Real-world Oriented Information Sharing Using Social Networks

Junichiro Mori
Graduate School of
Information Science
University of Tokyo
Tokyo, Japan
jmori@miv.t.u-tokyo.ac.jp

Tatsuhiko Sugiyama
UNICUS Co.,Ltd.
Yokohama, Japan
sugiyama@unicus.jp

Yutaka Matsuo
National Institute of Advanced
Industrial Science and
Technology
Tokyo, Japan
y.matsuo@carc.aist.go.jp

ABSTRACT

While users disseminate various information in the open and widely distributed environment of the Semantic Web, determination of who shares access to particular information is at the center of looming privacy concerns. We propose a real-world-oriented information sharing system that uses social networks. The system automatically obtains users' social relationships by mining various external sources. It also enables users to analyze their social networks to provide awareness of the information dissemination process. Users can determine who has access to particular information based on the social relationships and network analysis.

Categories and Subject Descriptors

H.5.3 [**Information Interfaces AND Presentation**]: Group and Organizational Interfaces—*collaborative computing, computer-supported cooperative work, web-based interaction*

General Terms

Design, Human Factors

Keywords

Information sharing, Social network

1. INTRODUCTION

With the current development of tools and sites that enable users to create Web content, users have become able to easily disseminate various information. For example, users create Weblogs, which are diary-like sites that include various public and private information. Furthermore, the past year has witnessed the emergence of social networking sites that allow users to maintain an online network of friends or associates for social or business purposes. Therein, data related to millions of people and their relationships are publicly available on the Web.

Although these tools and sites enable users to easily disseminate information on the Web, users sometimes have difficulty in sharing information with the right people and frequently have privacy concerns because it is difficult to determine who has access to particular information on such applications. Some tools and applications provide control over information access. For example, Friendster, a huge social networking site, offers several levels of control from "public information" to "only for friends". However, it provides only limited support for access control.

An appropriate information sharing system that enables all users to control the dissemination of their information is needed to use tools and sites such as Weblog, Wiki, and social networking services fully as an infrastructure of disseminating and sharing information. In the absence of such a system, a user would feel unsafe and would therefore be discouraged from disseminating information.

How can we realize such an information sharing system on the Web? One clue exists in the information sharing processes of the real world. Information availability is often closely guarded and shared only with the people of one's social relationships. Confidential project documents which have limited distribution within a division of company, might be made accessible to other colleagues who are concerned with the project. Private family photographs might be shared not only with relatives, but also with close friends. A professor might access a private research report of her student. We find that social relationships play an important role in the process of disseminating and receiving information. This paper presents a real-world oriented information sharing system using social networks. It enables users to control the information dissemination process within social networks.

The remainder of this paper is organized as follows: section 2 describes the proposed information sharing system using social networks. In section 3, we describe the application of our system. Finally, we conclude this paper in section 4.

2. INFORMATION SHARING USING SOCIAL NETWORKS

Figure 1 depicts the architecture of the proposed information sharing system. The system functions as a "plug-in" for applications so that external applications enable users to leverage social networks to manage their information dis-

GROUP'05, November 6-9, 2005, Sanibel Island, Florida, USA.

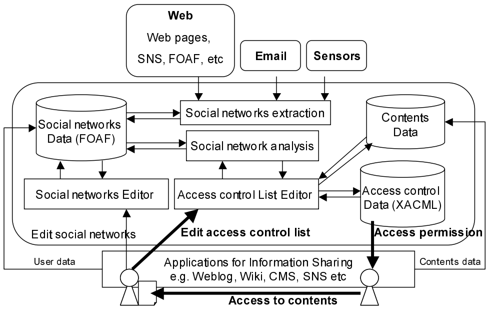


Figure 1: Architecture of the proposed information sharing system

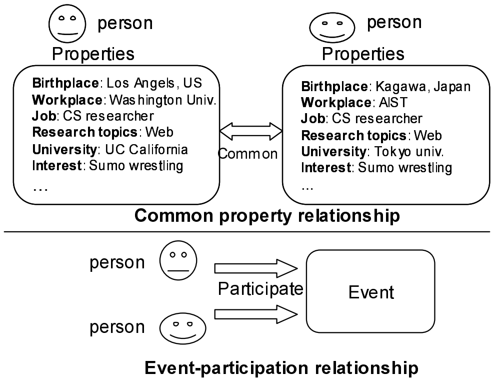


Figure 2: Two kinds of relationships

semination. A user can attach an access control list to his content using his social network when creating content on an application. Then, when the application receives a request to access the content, it determines whether to grant the request based on the access control list.

Because users determine the access control to information based on the social network, the system requires social network data. The system obtains users' social networks automatically by mining various external sources such as Web, emails, and sensor information; subsequently, it maintains a database of the social network information. Users can adjust the network if necessary.

The system enables users to analyze their social network to provide awareness of the information dissemination process within the social network. Using social relationships and the results of social network analyses, users can decide who can access their information.

Currently, the proposed system is applied to an academic society because researchers have various social relationships (e.g., from a student to a professor, from a company to a university) through their activities such as meetings, projects, and conferences. Importantly, they often need to share various information such as papers, ideas, reports, and schedules. Sometimes, such information includes private or confidential information that ought only to be shared with appropriate people. In addition, researchers have an interest in managing the information availability of their social relationships. The information of social relationships of an academic society, in particular computer science, is easily available online to a great degree. Such information is important to obtain social networks automatically.

Hereafter, we explain in detail how social networks are modeled, extracted and analyzed. Then we explain how users can decide to control information access using social networks.

2.1 Representation of Social Relationships

With the variety of social relationships that exist in the real world, a salient problem has surfaced: integration and consolidation on a semantic basis. The representation of social relationships must be sufficiently fine-grained that we can capture all details from individual sources of information in a way that these can be recombined later and taken as evidence of a certain relationship.

Several representations of social relationships exist. For example, social network sites often simplify the relationship as "friend" or "acquaintance". In the Friend of a Friend (FOAF) [1] vocabulary, which is one of the Semantic Web's largest and most popular ontologies for describing people and whom they know, many kinds of relationships between people are deliberately simplified as "knows" relations. A rich ontological consideration of social relationships is needed for characterization and analysis of individual social networks.

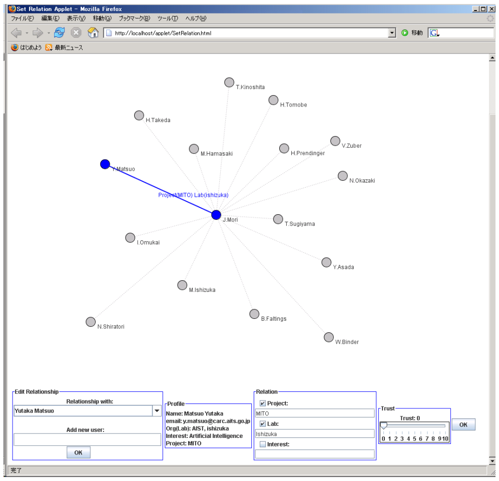
We define two kinds of social relationship (Fig. 2) [7]. The first basic structure of social relationship is a person's participation in an event. Social relationships come into existence through events involving two or more individuals. Such events might not require personal contact, but they must involve social interaction. From this event, social relationships begin a lifecycle of their own, during which the characteristics of the relationship might change through interaction or the lack thereof. An event is classified as *perdurant* in the DOLCE ontology [6], which is a popular ontology. For example, an event might be a meeting, a conference, a baseball game, a walk, etc. Assume that person A and person B participate in Event X. In that situation, we note that A and B share an *event co-participation relationship* under event X.

A social relationship might have various social roles associated with it. For example, a student-professor relationship within a university setting includes an individual playing the role of a professor; another individual plays the role of a student. If A and B take the same role to Event X, they are in a *same role relationship* under event X (e.g., students at a class, colleagues in a workspace). If A cannot take over B's role or vice versa, A and B are in a *role-sharing relationship* (e.g., a professor and students, a project leader and staff).

Another kind of social relationship is called a *common property relationship*. Sharing the same property value generates a common property relationship between people. For example, person A and person B have a common working place, common interests, and common experiences. Consequently, they are in a common property relationship with regard to those common properties.

2.2 Extraction of Social Networks

If two persons are in either an event co-participation relationship or a common property relationship, they often communicate. The communication media can be diverse:



Edit Relationship

Relationship with: Yutaka Matsuo

Add new user:

OK

Profile

Name: Matsuo Yutaka
email: y.matsuo@carc.aits.go.jp
Org(Lab): AIST, ishizuka
Interest: Artificial Intelligence
Project: MITO

Relation

☑ Project: MITO
☑ Lab: Ishizuka
☐ Interest:

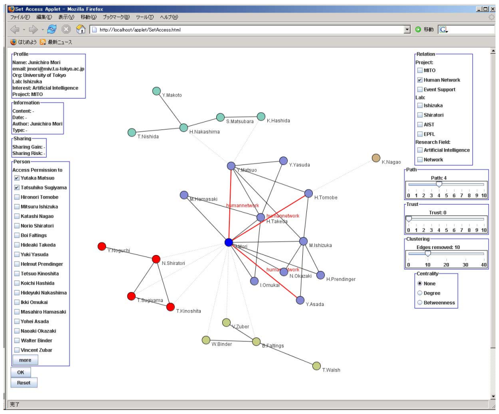
Trust

Trust: 0

0 1 2 3 4 5 6 7 8 9 10

OK

Figure 3: Editor for social relationships



Profile

Name: Junichiro Mori
email: jmori@miv.t.u-tokyo.ac.jp
Org: University of Tokyo
Lab: Ishizuka
Interest: Artificial Intelligence
Project: MITO

Information

Content: -
Date: -
Author: Junichiro Mori
Type: -

Sharing

Sharing Gain: -
Sharing Risk: -

Person

Access Permission to

☑ Yutaka Matsuo
☑ Tatsuhiko Sugiyama
☐ Hironori Tomobe
☐ Mitsuru Ishizuka
☐ Katashi Nagao
☐ Norio Shiratori
☐ Boi Faltings
☐ Hideaki Takeda
☐ Yuki Yasuda
☐ Helmut Prendinger
☐ Tetsuo Kinoshita
☐ Koichi Hashida
☐ Hideyuki Nakashima
☐ Ikki Omukai
☐ Masahiro Hamasaki
☐ Yohei Asada
☐ Naoaki Okazaki
☐ Walter Binder
☐ Vincent Zubar

more

OK

Reset

Relation

Project:
☐ MITO
☑ Human Network
☐ Event Support

Lab:
☐ Ishizuka
☐ Shiratori
☐ AIST
☐ EPFL

Research Field:
☐ Artificial Intelligence
☐ Network

Path

Path: 4

0 1 2 3 4 5 6 7 8 9 10

Trust

Trust: 0

0 1 2 3 4 5 6 7 8 9 10

Clustering

Edges removed: 10

0 10 20 30 40

Centrality

● None
○ Degree
○ Betweenness

Figure 4: Editor for analyzing social networks and assigning an access control list to content

face-to-face conversation, telephone call, email, chat, online communication on Weblogs, and so on. If we wish to discover the social relationship by observation, we must estimate relationships from superficial communication. The emerging field of social network mining provides methods for discovering social interactions and networks from legacy sources such as web pages, databases, mailing lists, and personal emails.

Currently, we use three kinds of information sources to obtain social relationships using mining techniques. From the Web, we extract social networks using a search engine and the co-occurrence of two persons' names on the Web. Consequently, we can determine the following relationships among researchers: Coauthor, Same affiliation, Same project, Same event (participants of the same conference, workshop, etc.) [8]. Coauthor and Same event correspond to an event coparticipation relationship. Same affiliation and same project correspond to a common property relationship. We are also using other sources such as email and sensors (we are developing a device that detects users within social spaces such as parties and conferences) to obtain social relationships.

Necessarily, the quality of information obtained by mining is expected to be inferior to that of manually authored profiles. We can reuse those data if a user has already declared his relationships in FOAF or profiles of social networking services. Although users might find it difficult and demanding to record social relations, it would be beneficial to ask users to provide information to obtain social relationships.

In addition to the relationship type, another factor of the social relationship is tie strength. Tie strength itself is a complex construct of several characteristics of social relations. It is definable as affective, frequency, trust, complementarity, etc. No consensus for defining and measuring them exists, which means that people use different elicitation methods when it comes to determining tie strength. For example, Orkut, a huge social networking service, allows description of the strength of friendship relations on a five-point scale from "haven't met" to "best friend", whereas other sites might choose other scales or terms.

In our system, we use trust as a parameter of tie strength. Trust has several very specific definitions. In [4], Golbeck describes trust as credibility or reliability in a human sense: "how much credence should I give to what this person speaks about" and "based on what my friends say, how much should I trust this new person?" In the context of information sharing, trust can be regarded as reliability regarding "how a person will handle my information". Users can give trust directly in a numerical value to a person in his relation. Alternatively, trust is obtainable automatically as authoritativeness of each person using the social network [8].

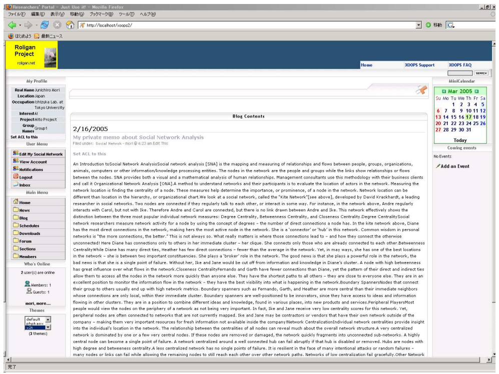
The obtained social network data are integrated as extended FOAF files and stored in database. Users can adjust networks if needed (Fig. 3). The social relationship and its tie strength become guiding principles when a user determines an access control list to information.

2.3 Social Network Analysis for Information Sharing

The system enables users to analyze their social networks to provide awareness of the information dissemination process within the social network.

Social network analysis (SNA) is distinguishable from other fields of sociology by its focus on relationships between actors rather than attributes of actors, a network view, and a belief that structure affects substantive outcomes. Because an actor's position in a network affects information dissemination, SNA provides an important implication for information sharing on the social network. For example, occupying a favored position means that the actor will have better access to information, resources, and social support.

The SNA models are based on graphs, with graph measures, such as centrality, that are defined using a sociological interpretation of graph structure. Freeman proposes numerous ways to measure centrality [2]. Considering a social network of actors, the simplest measure is to count the number of others with whom an actor maintains relations. The actor with the most connections, the highest degree, is most central. This measure is called *degreeness*. Another measure is *closeness*, which calculates the distance from each actor in the network to every other actor based on connections among all network members. Central actors are closer to all others than are other actors. A third measure is *betweenness*, which examines the extent to which an actor is situated among others in the network, the extent to which



Real Name Junichiro Mori
Location Japan
Occupation Ishizuka Lab. at Tokyo University
Interest AI
Project Mito Project
Group Names Group1

Blog Contents

2/16/2005

My private memo about Social Network Analysis

Filed under: Social Netwok - mori @ 6:23 am Edit This

Set ACL to this

An Introduction toSocial Network AnalysisSocial network analysis [SNA] is the mapping and measuring of relationships and flows between people, groups, organizations, animals, computers or other information/knowledge processing entities. The nodes in the network are the people and groups while the links show relationships or flows between the nodes. SNA provides both a visual and a mathematical analysis of human relationships. Management consultants use this methodology with their business clients and call it Organizational Network Analysis [ONA].A method to understand networks and their participants is to evaluate the location of actors in the network. Measuring the network location is finding the centrality of a node. These measures help determine the importance, or prominence, of a node in the network. Network location can be different than location in the hierarchy, or organizational chart.We look at a social network, called the "Kite Network"[see above], developed by David Krackhardt, a leading researcher in social networks. Two nodes are connected if they regularly talk to each other, or interact in some way. For instance, in the network above, Andre regularly interacts with Carol, but not with Ike. Therefore Andre and Carol are connected, but there is no link drawn between Andre and Ike. This network effectively shows the distinction between the three most popular individual network measures: Degree Centrality, Betweenness Centrality, and Closeness Centrality.Degree CentralitySocial network researchers measure network activity for a node by using the concept of degrees – the number of direct connections a node has. In the kite network above, Diane has the most direct connections in the network, making hers the most active node in the network. She is a 'connector' or 'hub' in this network. Common wisdom in personal networks is "the more connections, the better." This is not always so. What really matters is where those connections lead to – and how they connect the otherwise unconnected! Here Diane has connections only to others in her immediate cluster – her clique. She connects only those who are already connected to each other.Betweenness CentralityWhile Diane has many direct ties, Heather has few direct connections – fewer than the average in the network. Yet, in may ways, she has one of the best locations in the network – she is between two important constituencies. She plays a 'broker' role in the network. The good news is that she plays a powerful role in the network, the bad news is that she is a single point of failure. Without her, Ike and Jane would be cut off from information and knowledge in Diane's cluster. A node with high betweenness has great influence over what flows in the network.Closeness CentralityFernando and Garth have fewer connections than Diane, yet the pattern of their direct and indirect ties allow them to access all the nodes in the network more quickly than anyone else. They have the shortest paths to all others – they are close to everyone else. They are in an excellent position to monitor the information flow in the network – they have the best visibility into what is happening in the network.Boundary SpannersNodes that connect their group to others usually end up with high network metrics. Boundary spanners such as Fernando, Garth, and Heather are more central than their immediate neighbors whose connections are only local, within their immediate cluster. Boundary spanners are well-positioned to be innovators, since they have access to ideas and information flowing in other clusters. They are in a position to combine different ideas and knowledge, found in various places, into new products and services.Peripheral PlayersMost people would view the nodes on the periphery of a network as not being very important. In fact, Ike and Jane receive very low centrality scores for this network. Yet, peripheral nodes are often connected to networks that are not currently mapped. Ike and Jane may be contractors or vendors that have their own network outside of the company – making them very important resources for fresh information not available inside the company!Network CentralizationIndividual network centralities provide insight into the individual's location in the network. The relationship between the centralities of all nodes can reveal much about the overall network structure.A very centralized network is dominated by one or a few very central nodes. If these nodes are removed or damaged, the network quickly fragments into unconnected sub-networks. A highly central node can become a single point of failure. A network centralized around a well connected hub can fail abruptly if that hub is disabled or removed. Hubs are nodes with high degree and betweeness centrality.A less centralized network has no single points of failure. It is resilient in the face of many intentional attacks or random failures – many nodes or links can fail while allowing the remaining nodes to still reach each other over other network paths. Networks of low centralization fail gracefully.Other Network

Figure 5: Web site for sharing research information

information must pass through them to get to others, and consequently, the extent to which they are exposed to information circulation within the network. If the betweenness of an actor is high, it frequently acts as a local bridge that connects the individual to other actors outside a group. In terms of network ties, this kind of bridge is well known as Granovetter's "weak tie" [5], which contrasts with "strong tie" within a densely-closed group.

As the weak tie becomes a bridge between different groups, a large community often breaks up to a set of closely knit group of individuals, woven together more loosely according to occasional interaction among groups. Based on this theory, social network analysis offers a number of clustering algorithms for identifying communities based on network data.

The system provides users with these network analyses (Fig. 4) so that they can decide who can access their information. For example, if user wants to diffuse her information, she might consider granting access to a person (with certain trust) who has both high degreeness and betweenness. On the other hand, she must be aware of betweenness when the information is private or confidential. Clustering is useful when a user wishes to share information within a certain group.

3. APPLICATION

To demonstrate and evaluate our system, we developed a community site (Fig. 5) using communication tools such as Weblogs, Wikis, and Forums. By that system, studies from different organizations and projects can be disseminated and their information thereby shared. Users can share various information such as papers, ideas, reports, and schedules at the site. Our system is integrated into a site that provides access control to that information. Integrating our system takes advantage of the open and information nature of the communication tools. It also maintains the privacy of the content and activities of those applications.

Users can manage their social networks (Fig. 3) and attach the access control list to their content (e.g., Blog entries, profiles, and Wiki pages) using extracted social relationships and social network analysis (Fig. 4).

Once a user determines the access control list, she can save it as her information access policy for corresponding content. The access policy is described using extended eXtensible Access Control Markup Language (XACML) and is stored in a database. She can reuse and modify the previous policy if she subsequently creates a similar content.

One feature of our system is that it is easily adaptable to new applications because of its plug-and-play design. We are planning to integrate it into various Web sites and applications such as social network sites and RSS readers.

4. RELATED WORKS AND CONCLUSIONS

Goecks and Mynatt propose a Saori infrastructure that also uses social networks for information sharing [3]. They obtain social networks from users' email messages and provide sharing policies based on the type of information. We obtain social networks from various sources and integrate them into FOAF files. This facilitates the importation and maintenance of social network data. Another feature is that our system enables users to analyze their social networks. Thereby, users can control information dissemination more effectively and flexibly than through the use of pre-defined policies.

As users increasingly disseminate their information on the Web, privacy concerns demand that access to particular information be limited. We propose a real-world oriented information sharing system using social networks. It enables users to control the information dissemination process within social networks, just as they are in the real world. Future studies will evaluate the system with regard to how it contributes to wider and safer information sharing than it would otherwise. We will also develop a distributed system that can be used fully on the current Web.

5. REFERENCES

[1] D. Brickley and L. Miller. FOAF: the 'friend of a friend' vocabulary. http://xmlns. com/foaf/0.1/, 2004.

[2] L. C. Freeman. Centrality in social networks: Conceptual clarification, *Social Networks*, Vol.1, pp.215–239, 1979.

[3] J. Goecks and E. D. Mynatt. Leveraging Social Networks for Information Sharing *In Proc. of CSCW'04*, 2004.

[4] J. Golbeck, J. Hendler, and B. Parsia. Trust networks on the semantic web, *in Proc. WWW 2003*, 2003.

[5] M. Granovetter. Strength of weak ties, *American Journal of Sociology*, Vol.18, pp.1360–1380, 1973.

[6] C. Masolo, S. Borgo, A. Gangemi, N. Guarinno, and A. Oltramari. WonderWeb Deliverable D18, http://wonderweb.semanticweb.org/deliverable/D18.shtml

[7] Y. Matsuo, M. Hamasaki, J. Mori, H. Takeda and K. Hasida. Ontological Consideration on Human Relationship Vocabulary for FOAF. *In Proc. of the 1st Workshop on Friend of a Friend, Social Networking and Semantic Web*, 2004.

[8] Y. Matsuo, H. Tomobe, K. Hasida, M. Ishizuka. Finding Social Network for Trust Calculation. *In Proc. of 16th European Conference on Artificial Intelligence*, 2004.