Combining study of complex network and text mining analysis to understand growth mechanism of communities on SNS

OSAMU YAMAKAWA
Fukui Prefectural University, JAPAN

TAKAHIRO TAGAWA, HITOSHI INOUE
Kyushu University, JAPAN

KOICHI YASTAKE, TAKAHIRO SUMIYA
Hiroshima University, JAPAN

1. INTRODUCTION

Japanese 7 higher educational institutions in Fukui prefecture started to cooperate with each other in the middle of 2008. The aim of this project called "F-LECCS" is in order to build "One virtual university environment" on a computer network by using open source software such as SNS (OpenSNP), LMS (Moodle), and e-Portfolio (Mahara). This open platform allows all the users to access learning resources across the universities, and enables them to form intercollegiate learning communities as the "Communities of Practice (CoP). The center system of the CoP is the social networking service (F-LECCS SNS) and has started to work on the beginning of April, 2009. On the F-LECCS SNS, over 300 communities and 360,000 logins (370 logins/day) have been made by the end of March, 2011. The CoPs are growing well.

Our research interest is to understand the growth mechanism of communities on a SNS. Therefore, we combine the complex network analysis for understanding how network grows and the text mining analysis for understanding how people communicate.

2. COMPLEX SYSTEM ANAYSIS

Fig 1 shows a sample graph of the friend-network on F-LECCS SNS. On these networks, the network index, i.e.the network density, average degree, clustering coefficient, average path length and assortative mixing by degree, widely used for
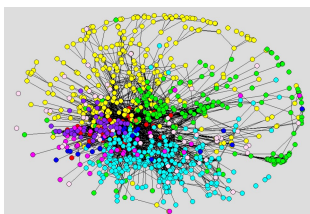

Fig.1 Friend-network on F-LECCS SNS

O.Yamakawa: Center for Arts and Sciences, Fukui Prefectural University, JAPAN, E-mail: yamakawa@fpu.ac.jp
T.Tagawa and H.Inoue: Computing and Communications Center, Kyushu University, JAPAN E-mail: tagawat@cc.kyushu-u.ac.jp , jin@cc.kyushu-u.ac.jp
K.Yasutake: Graduate Schoool of Social Sciences, Hiroshima University, JAPAN, E-mail: ystake@hiroshima-u.ac.jp
T.Sumiya: Information Media Center, Hiroshima University, JAPAN, E-mail: sumi@riise.hiroshima-u.ac.jp
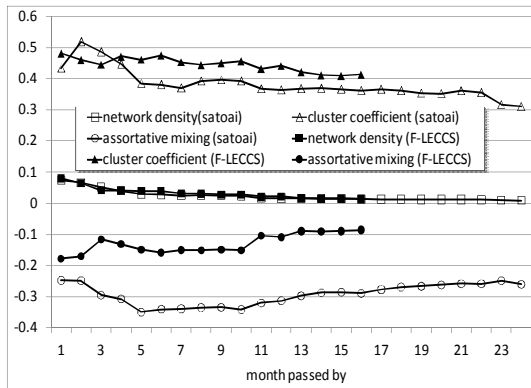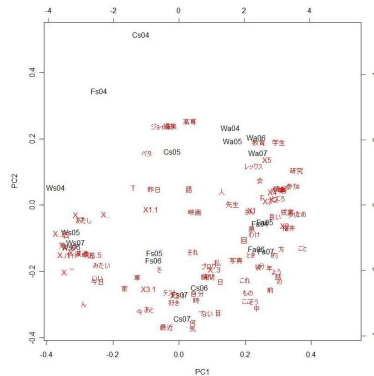
Fig.2 Variation of the network indexes

Fig.3 TF-IDF + principal component analysis

complex system analysis(Watts98), have been calculated by network analysis software pajek(Nooy et al. 2005) and igraph package for statistical software R.

In order to compare other friend-network on SNS, "Satoai-SNS" have been also analyzed. Satoai is the cross-university project in Sikoku area in JAPAN. The log data of satoai project for 24 months and F-LECCS project for 16 months data have been analyzed. The variation of the network indexes are shown in Fig.2.

Most indexes have same tendency except the assortative mixing by degree. This result means Satoai's network contains more star type sub-networks than F-LECCS's. In another words, F-LECCS has weaker tendency to connect each other depend on degree than Satoai.

3. TEXT MINING ANAYSIS

In order to understand the interaction of communication, the blogs on F-LECCS SNS have been analyzed by text mining method. Fig.3 shows the result of TF-IDF and principal component analysis. We have chosen the 100 important words in blogs by TF-IDF method and analyzed the words by the principal component analysis. Users have been divided into groups by their attributions i.e. universities, teachers or students. In fig.3, red characters are Japanese words and black ones are the groups at certain month. For example, Fs04 indicate University "F", Students "s", and April "04". Fig.3 shows us that the groups Wa, Fa and Ws stay nearly same position. However, the groups Cs and Fs moved from upper part to lower part. This result means that the contents of blogs in some groups did not change and those of the other groups changed for 4 months at the view point of the frequency in the use of important word.

4. CONCLUSION

The complex network analysis is used to understand the formal side of communities and the text mining analysis is for the interactive side of communities. We try to combine the two analyses to understand the growth mechanism of communities on SNS.

REFERENCES

Watts, D and Strogatz, S. (1998), Collective dynamics of 'small-world' networks,. Nature 393, 440-442.

Nooy, W. de, Mrvar, A., Batagelj, V. (2005). Exploratory Social Network Analysis with Pajek. Structural Analysis in the Social Sciences 27, Cambridge University Press.