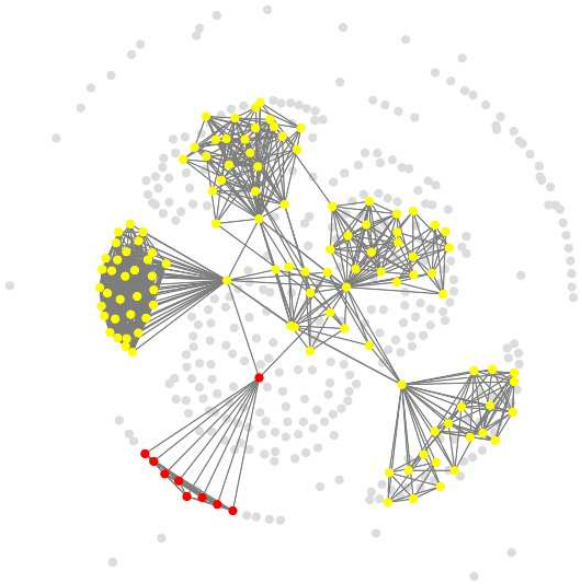
5. Evaluation of our method

It is general that methods of community analyses are evaluated by the ratio of the number of edges inside the community and that of inter communities. It is no doubt to consider the larger the ratio becomes, the better the community detection method is. Using an ordinary evaluation, we obtain the internal and inter ratios as 0.61, 0.01 in tomocom.jp and 0.62, 0.01 in FLECCS respectively.

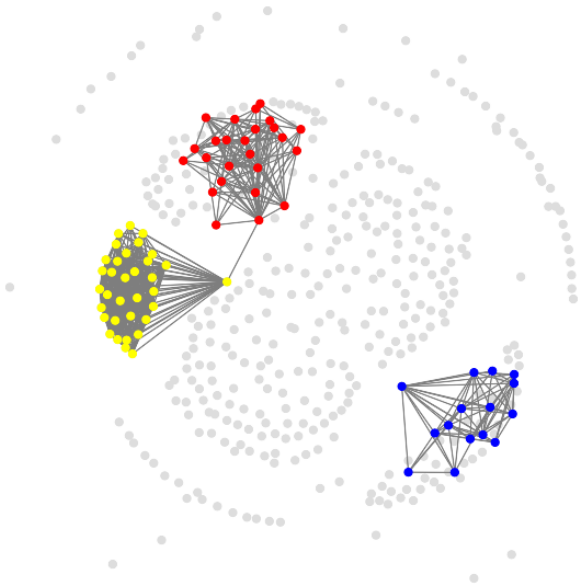
In this paper our objective data is in social media and we use overlapping method, we carry out the evaluation based on the similarity of network attributes of nodes. In this process we make use of the method which is developed for big data analysis[17]. An example of visualization with this method is shown in Fig.3. We can expect the evaluation becomes more reliable using more than one index. Detail results will be shown in the conference.

6. Conclusion

In order to analyze social networks, an improved method for Clique Percolation Method(CPM), which is one of the effective community analysis with overlapping nodes is proposed. Using this method, which is called pseudo ACPM, network analysis of friendship networks on SNS sites for college students is carried out. As the number of lack of nodes to fuse two cliques decreases, it is confirmed that small communities inside large communities can be

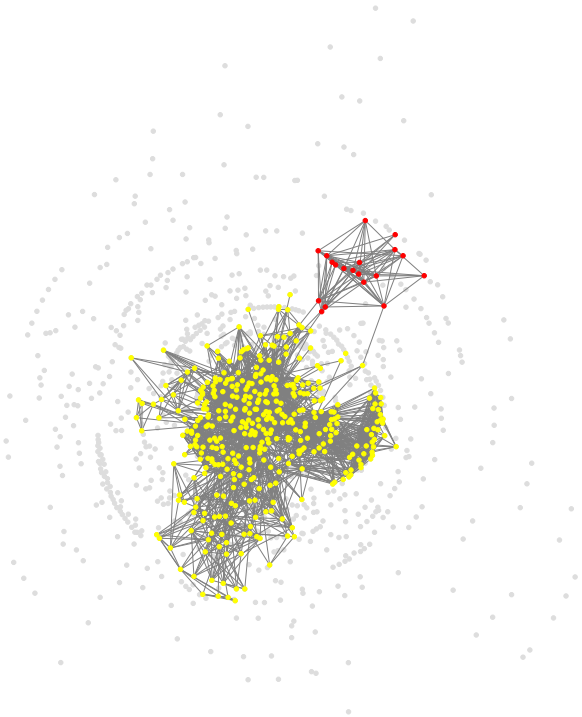


(a) $\sigma = 5$

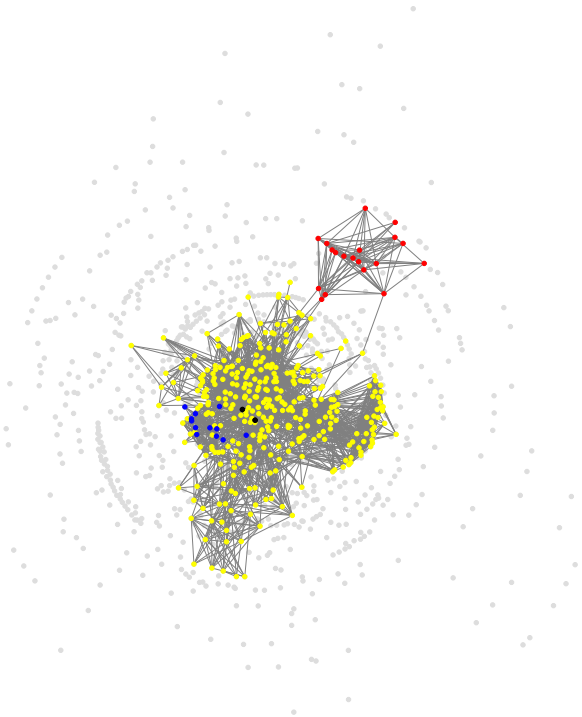


(b) $\sigma = 3$

Figure 1: Communities in tomocom.jp at 31-Dec-2010. Yellow, red, blue and gray circles represent the first, the second, the third largest communities and other nodes respectively. There are no overlapping nodes among them.



(a) $\sigma = 5$



(b) $\sigma = 3$

Figure 2: Communities in FLECCS at Jul-2010. Yellow, red, blue and gray circles represent the first, the second, the third largest communities and other nodes respectively. Overlapping nodes between the first and the third communities are shown with black circles.

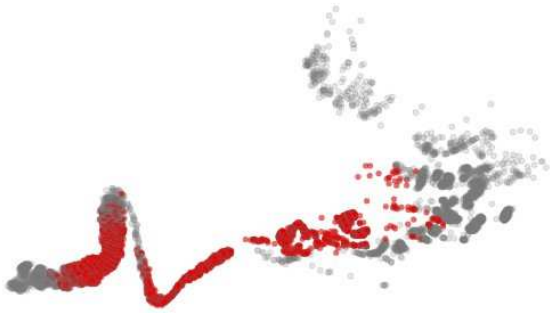


Figure 3: An example of big data analysis result. All similar node come closer and form some clusters.

detected. The differences between two SNS sites coming from their system, registration or invitation, are also clarified.

In order to evaluate our method, we use the method for big data analysis in addition to the ordinary evaluation. With that method we can evaluate effectively our results in a different way.

Acknowledgments

This work was partly supported by JSPS KAKENHI(Granted Numbers JP17H01729, JP17K01781, JP15K00330 and JP15H03699). Besides this work was also partly supported by the Center of Innovation Program from the Japan Science and Technology Agency (JST), and JST CREST(Grant Number JPMJCR1402), and MIC SCOPE(Strategic Information and Communications R&D Promotion Programme); Grant Number: 172302009. Furthermore, the authors would like to thank Dr. Takaeda, toor Inc., for providing us the useful analysis tool “BigDAT@viewer”.

References

[1] R. Albert and A. L. Barabási, “Statistical Mechanics of Complex Networks,” *Rev. Mod. Phys.*, Vol. 74, pp. 47–97, 2002.

[2] S. Milgram, “The Small World Problem,” *Psych. Today,*, Vol.1, pp. 61–67, 1967.

[3] T. Travers and S. Milgram, “An experimental study of the small world problem,” *Sociometry*, Vol.32, pp. 425–443, 1969.

[4] C. Korte and S. Milgram, “Acquaintance links between White and Negro populations, Application of the small world method,” *J. Personality and Social Psychology*, Vol.15, No.2, pp. 101-108, 1970.

[5] K. Yuta, N. Ono and Y. Fujiwara, “Structural Analysis of Human Network in Social Networking Services,” *IPSJ Journal*, Vol.47, No.3, pp. 865-874, 2006.

[6] F. Toriumi, H. Yamamoto, H. Suwa, I. Okada, K. Izumi and Y. Hashimoto, “Comparison Analysis among Large Amount of SNS Sites,” *Trans. Japanese Soc. Artificial Intelligence*, Vol.25, No.1, pp. 78–89, 2010.

[7] F. Toriumi and K. Ishii, “Simulation of Encouragement Methods for SNS based on User Behavior Model,” *Proc. of 3rd World Congress on Social Simulation*, 2010.

[8] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proc. Natl. Acad. Sci. U.S.A.*, Vol.99, No.12, pp. 7821–7826, 2002.

[9] M. E. J. Newman, “Fast algorithm for detecting community structure in networks,” *Phys. Rev. E*, Vol.69, 066133, 2004.

[10] A. Clauset, M. E. J. Newman and C. Moore, “Finding community structure in very large networks,” *Phys. Rev. E*, Vol.70, 066111, 2004.

[11] G. Palla, I. Derényi, I. Farkas and T. Vicsek, “Uncovering the overlapping modular structure of protein interaction networks,” *FEBS JOURNAL*, 272, pp. 434–434, Suppl. 1, 2005.

[12] A. Fortunato, “Community detection in graphs,” *Physics Report*, Vol. 486, Issues 3-5, pp. 75–174, 2010.

[13] A. Tanaka, “Proposal of Alleviative Method of Community Analysis with Overlapping Nodes,” *Proc. The Seventh IEEE Int. Conf. on Social Computing and Networking*, C.1.30-5, 2014.

[14] A. Tanaka and M. Tomochi, “Emergence of Hierarchical Small-world Property in SNS for College Students,” *Proc. Int. Symp. on Nonlinear Theory and its Application*, pp. 98–101, 2012.

[15] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, Vol.393, pp. 440–442, 1998.

[16] A. L. Barabási and R. Albert, “Emergence of Scaling Networks,” *Science*, Vol. 286, pp. 509–512, 1999.

[17] http://www.cybernet.co.jp/toorpia/