

# Metadata Description for Analyzing Web Social Networks in Academic Communities

**Takeru Miki, Saeko Nomura, Toru Ishida**  
Department of Social Informatics, Kyoto University,  
Kyoto, Japan  
{miki-t, saeko}@kuis.kyoto-u.ac.jp, ishida@i.kyoto-u.ac.jp

## Abstract

To discover how precisely links among researchers' Web sites capture social ties in academic communities, this study conducts Web citation analysis. We assessed 3,878 computer scientists' Web sites and developed Web link ontology that defines various classes of relationships among researchers. We applied annotating metadata to the links among frequently cited (linked) 200 researchers' sites and found: 1) by extracting intellectual ties, cocitation analysis finds more fine-grained academic communities, 2) sociocognitive ties analysis shows a community being tied with a specific class of relationship, 3) researchers in the same field are gathered closely and gatekeepers who bridges research fields appear.

## 1 Introduction

Necessity of "understanding the sociology of Web contents creation" [Broder *et al.*, 2000] is pointed out. Our study aims at making clear the structure of social community reflected on the Web and at proposing new use of Web as describing and analyzing information on social aspect.

We introduce Web citation analysis, an application of bibliographical citation analysis to the Web. In this study, the academic community is selected as a subject. Since the Web link structures have rather vague semantics [Goldbeck and Hendler, 2004] compared to bibliographical data, challenge is to clarify how precisely Web citation analysis captures ties in the academic community.

We propose ontology that defines classes of social relationships that is represented by links on researchers' Web sites, using the semantic Web technology. Here, we describe detailed class of researchers' relationships, and show the results of analyses on both intellectual ties which represent bibliographical citation relationships, and sociocognitive ties which represent social acquaintanceships.

We collected and analyzed the bibliographical citation data and personal Web sites created by computer scientists. They are obtained from *CiteSeer*<sup>1</sup>, an online

Table 1. Link Ontology

Link Class	Description	#Links
0 Link	Root class	220
1 Intellectual Link	Reference information	44
1.1 Research Reference	Research-related reference	153
1.1.1 People	Link to researcher	666
1.1.2 Work	Link to research work	624
1.1.3 Society	Link to journal or conference	90
1.1.4 Utility	Link to tools for research	20
1.2 Teaching Reference	Education-related reference	836
1.3 Private Reference	Links in non-professional context	26
2 Sociocognitive Link	Social relationship	50
2.1 Collaborator	Collaborator	41
2.1.1 Research Collaborator	Research collaborator	62
2.1.1.1 Project	Collaborator on some project	138
2.1.1.2 Coauthor	Coauthor	312
2.1.2 Teaching Collaborator	Collaborator on education	78
2.1.2.1 Invited Lecturer	People invited to class	19
2.1.3 Academic Committee	Co-organizer	185
2.2 Educational Relationship	Supervisor or student	0
2.2.1 Supervisor	Student to supervisor	106
2.2.2 Student	Supervisor to student	17
2.3 Organizer-Participant	Organizer or participant	0
2.3.1 Organizer	Organizer	57
2.3.2 Participant	Participant	44
2.4 Private Acquaintance	Informal relationship	0
2.4.1 Friend	Friendship	7
2.4.2 Family	Kinship	9
Total		3,804

citation library system. All articles list in the 'Computer Science Directory' and associated authors' Web sites were collected in November 2001. As a result, 3,878 researchers that had personal Web sites were targeted to be analyzed.

## 2 Link Ontology

To describe link semantics as metadata, we developed a prototype of ontology for link classification according to site creators' social relationship that a link indicates. The ontology defines a class hierarchy that classifies links. Table 1 shows a part of the ontology. It is developed incrementally and simultaneously with attaching metadata to 3,804 actual Web links.

The ontology is designed so as to divide linkage (<0 Link>) firstly into <1 Intellectual Link> and <2 Sociocognitive Link>. The former is the class of linkage that simply refers to knowledge about things or creators, while the latter is that of linkage that shows some social ties between the creators of linked two Web pages. The

<sup>1</sup> <http://citeseer.nj.nec.com/>

class <Intellectual Link> is classified further by the context of Web page topic like “research” or “education.” As for <Sociocognitive Link> typical kinds of links observed in academic Web sites are organized into a hierarchy. The major two class of human relationship are <Collaborator> (people who have worked together) and <Educational Relationship> (supervisor-student). The links showing that someone joined an international conference are interpreted as <Organizer-Participant> relationship. <Private Acquaintance> means informal ties like kinship or friendship.

Actual classification is based on some clues such as the anchor-text of a link, the surrounding text, and/or the contents of the page that the link points to.

### 3 Metadata-based Social Network Analysis

Using the ontology, metadata are attached manually to frequently linked 200 researchers’ Web sites in the gathered 3,878 sites. This chapter shows about two metadata-based analytical methods and their results to overcome vagueness of Web links.

#### 3.1 Analysis of Intellectual Ties

We applied author cocitation analysis to the link network to cluster sites’ (articles’) authors into research fields. The result is visually mapped to the network, which shows whether the network structure reflects academic fields. We confirmed that the cocitation result of only on <Intellectual ties> links extracts more detailed and fine-grained fields than that of all Web links.

#### 3.2 Analysis of Sociocognitive Ties

In order to extract relationships with bi-directional recognition, <Sociocognitive Link> and its subclasses except for <Organizer-Participant> are picked up as network edge and then the network is calculated. Figure 1 shows the largest connected component of the extracted network. An edge indicates existence of social acquaintanceship (not information reference) of people, and the line shape shows the class of the relationship. Clusters extracted by cocitation analysis are mapped onto the network by coloring nodes.

The analytical result shows interesting phenomena on the community. Topological groups on the social network are closely related to research fields. The gatekeepers interconnecting fields are also found clearly.

In addition, some groups connected by a specific kind of edges are found (e.g. T. Sandholm and his neighbors). Thus, network analysis of annotated link shows more intelligible status of the society.

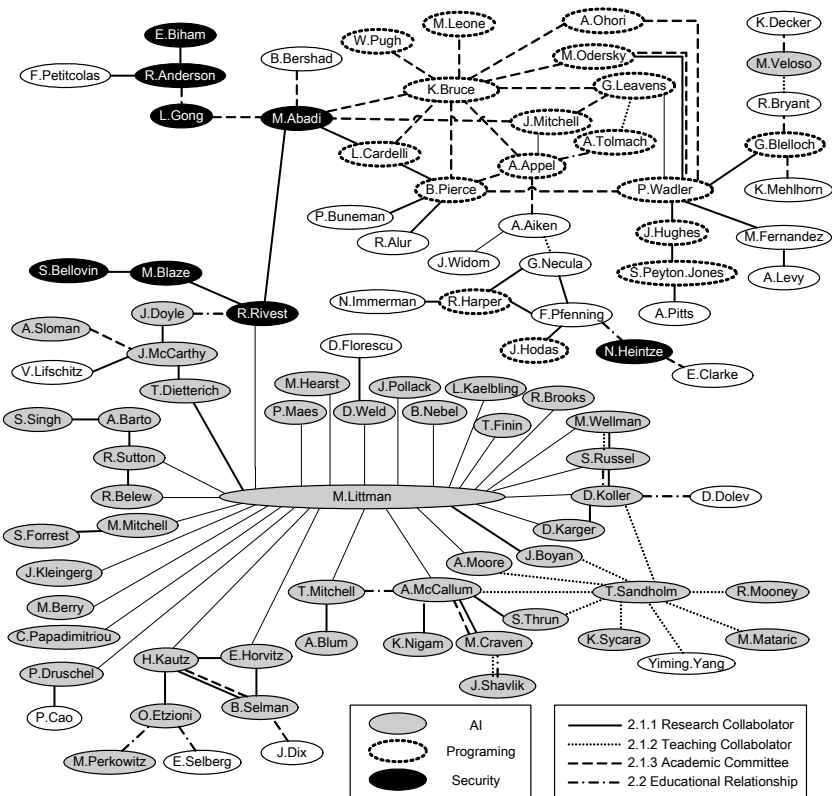


Figure 1. The Social Network with Edges of Bidirectional Acquaintance Links

### 4 Conclusion

We developed link ontology of social network. Analysis of annotated links in 200 researchers’ sites gives better understanding of their society. We found that 1) Cocitation analysis of referential links finds more detailed fields; 2) Annotated sociocognitive links show some community tied with particular kinds of relationships; 3) Gatekeepers being on the bridge points of research fields are definitely shown.

This study shows that a Web-based social network tool is an interesting application of semantic web, although understanding of correspondence to real world network remains as the future work.

### References

[Broder *et al.*, 2000] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. In *Proceedings of the ninth WWW Conference*, pages 309-320, 2000.

[Goldbeck and Hendler, 2004] Jeniffer Goldbeck and Jim Hendler. Accuracy of Metrics for Inferring Trust and Reputation in Semantic Web-based Social Networks. In *Proceedings of the fourteenth International Conference on Knowledge Engineering and Knowledge Management*, 2004.