An Integrated Method for Social Network Extraction

Masahiro Hamasaki(1) Yutaka Matsuo(1) Keisuke Ishida(1) Tom Hope(1) Takuichi Nishimura(1) Hideaki Takeda(2)
(1) National Institute of Advanced Industrial Science and Technology (AIST), JAPAN
(2) National Institute of Informatics (NII), JAPAN

1. Background and Purpose

A social network can become bases for information infrastructure in the future. It is important to extract social networks that are not biased. Providing a simple means for users to register their social relation is also important.

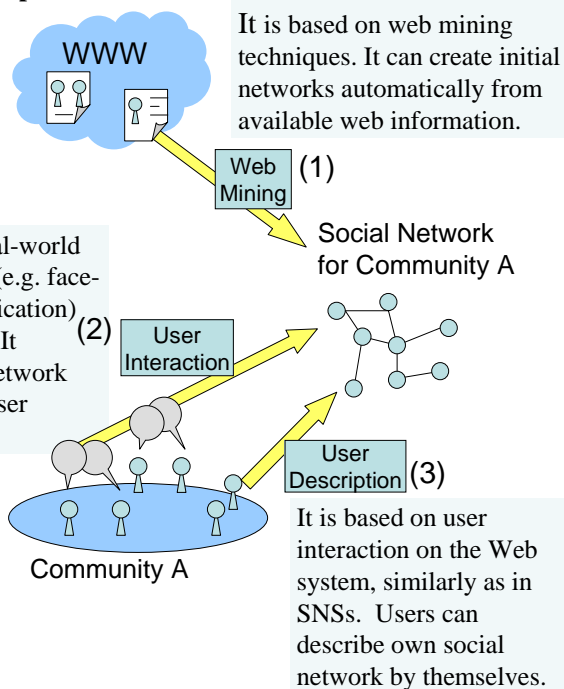There are several ways to obtain social networks:
- FOAF (Friend-of-a- Friend)
- Social Networking Service
- Automatic detection of relation (e.g., from e-mail, schedule)
- Web citation information
- Observing persons' behaviors in wearable devices, etc.

➡ **Whichever method we will take for obtaining a social network, it suffers from some degree of flaws**
- SNS data and FOAF data, which are based on self-reporting, suffer from data bias and sparseness.
- Automatically obtained networks do not properly record relationships of novices, students, and other "normal" people.
- Using wearable devices are constrained by device-specific characteristics: they might have detection errors, limitation of detection scopes, and biased usage by users.

Our Proposal

We propose a new method that is a combination of three methods.
(1) **Web mining**
(2) **User interaction in the real world**
(3) **User description**

It is based on web mining techniques. It can create initial networks automatically from available web information.

WWW

Web Mining (1)

Social Network for Community A

It is based on real-world user interaction (e.g. face-to-face communication) in communities. It extracts social network with capturing user interactions.

(2) User Interaction

User Description (3)

Community A

It is based on user interaction on the Web system, similarly as in SNSs. Users can describe own social network by themselves.

2. Three Extraction Methods

(1) Web Mining

It extracts relationship among users using cooccurrence on the Web. The simplest approach is to measure the relevance of two nodes based on the number of retrieved results obtained by a search engine query.
- "Masahiro Hamasaki" and "Yutaka Matsuo" : 82hit
- "Masahiro Hamasaki" and "Tim Berners Lee" : 4hit
- "Yutaka Matsuo" : 13,800hit
- "Tim Berners Lee" : 4,830,000hit

We can find the relevance of "M. Hamasaki" and "Y. Matsuo" is stronger than "M. Hamasaki" and "Tim B. Lee".

The R(x,y) represents the relevance of nodes x and y

$$R(X,Y) = \begin{cases} \dfrac{|X \cap Y|}{min(|X|,|Y|)} & \text{if } |X| > k \text{ and } |Y| > k, \\ 0 & \text{otherwise} \end{cases}$$

➡ **It can obtain social networks automatically**

(2) User interactions on the real world

Information kiosks are in the conference room. Each information kiosk has some IC card readers and Users can view social networks among them if two or three participants place IC cards there together. It can serve as a real world name card exchange. The system captures such a behavior as a creating relationship.

IC card

Social network among two users

Two Users at an Information Kiosk

➡ **It can capture new relationship which is created in the event space**

(3) User interaction on the Web system

Users can register their friends and acquaintances, as they can with other SNSs. But it does not need mutual certification to add links among users because every author is not active user.

➡ **User can modify and complement social network**

My Page

An Integrated Method for Social Network Extraction

Masahiro Hamasaki(1) Yutaka Matsuo(1) Keisuke Ishida(1) Tom Hope(1) Takuichi Nishimura(1) Hideaki Takeda(2)

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

(2) National Institute of Informatics (NII), JAPAN

3. Polyphonet Conference

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

System Functions:
-Displaying social network among participants
-Information provision, scheduling support, and recommendation
 about presentations

System Purpose:
-Efficient communication and information exchange through
 finding organic links among researchers and presentations on the web

System overview

The system is a web application; users can access it via PC or information kiosk.

(a) The system extracts social networks from the web (We call '*Web links*')

(b) Users use an online timetable and register their social networks. (We call '*Know links*')

(c) They can view mutual social relationships. Information kiosks capture their interaction. (We call '*Touch links*')
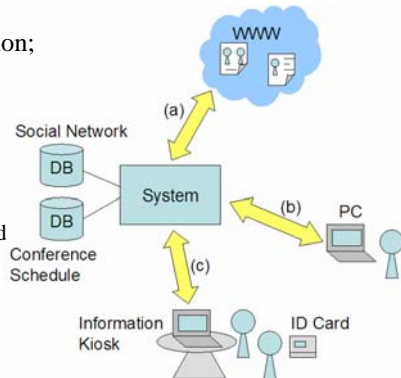
Applications with Social Networks

Social Network as a Context

The system has a presentation page. It has a surrounding social network and users can view those associated persons. A presentation page shows a social network among authors and users who are interested in that presentation.

Social Network Search

Users can seek shortest paths among users and see details of their mutual relationships.

Information Recommendation

Users can seek shortest paths among users and see details of their mutual relationships.

4. Field Test

We operated our system at a three-day Japanese domestic conference (JSAI2006). That conference had 297 presentations and 579 authors (including co-authors). About 500 participants joined that conference.
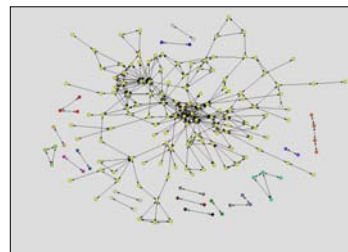
- The number of active users: 217
  - Users who added *know link* : 94
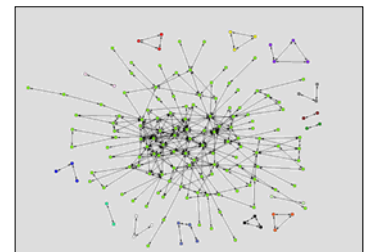 - Users who added *touch link*: 162
- The number of links
  - Web link : 34,880
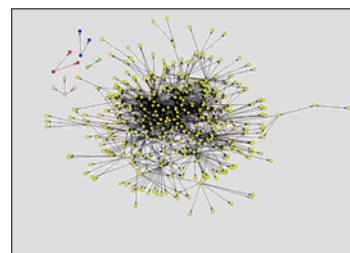 - Know link : 1,326
  - Touch link : 288

Social Networks

(1) *Web links*

(2) *Touch links*

(3) *Know links*

(1)*Web links*: There are several sub-communities exist, where nodes are especially densely connected.
(2)*Touch-link*: There is only a small dense part in the network. Users do not use information kiosk exhaustively with their friends.
(3)*Know-link*: It tends to be more concentrated. Several participants have a large number of (out-)edges.

User Categories and Social Networks

The more authoritative people (with lots of hit count) tend to have more number of *Web links*. Active middle-authoritative users use *Know links* the most. The less authoritative users use *Touch links* the most.
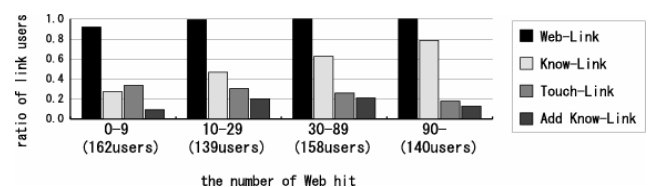
5. Conclusion

Our proposed method can realize unbiased extraction of social networks and provide a simple means to register social networks.
-Web mining can extract social networks from the community evenly
-Real world interaction support can provide an easy way to register
 social networks