

Semantic Web Link Analysis to Discover Social Relationships in Academic Communities

Takeru Miki, Saeko Nomura, Toru Ishida
Department of Social Informatics, Kyoto University,
Kyoto 606-8501, 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 the ties of academic communities, this study conducts Web citation analysis. 3,878 Web sites of computer scientists and their 8,811 articles are assessed. 200 most frequently cited researchers from them are subjected to Web author cocitation analysis and compared to bibliographical data. As a result, a cluster of researchers on the Web becomes a superset of bibliographical one, that is, a community representing a larger research field. We then develop Web link ontology that defines various relationships among researchers to apply metadata annotation to the links among 200 sites. Findings: 1) since intellectual ties analysis separates off sociocognitive ties, it results into a better approximation of bibliographical author cocitation, 2) since sociocognitive ties analysis can utilize various links not only coauthor networks, diverse kinds of social connections can be found, 3) merging intellectual ties and sociocognitive ties analyses detect gatekeepers who bridges research fields on social networks.

1. Introduction

The information distribution on the Web plays more and more important role in the social activities of humans. Although finding community as a set of Web pages that have relevant contents has been often studied [8][13], the research which regards links as representation of social relationship between Web site creators and extract community as a social network from large amount of personal Web sites has been just started to be thrown light on [5][6][10]. Our study copes with this problem.

This paper introduces Web citation analysis to the academic community. Web citation analysis is an application of bibliographical citation analysis to the Web. Since the Web link structures have rather vague semantics compared to bibliographical citations [2][7],

the first challenge is to clarify how precisely Web citation analysis captures ties in academic community.

First, to examine if and how exactly Web links among researcher's Web sites represent the intellectual ties in academic communities, Web author cocitation analysis [16] is applied and compared to bibliographical author cocitation data.

Next, to extract more detailed social relationships from the Web, we propose ontology that defines classes of social relationships represented by links on researchers' sites applying the technology of the semantic Web [1], as the basis necessary for extracting substructures of community by qualified links.

As an example of such studies, FOAF¹ has been proposed. Goldbeck *et al.* [5][6] discuss the applicability of social network analysis to the semantic web for ontological trust specifications, extending FOAF ontology. This paper describes detailed classes of human relationship, and shows the results of analyses on both intellectual and sociocognitive ties.

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.

2. Dataset

2.1. Data Source

Bibliographical citation data are obtained from *CiteSeer*² [9], an online library system that collects freely available articles published on the Web. All articles listed in the 'Computer Science Directory'³ were collected in November 2001. In *CiteSeer*, the URL of the author's Web site provided by *HPSearch*⁴ is accompanied to each article. Using robots, we

¹ <http://www.foaf-project.org/>

² <http://citeseer.nj.nec.com/>

³ <http://citeseer.nj.nec.com/directory.html>

⁴ <http://hpsearch.uni-trier.de/>

Table 1. Data Statistics

Data source	<i>CiteSeer</i> Computer Science Directory
Research domain	Computer Science and Engineering
Date of data sampling	November, 2001
Number of <i>papers</i> gathered	8,811
Number of <i>researchers</i> (single or first authors in the papers)	5,220
Number of <i>researchers who have Web sites</i>	3,878
Number of <i>total Web pages</i> in gathered Web sites	273,404
Number of <i>Web links between researchers</i>	6,263
Number of <i>bibliographical citations between researchers</i>	27,840
Number of <i>coauthorships between researchers</i>	4,852

*Citations (links) are counted as directed edges, while coauthorships undirected.

followed these hyperlinks and collected all Web sites created by first authors. Of the 5,220 researchers listed in 8,811 articles, 3,878 researchers had personal Web sites. These 3,878 researchers are used as the analysis targets. Table 1 describes detailed statistics of the data.

2.2. The Unit of Data

Given the context of Web citation analysis, we did not pay much attention to individual Web pages, only entire Web sites. To be specific, we defined a Web site as a set of URLs that share a common prefix. For example, since following URLs have common prefix “<http://webscience.edu/~smith/>”, they are regarded as parts of one Web site.

- <http://webscience.edu/~smith/cv.html>
- <http://webscience.edu/~smith/course/2002.html>

In addition, multiple Web sites managed by the same researcher for some reason (e.g. both old and new Web sites are maintained or a researcher has multiple affiliations) are treated as a single Web site. The number of links from a site to another is normalized as 0 or 1, even if multiple links exist between these sites. For, in our observation, multiple linking often depends on authoring style (e.g. linking from CV, project, publication list, or them all?) and is not regarded as a good measure of cocitation strength. Thus cocitation count on the Web is calculated as the number of sites that cocite a site pair.

2.2. Data Correction and Author Identification

Paper titles and author names in *CiteSeer* are extracted by automatic text analysis. And the URLs of the researchers’ Web sites are based on the assumption of search engine results. For this reason, *CiteSeer* contains erroneous data. Accordingly, we manually checked and corrected all author names and URLs in November 2001.

In accordance with prior works (such as [11]), we differentiated the author names, by automatically normalizing them to “first name initial + last name,” such as “T.White.” We manually differentiated those who had the same first name initial and the same last name, but had different first names or different middle names. For these cases, we gave authors different names to distinguish them, such as “Alan.Show” / “Anthony.Show,” or “John.F.Smith” / “John.A.Smith”.

3. Discovery of Intellectual Ties by Web Author Cocitation Analysis

3.1. Bibliographical Author Cocitation Analysis

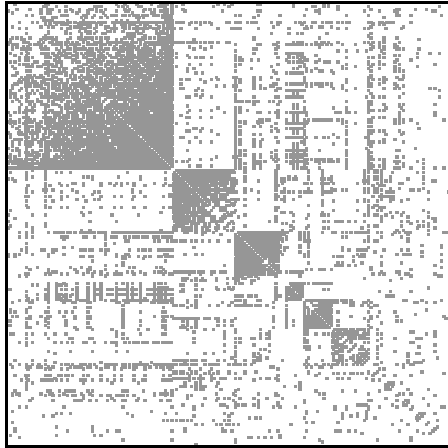
Bibliographical citation analysis [4][17] is a methodology to measure the value of researchers and their literary works in the community. Bibliographical cocitation analysis [15][16] mines related academic fields and hot topics.

In bibliographical cocitation analysis [15], when both article a_1 and a_2 are cited by the same article a_3 , the article pair (a_1, a_2) is called co-cited by a_3 . The number of articles that co-cite (a_1, a_2) is called cocitation count of (a_1, a_2) , and measures the strength of the relationship between articles a_1 and a_2 .

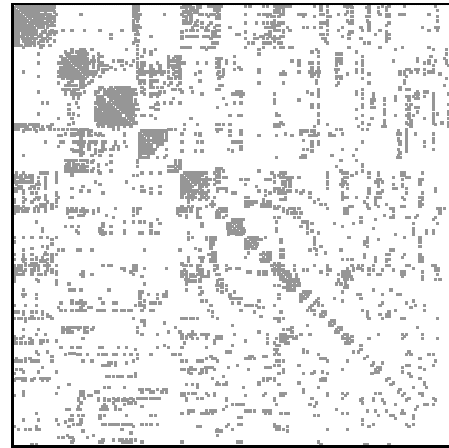
Bibliographical author cocitation analysis [16] is a kind of cocitation analysis. However the unit of bibliographical author cocitation analysis is not each article but a set of articles written by the same author. Here the author of an article indicates its major author, that is, the only or the first author. When article a cites an article written by author A_1 and one by A_2 , the author pair (A_1, A_2) is called co-cited by a . The number of articles that co-cite (A_1, A_2) is a measure of the strength of the relationship between authors A_1 and A_2 .

Author cocitation analysis uses statistical factor analysis to find a set of authors connected by cocitations to each other; it consists of three phases:

- 1. Author Selection:** Select n authors as the analysis targets, namely, author set $A = \{A_1, A_2, \dots, A_n\}$; where n is the number of selected authors. Generally, in order to make the result more reliable, authors cited by many articles are selected, that is, some threshold is determined as the minimum citation



(a) Web Author Cocitation Matrix



(b) Bibliographical Author Cocitation Matrix

Fig. 1. Web and Bibliographical Author Cocitation Matrix Sorted by Cluster and Factor Loading

count and only the authors whose citation count is equal to or more than the threshold are chosen.

2. **Obtaining Cocitation Counts:** Calculate cocitation counts of all author pairs $P = \{(A_i, A_j) \mid A_i, A_j \in A, i \neq j\}$ from the data. The result is represented by cocitation matrix C_{CC} :

$$C_{CC} = (c_{ij}); c_{ij} = c_{ji} = \text{CocitationCount}(A_i, A_j)$$

C_{CC} is the adjacency matrix of the network whose vertex is an author and whose edge means an author pair, weighted by the cocitation count of the pair.

3. **Principal Component Analysis:** Divide the authors into clusters by principal component analysis. The rows of the $n \times n$ cocitation matrix are samples, while columns are variables. Calculate principal components of the variables by varimax rotation; the principal component to which a variable (an author) is most strongly correlated represents the cluster that the author belongs to.

3.2. Web Author Cocitation Analysis

Author cocitation analysis is applied to the Web based on the methodologies by White and others [16]. They chose the authors whose articles were most frequently cited as the targets of analysis. In order to compare the results yielded by bibliographical author cocitation analysis and by Web author cocitation analysis, we calculated the product of citation counts on bibliography and those on the Web, and selected the top 200 researchers according to these scores.

Next, their Web and bibliographical cocitation counts are analyzed and then the researchers are classified into clusters. The identity of the clusters so generated was investigated in two ways: First, correspondence between the Web author cocitation

clusters and the bibliographical ones is examined. Second, research fields represented by the Web and bibliographical clusters are looked up by looking the categories of *CiteSeer* directory that contains any article written by a researcher classified into the cluster.

3.3. Relation between Web and Bibliographical Author Cocitation Clusters

Figure 1 shows the cocitation matrices in which researchers are brought together by cluster. Author A_i is represented by the i th row and the i th column. Authors are sorted first by the cluster number to which they belong, and then by the order of factor loadings. A colored cell indicates a cocited author pair.

In the cocitation matrices researchers who belong to the same cluster fall on the diagonal. They are densely connected to each other by cocitation relationship. 6 clear clusters appear in the matrix generated by Web author cocitation analysis, while 10 smaller clusters appear in the bibliographical one.

To reveal the relation of clusters of Web author cocitation and bibliographical author cocitation, we examined which Web author cocitation cluster the researchers belonged to who also belonged to a specific bibliographical author cocitation cluster. Table 2 shows the clusters that contain 2 or more authors from the 1st to 15th principal component. The values in the table's cells are the percentage of researchers in a bibliographical cluster who are contained in the specific Web cluster. Authors who have positive factor loadings to the principal component construct a class distinct from those who have negative ones. Symbols "+" / "-" on the right side of cluster number represent the sign of factor loadings.

Table 2. Cluster Correspondence between Web and Bibliographical Author Cocitation

		Web Cluster													total	
		1-	2+	3+	4-	5-	6+	8+	8-	10+	13+	15-	other			
Bib Cluster	1+	95	0	0	0	0	0	0	0	0	0	0	0	0	5	20
	1-	0	64	18	0	0	0	0	0	0	0	0	0	18	0	11
	2+	0	50	17	0	0	0	0	0	33	0	0	0	0	0	6
	2-	12	24	0	0	0	35	0	0	0	0	0	0	29	0	17
	3+	100	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	4+	0	0	100	0	0	0	0	0	0	0	0	0	0	0	13
	4-	0	100	0	0	0	0	0	0	0	0	0	0	0	0	6
	5+	62	0	0	23	15	0	0	0	0	0	0	0	0	0	13
	5-	50	0	0	0	0	0	0	0	0	0	0	0	50	0	2
	7+	0	0	0	0	60	20	0	0	0	0	0	0	20	0	5
	7-	0	0	14	0	0	57	14	0	0	0	0	0	14	0	7
	9+	100	0	0	0	0	0	0	0	0	0	0	0	0	0	6
	10-	100	0	0	0	0	0	0	0	0	0	0	0	0	0	5
	11-	75	0	0	0	0	0	0	0	0	25	0	0	0	0	4
	12+	100	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	13+	0	0	0	33	33	0	0	17	0	0	0	0	17	0	6
	14+	20	0	40	20	0	0	0	0	0	0	0	0	20	0	5
15+	50	0	50	0	0	0	0	0	0	0	0	0	0	0	2	
other	35	12	5	3	9	9	2	2	3	2	5	15	15	0	66	
total	75	28	23	8	13	17	2	2	4	2	3	23	200			

From Table 2 following facts are found out. In most bibliographical author cocitation clusters, the researchers in the cluster belong to a particular Web cocitation cluster. For instance, 95% of researchers in bibliographical cluster '1+' are also contained in Web cluster '1-.' It is also found that for 6 bibliographical clusters, 100% of the members fall into one Web cluster. Table 2 confirms that every bibliographical cluster is a subset of one of the Web clusters.

3.4. Community Represented by Author Cocitation Clusters

We next examined what properties the extracted Web and bibliographical clusters have based on article classification from the Computer Science Directory entries in *CiteSeer*. Both Web author cocitation clusters and bibliographical ones show the communities whose characteristics are sufficiently explained by the directory of *CiteSeer*. For example, at the positive end of the first cluster (cluster '1+'), 69% of authors are contained in the "Artificial Intelligence" directory, and 63% in "Machine Learning." Thus the clusters extracted by Web author cocitation analysis well represent particular research areas.

In bibliographical author cocitation analysis, on the other hand, in cluster '1+' 80% of researchers are classified in "Artificial Intelligence," 80% in "Information Retrieval," and 75% in "Machine Learning." The bibliographical author cocitation clusters tend to identify more distinctly, than Web clusters, which researchers contained in each cluster belong to the same research community.

Cocitation with regard to bibliography occurs after a rigorous selection of references based on the contents

of the article. The extracted research area, therefore, represents a smaller specialized domain. On the other hand, the semantics of links on the Web are not as strict as those seen in bibliographical citations. Actually, some Web author cocitations are generated by contents such as "Links to AI Researchers" in our data. That's why Web author cocitation yields more general research fields or "Topics."

4. The Ontology of Links in Academic Community

An ontology, which defines what links mean, is necessary to describe link semantics as metadata. Therefore we developed a prototype of the ontology for link classification according to the creators' social relationships that links indicate. It defines a class hierarchy that classifies links and some relationships such as symmetry or inverse. It is a basis of realizing description of communities connected by links and agents that analyze them.

4.1. The Development Process of the Ontology

The ontology is developed incrementally and simultaneously with attaching metadata to 3,804 actual Web links, which point the most frequently linked and cited 200 researchers' sites.

The process of classifying links based on the ontology, and of modifying the ontology to enable it to classify typical links in actual Web pages, is repeated iteratively. Some classes are found thoroughly, and then their superclass is extracted as the common concept of them. For instance, "coauthorship" is observed in some publication list pages and "coorganizer" in some workshop CFP pages. The two classes of human relationship, which involve working together in the research activity, are important elements in the social network. Therefore superclass "collaborator" is derived from them.

The ontology works as the basis of describing what links mean in the domain of academic Web sites. Described metadata makes it possible to pick up the links that represent some social relationship and to analyze in detail how link structure reflects social structure.

4.2. The Specification of the Link Ontology

Table 3 describes the class hierarchy of the developed ontology together with the number of links that is classified to each class.

For the purpose of analyzing researchers' social relationships, the ontology is designed so as to divide

Table 3. The Link Ontology And the Number of Links Belonging to Each Class

Link Class	Note	#Links
0 Link	Root class	220
1 Intellectual(Referential) Link	Link referring 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	6
1.1.3.1 Journal	Link to journal	7
1.1.3.2 Conference/Workshop	Link to conference or workshop	77
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	1
2.1.3.1 Journal	Journal co-editor	3
2.1.3.2 Conference/Workshop	Conference/workshop co-organizer	181
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	0
2.3.1.1 Journal	Author to editor	4
2.3.1.2 Conference/Workshop	Participant to organizer	53
2.3.2 Participant	Participant	0
2.3.2.1 Journal	Editor to author	1
2.3.2.2 Conference/Workshop	Organizer to participant	43
2.4 Private Acquaintance	Informal relationship	0
2.4.1 Friend	Friendship	7
2.4.2 Family	Kinship	9
Total		3,804

linkage (<0 Link>) firstly into <1 Intellectual Link> and <2 Sociocognitive Link>. The former is the class of linkage that simply refer to knowledge about things or persons, while the latter is that of linkage that shows some social ties between the creators of linked two Web pages. Classification can be judged according to the anchortext of the link, the surrounding text around it, and/or the contents of the page that the link points to.

Links as Reference to Knowledge

Class <1 Intellectual Link> is firstly divided, based on what page it is from. (<1.1 Research Reference> is a linkage from some research-related page.) Further, according to what the link points to, <1.1> has four subclasses. (<1.1.1 People> is a link to a researcher.)

Table 3 clarifies that links that represent intellectual ties are about 70% of all Web links. Also, many links that represent intellectual ties relate to educational activities, such as a pointer to textbook information. This causes Web author co-citation analysis extracting

bigger clusters than bibliographical author cocitation analysis did as we described in Section 3.

Links as Representation of Social Relationship

The subclasses of <2 Sociocognitive Link> are defined according to what social relation lies between sites' authors. For example, <2.1.1.2 Coauthor> observed in some publication pages and so on, and <2.1.1.1 Project Collaborator> seen in project or research group pages, both represent <2.1.1 research collaborator>. (Please refer the table for details.)

Most of these social relationships involve mutual recognition, since there is no uni-directional friendship. But <2.3 Organizer-Participant> may go without any face-to-face communication; it is rather special social relationship. On the other hand, an intellectual link pointing to people is always a uni-directional relationship ("I know him"). This relationship might occur mutually by chance ("I know him and also he

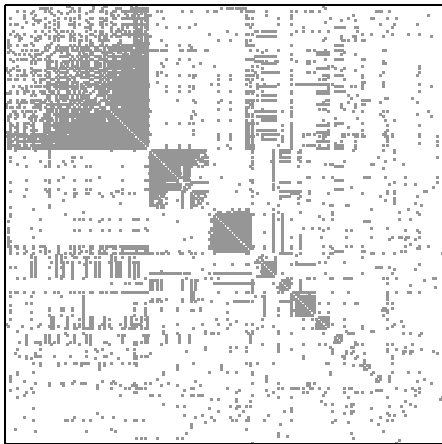


Fig. 2. Web Author Cocitation Matrix from <Intellectual Link> Only

Table 4. Cluster Correspondence between the Analysis Using All Links and That Using <Intellectual Link> Only

		Web Cluster (all links)													total
		1-	2+	3+	4-	5-	6+	8+	8-	10+	13-	15-	other		
Web cluster (1 links)	1-	95	0	0	3	0	2	0	0	0	0	0	0	0	63
	2+	0	92	0	0	0	4	0	4	0	0	0	0	0	26
	2-	0	0	100	0	0	0	0	0	0	0	0	0	0	16
	3+	0	0	75	0	0	25	0	0	0	0	0	0	0	4
	4-	30	0	10	60	0	0	0	0	0	0	0	0	0	10
	5-	0	18	9	0	45	9	18	0	0	0	0	0	0	11
	6-	0	0	0	0	100	0	0	0	0	0	0	0	0	6
	7+	67	0	0	0	0	0	0	0	0	0	17	17	0	6
	8+	0	0	0	0	0	100	0	0	0	0	0	0	0	2
	8-	25	0	0	0	0	0	0	25	0	0	0	0	0	50
	9-	33	0	0	0	0	50	0	0	0	0	0	0	17	6
	11-	0	0	0	0	0	40	0	0	0	20	0	40	0	5
	12+	0	0	0	0	0	0	0	0	0	0	0	100	0	2
	13+	0	0	0	0	0	0	0	0	67	0	0	33	0	3
	14+	100	0	0	0	0	0	0	0	0	0	0	0	0	2
other	12	8	4	0	4	19	0	0	8	4	8	35	0	26	
total	75	28	22	8	12	16	2	2	4	2	3	18	192		

knows me”), but it is essentially different from bi-directional social acquaintanceship.

Here, Table 3 shows that Web links that represent sociocognitive tie are not limited as coauthorships, but are varied. In the described relationships, the rate of coauthorships, which have been analyzed traditionally, is not so large. Social network analysis of Web links with metadata is effective since we can find diverse kinds of human connections. Web citation analysis has a potential to reveal social networks which bibliographical coauthor network analysis [11] cannot find thoroughly.

5. Metadata-based Social Network Analysis of Web Links

5.1. Analysis of Intellectual Ties

Using the ontology, metadata are attached manually to the links in the 3,878 Web sites gathered. (The number of links classified to each class is shown in Table 1.)

Web cocitation analysis is applied to the network. It is already shown that research fields can be extracted by clustering researchers with Web author cocitation analysis. The result is mapped to the network to show whether the network structure reflects academic fields.

Fig. 2 shows Web author cocitation clusters obtained by analyzing intellectual ties (links classified into <1 Intellectual Link>). Table 4 compares the characteristics of clusters in Fig. 2 to original Web author cocitation clusters (Fig. 1 (a)). Here we can understand that cocitation analysis on intellectual ties divides original clusters into detailed ones.⁵

5.2. Analysis of Sociocognitive Ties

Closely related groups and gatekeepers among them are seen in a social network written as a graph. Such phenomena have been examined, for example, coauthor network analysis. With metadata we can pick up sociocognitive ties from Web links and obtain a social network to analyze.

In order to extract relationships with bi-directional recognition from the links among 200 researchers, <2 Sociocognitive Link> except for <2.3 Organizer-Participant> is picked up and the network is calculated. Fig. 3 is the largest connected component of the calculated network. The line type of edges represents the class of relationship <2.1.1 Research Collaborator>, <2.1.2 Teaching Collaborator>, <2.1.3 Academic Committee>, and <2.2 Educational Relationship>.

The network of 200 people is divided into its connected components; then the network of 94 people (47% of the whole) is obtained as the largest. The rest of the components consist of 10 or less nodes.

Then the betweenness of each node in the network is calculated. Betweenness indicates a node’s influential power to control information flow. It is defined as the number of shortest paths of node pairs that pass through the node [3][14]. Table 5 shows the value of betweenness of twenty people with high scores, which are shown as the percentage of them in the number of all shortest paths. Relatively high scores are

⁵ We also tried to extract Web author cocitation clusters by using links classified into <1.1.2 Work>, which can be considered as bibliographical citations. Though we obtained fairly precise clusters, since the number of links is not large enough, we could not find a clear relationship between clusters obtained from <1.1.2 Work> and those from all Web links.

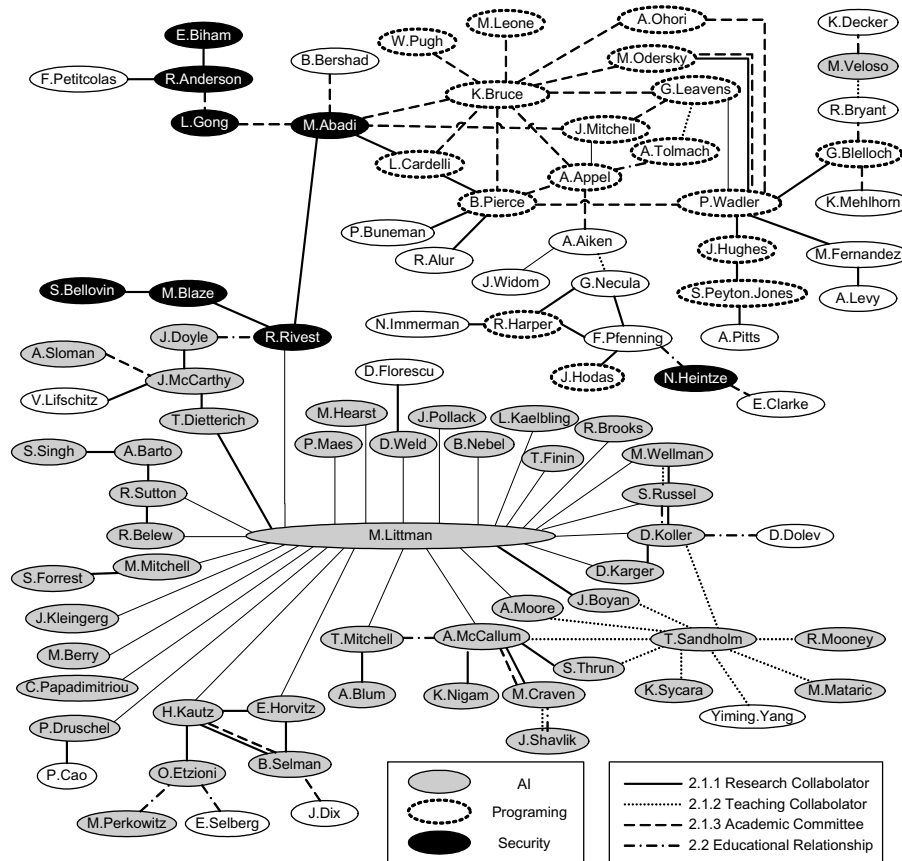


Fig. 3. The Social Network with Edges of Bidirectional Acquaintance Links

concentrated in three people (53 to 70%); this tendency meets with that of typical social network [12].

5.3. Merging Analyses of Intellectual and Sociocognitive Ties

We map the clusters obtained from Web author cocitation analysis onto the social network obtained from sociocognitive ties. Such subnetworks show that research community is constructed by diverse social relationships based on various research and educational activities. For example, the color and the line type of nodes in Fig. 3 represent the major three clusters obtained by Web author cocitation analysis. Each cluster corresponds to artificial intelligence, programming, and security area.

We analyzed what community makes up the network by examining the class of edges. The programming community appears on the upper side of the figure, while the AI community on the lower. And on the left-upper side the security group takes place.

We then examine how the gatekeepers of research community appear on the network by calculate the betweenness of each nodes. We found that the network

Table 5. Betweenness Value of Each Node

Betweenness	Name	Betweenness	Name
71.6%	M.Littman	10.2%	A.McCallum
53.4%	M.Abadi	8.4%	G.Belloch
53.0%	R.Rivest	8.3%	H.Kautz
27.7%	K.Bruce	7.2%	G.Leavens
22.2%	P.Wadler	6.7%	T.Sandholm
18.6%	A.Appel	6.4%	F.Pfenning
16.1%	A.Aiken	6.0%	L.Gong
13.3%	B.Pierce	4.3%	L.Cardelli
12.4%	J.Mitchell	4.3%	J.McCarthy
12.2%	G.Necula	4.3%	O.Etzioni

of academic social relationships contains subgroups, which correspond to research fields and that they are interconnected by some gatekeepers.

In the programming community, for example, R. Rivest and M. Abadi occupy the position that connects the programming and the AI fields. The two nodes also have high betweenness; the critical positions in the network are well represented by betweenness analysis as gatekeepers.

What kind of relationships do the researchers who connect fields have? M. Abadi has connections of coauthorship <2.1.1.2> and workshop coorganizer <2.1.3.2>. In this case the bridge is the researcher whose field is related to both areas. On the other hand, R. Rivest is connected two AI researchers (M. Littman and J. Doyle), but they are not found to be a research collaborator. The link from M. Littman to R. Rivest is listed in “Friends and Colleagues” links and the link from J. Doyle to R. Rivest is “Supervisor” link. There is good possibility that the field boundary is linked by more general human relationship instead of relationship based on clear relevance of research such as “Coauthors.” By analyzing Web links with semantic metadata, we can reveal more precise human relations than bibliographical citation analysis.

6. Conclusion

We applied bibliographical citation analysis to the Web to extract social relationships among peoples and improved the precision of results by using semantic Web techniques. By author cocitation analysis to the 200 authors most cited both on the Web and in bibliographies, it is found:

- The clusters extracted by Web author cocitation analysis reflect ties in academic communities. Web clusters have a tendency to be a superset of bibliographical author cocitation clusters.
- Further, we developed the ontology that defines classes of social relationships represented by links by investigating actual Web pages and finding out human relationship. The ontology, which consists of hierarchy of 34 classes, firstly classifies link into reference to knowledge (intellectual ties) and representation of social relationships (sociocognitive ties). It enables us to embed social relationship among Web page creators in existing Web. The findings, of which we applied the annotating metadata to the links among 200 researchers’ Web sites, were the following three points;
- Analysis using intellectual ties separates off sociocognitive ties and leads a better approximation of bibliographical author cocitation clusters. However, since intellectual ties include not only research references but also educational references, the resulted clusters represent broader research fields than bibliographical author cocitation clusters.
 - Analysis using sociocognitive ties can use various links not only coauthor networks, and it can find diverse kinds of social connections.
 - Application of intellectual ties analysis in combination with sociocognitive ties analysis

finds gatekeepers who bridges research areas are detected on social networks.

Social network analysis with metadata could be put to practical use: e.g. embedding social relationship to the Web links will realize an interesting social networking tool on a world wide scale.

References

- [1] Berners-Lee, T., Hendler, J., Lassila, O., “The Semantic Web”. *Scientific American*, vol.284, no.5, pp.34-43, May 2001.
- [2] Brin, S. and Page, L., “The anatomy of a large scale hypertextual Web search engine”, In *proc. 7th WWW Conference*, 1998.
- [3] Fielder, M., “Algebraic connectivity of graphs”, *Czechoslovak Mathematical Journal*, 23, 298-305, 1973.
- [4] Garfield, E., “Citation Indexing: Its theory and application in Science”, Wiley: New York, 1979.
- [5] Goldbeck, J., Parsia, B., & Hendler, J., “Trust networks on the semantic web”, In *proc. 12th WWW Conference*, 2003.
- [6] Goldbeck, J., & Hendler, J., “Inferring reputation on the semantic web”, In *proc. 13th WWW Conference*, 2004.
- [7] Kleinberg, J. M., “Authoritative sources in a hyperlinked environment”, *Journal of the ACM*, 46(5), 604-632, 1999.
- [8] Larson, R., “Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace”, *American Society for Information Science and Technology*, 1996.
- [9] Lawrence, S., Giles, C. L., Bollacker, K., “Digital libraries and autonomous citation indexing”, *IEEE Computer*, 32(6), 67-71, 1999.
- [10] Matsuo, Y., Tomobe, H. Hashida, K. and Ishizuka, M. “Finding social network for trust calculation”, In *proc. ECAI2004*, 510-514, 2004.
- [11] Newman, M. E. J., “The structure of scientific collaboration networks”, In *proc. National Academy of Sciences of the United States of America*, 98, 404-409, 2001.
- [12] Newman, M. E. J., “Who is the best connected scientist?: a study of scientific coauthorship networks”, *Physical Review*, E64 (016131), 2001.
- [13] Pitkow, J. and Pirolli, P., “Life, death, and lawfulness on the electronic frontier”, In *proc. CHI 97*, 22-27, 1997.
- [14] Schvaneveldt, R. W., Durso, F. T., & Donald, W. D., “Network structures in proximity data”, *The Psychology of Learning and Motivation*, 24, 249-284, 1989.
- [15] Small, H., “Co-citation in the scientific literature: a new measure of the relationship between two documents”, *Journal of the American Society for Information Science*, 265-269, 1973.
- [16] White, H. D. and Griffith, B. C., “Author cocitation: a literature measure of intellectual structure”, *Journal of the American Society for Information Science*, 163-172, 1981.
- [17] White, H. D., and McCain, K. W., “Bibliometrics”, *Annual Review of Information Science Technology*, 24, 119-165, 1989.