

Community Focused Social Network Extraction

Masahiro Hamasaki¹, Yutaka Matsuo¹, Keisuke Ishida¹, Yoshiyuki Nakamura¹,
Takuichi Nishimura¹, and Hideaki Takeda²

¹ National Institute of Advanced Industrial Science and Technology (AIST),
Tokyo, Japan

² National Institute of Informatics (NII),
2-1-2 Hitotsubashi, Chiyoda-ku Tokyo, Japan

Abstract. A social networking service can become the basis for the information infrastructure of the future. For that purpose, it is important to extract social networks that reflect actual social networks which users have already had. Providing a simple means for users to register their social relations is also important. We propose a method that combines various approaches to extract social networks. Especially, three kinds of networks are extracted: user-registered *Know-link* networks; Web-mined *Web-link* networks; and face-to-face *Touch-link* networks. This paper describes the combination of social network extraction for an event-participant community. Analyses on the extracted social networks are also presented.

1 Introduction

This paper presents an integrated method for social network extraction. Social networks play important roles in our daily lives. Social networks overwhelmingly influence our lives without our knowledge of their implications. Many applications use social networks [12]. In the context of the Semantic Web, social networks are crucial to realize a web of trust, which enables the estimation of information credibility and trustworthiness [3][6]. Ontology construction is also related to social networks [9].

People pay attention to social networks not only for academic study, but also for commercial services. Social networking services (SNSs) have become popular. Friendster³ and Orkut⁴ are among the earliest and most successful SNSs. An interesting aspect of SNSs is that a users can visualize social networks (acquaintance lists) in addition to others' personal attributes (e.g., name, affiliation, hobby). Acquaintance lists reveal information about users' personalities. On the other hand, acquaintance lists are proof that acquaintances can track activity within the SNS. We can expect that the purview of acquaintances that such lists offer serves to restrict anti-social behavior within a community (e.g., assuming false names, abusive language). In fact, a famous Japanese SNS called mixi⁵

³ <http://www.friendster.com/>

⁴ <https://www.orkut.com/>

⁵ <http://mixi.jp/>

has three million users; 70% of them are active users. In all, mixi has 490,000 communities (BBS), but maintaining communications on online communities presents some difficulties [4]. An SNS that manages and stores social networks can become a base of our future information infrastructure.

Currently, we are doing an *Event Space Information Support Project*. We have targeted event spaces such as expositions or conventions because such events involve rich contents and many attendees. This project is intended to activate the community by supporting real-world-based interaction using Ubiquitous and Semantic Web technologies. Ishida [5] explained that community computing includes five functions for encouraging social interaction in communities: knowing each other, sharing preferences and knowledge, generating consensus, supporting everyday life, and assisting social events. We try to realize such functions using social networks.

What is a difference between an ordinary SNS and an SNS for an event participants' community? All SNS users use SNS. Nevertheless, not every event participant uses it. Furthermore event participants' communities have already had social networks and are created in event spaces and other places. Our necessary SNS should reflect actual social networks in that community. It requires a method to obtain social networks not only from user registration but also others. We called this challenge *Community Focused Social Network Extraction*.

We propose a new method to extract social networks. It is a combination of three methods. We targeted an academic conference as the first trial community for our proposed method and developed a system that has our proposed method. We operated our system at some academic conferences. We have analyzed their respective social networks. This paper describes characteristics of three means of extraction of social networks and discusses the effectiveness of their combination.

2 Community Focused Social Network Extraction

2.1 Approach of Social Network Extraction

This section presents a summary of methods to obtain social networks. Several means exist to obtain social networks: Friend-of-a-Friend (FOAF) is a vocabulary to describe information on a person and their relation to others. We can collect FOAF files and obtain a FOAF network [2][8]. Users create both SNS data and FOAF data themselves.

On the other hand, automatic detection of relations is also possible from various sources of online information such as e-mail archives, schedule data, and Web citation information [1][14][10]. Especially in some works, social networks are extracted by measuring the co-occurrence of names on the Web using a search engine [8][7].

Another means has been explored to obtain social networks: observing persons' behaviors using ubiquitous and wearable devices [11].

Whichever method is taken for obtaining a social network, it suffers from some flaws. For example, SNS data and FOAF data, which are based on self-reporting, suffer from data bias and sparseness. Users might name some of their

work acquaintances, but they might not include private friends. Some name hundreds of friends, while others name only a few. People create SNSs data by selecting registered users on the SNS and FOAF data by naming others freely. Sparsity in FOAF data is more serious than SNSs data. Automatically obtained networks, e.g., Web-mined social networks, provide a good view of prominent persons, but they do not properly record relationships of novices, students, and other "normal" people. Social networks observed using wearable devices are constrained by device-specific characteristics: they might have detection errors, limitation of detection scopes, and biased usage by users.

2.2 Our Proposed Method

This paper presents a method to extract social networks for a specific community. In this case, we target an academic conference participants' community. For extracting social network from a community, it is important to obtain cooperation from community members. For that reason, we infer that it is important to create the initial network without demanding personal information input from the user to make the whole social network useful [13]. The initial network should be modified according to user interaction to the system. For example, the initial information from the user is from the web system that has a click button to show that a user knows this person or is interested in some content, which preference might resemble the preferences of other users. Such information from web systems is added to the social network. Real-world-based user interaction information is also added, e.g., if three users used the same table together and the same demonstration was visited simultaneously by two other users.

Figure 1 shows our proposed methods. Ordinarily, a user registers a social network (1) in an SNS. We have proposed methods to extract social networks using user interaction (2) and web mining (3). Conventional methods cannot extract social networks among a community automatically and improve them using user-system interactions. We proposed a combined social network extraction method that includes many web services based on community interests and contents and onsite systems that have been deployed in real-world space for supporting mobile users in the site.

3 POLYPHONET Conference

We developed *POLYPHONET Conference* (hereafter, *POLYPHONET*) that has our proposed method. The system is a community support system whose target is an academic conference.

POLYPHONET Conference has functions as a social networking service and a conference scheduling system. A user can find research topics that a researcher is exploring or with whom she is working. In the scheduling part, a user can register interesting presentations (papers, demos and posters)

Our proposed method thereby comprises three means to extract social networks. The first is based on web mining techniques. This method can create initial networks automatically from available web information. The second is based

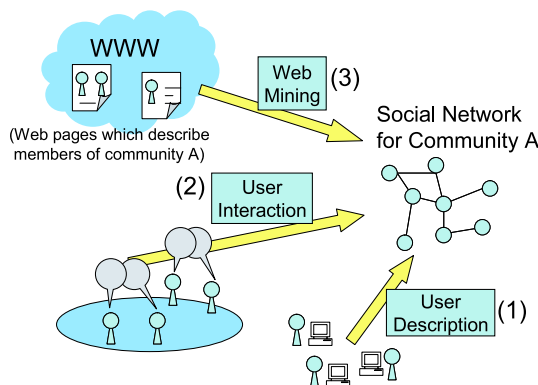


Fig. 1. Proposed Method

on real-world-based user interaction in communities. Furthermore, the method captures user interactions in the conference room. The last is based on user interaction on the web system. It gathers user clicks of acquaintance buttons: it is similar to SNS. We call the first a *Web-link* and the second a *Touch-link*; the last is a *Know-link*.

A *Web-link* is extracted from the Web using Web Mining techniques. We applied the web mining method based on method [7] to extract a social network among participants. That method is based on measuring the relevance of two nodes based on the number of retrieved results obtained by a search engine query.

Users register a *Touch-link* via an information kiosk. We set several information kiosk in a public space and deliver each participant an ID card as a name card. They can view social networks among them and compare personal schedule if two or three participants place ID cards there together. Then the social-tie "We meet and see social networks together" is added to *POLYPHONET* automatically. It seemed to provide easy and a good balance between privacy and effectiveness because user understands how and to whom the relation will be created.

Users on the web system can register a *Know-link*. A user can make an addition to the "I-know" list when that user finds an acquaintance. At that time, the acquaintance is also added to the acquaintance's "I'm-known-by" list.

The system has a portal page that is tailored to an individual user, called *my page*. The user's presentations, bookmarks of presentations, and registered acquaintances are shown along with the social network extracted from the Web. It helps users to register *Knows-link* easily by seeing the *Web-links*.

4 Field Test

We tested our system at 17th, 18th and 19th Annual Conferences of the Japan Society of Artificial Intelligence (JSAI2003, JSAI2004, and JSAI2005) and at

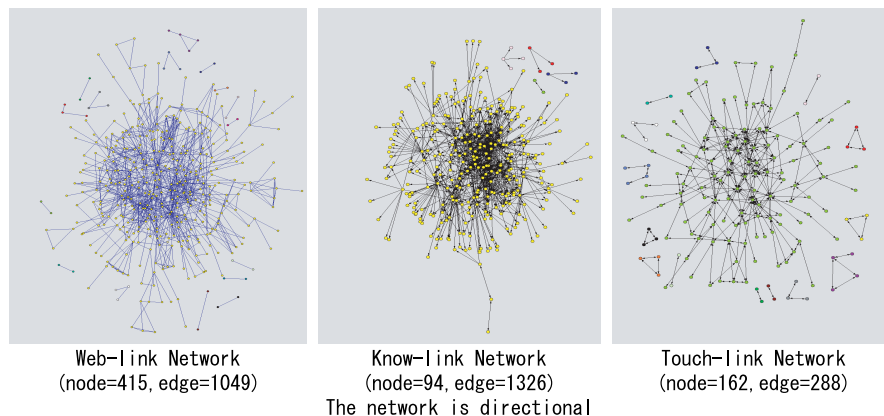


Fig. 2. Three Types of Social Networks.

The International Conference on Ubiquitous Computing (UbiComp 2005) and analyzed obtained social networks. In this paper, we show some results of analyses on obtained networks from JSAI2005.

That conference included 297 presentations and 579 authors (including co-authors). About 500 participants joined that conference: its size was similar to that of a conference at which we tested the previous system. The system started at 5th June. In all, our system's users were 217.

POLYPHONET extracts social networks among participants using three methods. First, the system extracts them using Web mining technique as initial data (*Web-link*). Second, users can register their own social networks with our system (*Know-link*). Third, two or three users can register their face-to-face meeting through information kiosks (*Touch-link*).

As a result, *Web-link* has 484 nodes and 34,880 links. *Touch-link* has 162 nodes and 288 links. *Know-link* has 94 nodes and 1,326 links. Figures 2 show the respective networks. In these figures, the *Web-link* network threshold is controlled to reduce the number of edges and allow clear visualization.

Web-links are more numerous than others, which serves our system well. It uses *Web-links* as initial data of social networks. It is therefore desirable that every user has initial personal data. The number of users of *Touch-link* is larger than *Know-link* even though users can create *Touch-link* for only three days. It indicates *Touch-link* can provide an easy way to register social networks for community members.

About half of the *Know-links* are common with *Web-links*, indicating the initial network validity. From user log analysis, 52% of the *Know-links* are registered from a user's *Web-links* page. This suggests that at least in *POLYPHONET*, the *Web-link* contributes to set *Know-links* efficiently.

Figure 3 shows the number of Web hit (by putting a person's name to a search engine) versus the number of three kinds of links. The more authoritative

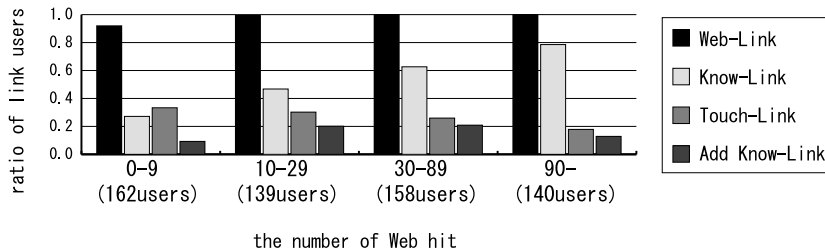


Fig. 3. Ration of Link Users at JSAI2005. 'Know-Link' means users who have undirected *Know link* and 'Add Know-Link' means users who added *Know link*.

people (with lots of hit count) tend to have more number of *Web links*. While the most authoritative people do not use *Knows links* the most; active middle-authoritative users use the most. They may know well about the community, and feel interesting. *Touch links* are used by the less authoritative users more; especially, the persons with the same level of authoritative are likely to have meets link. It is natural because persons who have fewer acquaintances want more acquaintances, and people are likely to meets people with the same social level.

Community focused social network extraction should reflect social network of a target community. Our target community is participants of an academic conference and it has various members. Results of this field-test shows our proposed method extracts social network from such various members.

Every *Web-link*, *Touch-link*, and *Know-link* has own characteristic even though they mean same social relationship. When we make applications with three links, we should improve a way to integrate them for each application. For example, *Web-link* is more effective when we use social network as context like a presentation page in *POLYPHONET* because *Web-link* can cover relationship of authoritative people. When we use social network for communication support, it seems that *Touch-link* and *Know-link* is more effective because they are created by active users. Especially, *Touch-link* is a key because actually a person meets others on site, the person feel easy talking again to introduce someone.

5 Conclusion

This paper presents community-focused social networks and extraction methods. We implemented our proposed method to *POLYPHONET Conference* and applied it to the academic conference. Results of field-testing show that our proposed method can realize unbiased extraction of social networks and provide a simple means to register social networks for users.

Community Focused Social Network Extraction is important for using SNS as an information infrastructure of one community. Social network is related to Semantic Web technology. If SNS is approved as an information infrastructure,

the achievement of advanced information support by Semantic Web technology becomes possible, too. I hope this research becomes the help. Future studies will address network integration methods, especially how to integrate various kinds of edges, and novel applications using social networks for supporting communities.

6 Acknowledgments

This research has been supported by NEDO (New Energy and Industrial Technology Development Organization) as the project ID of 04A11502a.

References

1. L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social Networks*, 25(3):211–230, 2003.
2. T. Finin, L. Ding, and L. Zou. Social networking on the semantic web. *The Learning Organization*, 2005.
3. J. Golbeck and J. Hendler. Inferring trust relationships in web-based social networks. *ACM Transactions on Internet Technology*, 2005.
4. J. Grudin. Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM*, 37(1):99–105, 1994.
5. T. Ishida, editor. *Community Computing: Collaboration over Global Information Networks*. John Wiley and Sons, 1998.
6. P. Massa and P. Avesani. Controversial users demand local trust metrics: an experimental study on epinions.com community. In *Proceedings of AAAI-05*, 2005.
7. Y. Matsuo, J. Mori, M. Hamasaki, H. Takeda, T. Nishimura, K. Hashida, and M. Ishizuka. Polyphonet: An advanced social network extraction system. In *Proceedings of WWW2006*, to appear.
8. P. Mika. Flink: Semantic web technology for the extraction and analysis of social networks. *Journal of Web Semantics*, 3, 2005.
9. P. Mika. Ontologies are us: A unified model of social networks and semantics. In *Proceedings of ISWC2005*, 2005.
10. T. Miki, S. Nomura, and T. Ishida. Semantic web link analysis to discover social relationship in academic communities. In *Proceedings of SAINT2005*, 2005.
11. A. Pentland. Socially aware computation and communication. *IEEE Computer*, 2005.
12. S. Staab, P. Dmingos, T. Finin, P. Mika, A. Joshi, J. Golbeck, A. Nowak, L. Ding, and R. R. Valleecher. Social network applied. *IEEE Intelligent systems*, pages 80–93, 2005.
13. H. Takeda and I. Ohmukai. Building semantic web applications as information/knowledge sharing systems. In *Proceedings of End User Aspects of the Semantic Web*, 2005.
14. J. Tyler, D. Wilkinson, and B. A. Huberman. Email as spectroscopy: automated discovery of community structure within organizations. *Communities and technologies*, pages 81–96, 2003.