Engineering approach to assist the development of National Innovation System

Ichiro Sakata

Todai Policy Alternative Research Institute, The University of Tokyo
Innovation Policy Research Center, Graduate School of Engineering, The University of Tokyo, Japan

Abstract
In these days, the planner of scientific and technological research has to grasp the broader coverage of research, and make decisions on effective investment in promising and emerging technologies especially under circumstances of limited resources. In such a situation, traditional expert-based approach is time-consuming and subjective, and therefore computer-based approach is expected to supplement expert-based approach. Innovation Policy Research Center at the University of Tokyo has developed the computer-based approach to comprehend, science, technology, industry, and market structures and trends for effective and efficient decision making. In this presentation, we demonstrate the effectiveness of our tools by taking energy research, regional cluster, and web community as examples.

I. Introduction

A prominent feature of modern activities is in the creation, dissemination, and application of scientific knowledge. The transition to a knowledge-based society is recently much publicized as a key concept for solving social problems and also as seeds of industrial innovations to drive further economic development. The importance of new knowledge-intensive technologies is widely recognized and has been growing consistently. In this knowledge-based society, scientific activities are playing an increasingly important role and expecting to play more. Therefore, academic articles as the outcome of scientific activities have gained the increasing interest of not only scholars at universities and research institutes but also engineers and policy makers in industry and government domains.

Although investment for research and development (R&D) is an inevitable first step to support scientific activities and to promote technological innovation, the rapid pace of science and technology (S&T) growth and globalization has substantially increased the complexity of S&T

management. It is not a rudimentary task for R&D managers and policy makers to make decisions on effective investment in promising and emerging technologies by selecting them from a pile of plausible candidates especially under circumstances where total budget is constrained.

On the other hand, the process of innovation is becoming fast and complex (Fig. 1). Traditional linear model from basic science, applied research, industrial application, and then social impact including economic impact seems to be not valid. The process has feedback and interaction within the different layer, the speed of it at each layer becomes faster, and some processes are skipped.

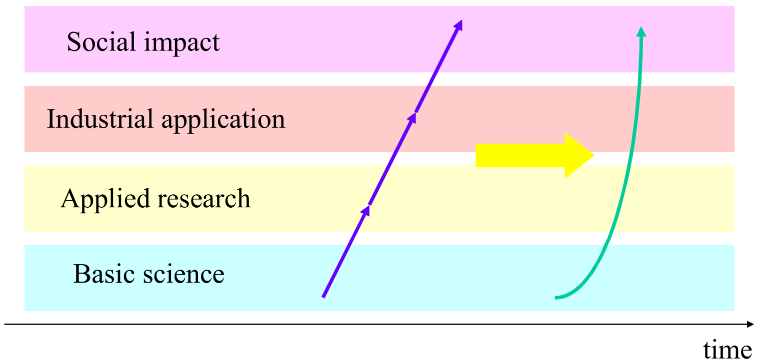

Figure 1. Non-linear innovation model.

Under those circumstances, tools for planners and decision makers engaging in innovation policy is essential and critical to support collecting and analyzing academic, industrial, and market information from a vast of information source such as data bases of scientific papers and patents and the web to synthesize effective R&D policy and plan. Innovation Policy Research Center (IPR-CTR) at the University of Tokyo (UT) has developed the computer-based approach to comprehend, science, technology, industry, and market structures and trends for effective and efficient decision making. In this presentation, we demonstrate the effectiveness of our tools by taking energy research, regional cluster, and web community as examples. In the next section, we illustrated our method depicting academic landscape from existing academic publications whose size is on the order of billion. Case study in energy sector and application of our tool for S&T roadmapping are shown. In section 3, we analyzed industrial structure of regional clusters in Japan. We developed an analyzing tool for industrial network structures, and analyzed selected 18 clusters. Networking mechanism in region of Japanese firms is discussed. In section 4, we visualized community structure of scholars that is usually referred to “invisible college” by on our tool.

2. Academic landscape

Scientific knowledge recorded in academic publications is of primary source for generating and stimulating innovation. For example, a great progress and personalization in computers and internets are driven by innovative breakthrough both in fundamental and applied science such as physics, electronics, materials science, computer science, and information science. At these days, our daily life is in debt with fruitful outcome of scientific activity. Therefore, most of innovation is expected to emerge among a variety of technological breakthrough or a combination of them. Scientific publications constitute the generally accepted, although not always perfect, major output of the scientific activity stimulating technological innovations. So, it is a required and an essential task for R&D managers and policy makers to detect emerging research fronts among a pile of scientific publications. However, such a task becomes highly laborious and difficult as each research domain becomes of specialized and segmented nature of scientific research and also because of the rapid increase of scientific publications [1].

We perform a series of studies to comprehend overall structures of a knowledge domain and detect emerging knowledge domains by analyzing the citation network of scientific publications. For example, Kajikawa et al. investigated the current structure of sustainability science which is emerging but still nascent stage of the development by combined approach of citation network analysis (CNA) and natural language processing (NLP), and discussed the difference between them [2]. Takeda et al. analyzed the detailed structures of emerging knowledge domains by using recursive clustering method of citation networks [3,4]. Kajikawa et al. detected emerging research topics in the energy sector [5-7]. Shibata et al. proposed new approach to predict future core papers in emerging knowledge domain by using topological measures [8]. In addition to such CNA works, Kajikawa performed NLP to extract the essential semantic relationships in scientific papers [9]. Yarime at al. analyzed co-authorship networks in sustainability science, and found the existence of regional collaboration pattern [10-12]. In this section, we pick up a case study in energy sector [5], and illustrate a brief methodology and example of the analyzed result.

2.1. Mehodology

Data

In this paper, we focus on energy research because of its importance. Sustainable energy and renewable energy have been widely accepted as a key concept for our common future. Nowadays, it is often warned that our current energy system is not sustainable. Therefore, we must promote scientific activities and invest on emerging and promising technologies for our future sustainable society.

Our analysis is citation-based approach. We construct networks where nodes are papers and links are citations, and then divide them by clustering. By analyzing the publication years of papers in each cluster, we depict the evolution of the citation clusters. The first step to perform computer-assisted forecasting is to build relevant corpus. We collected citation data of energy-related publications from the Science Citation Index (SCI) compiled by the Institute for Scientific Information (ISI). We used the Web of Science, which is a Web-based user interface of ISI's citation databases. Journal Citation Report (JCR) by ISI was also used to collect energy-related papers.

Our corpus consists of two parts. One is an entire energy research. Another is three sets of emerging subdomains in energy research, i.e., fuel cells, solar cells, and biomass. We collected energy-related papers based of the category of the journal where they were published. We collected papers published in journals categorized in Energy and Fuels of JCR. Bibliographic records of 152,514 papers published in the 68 journals in the category. To construct the corpus for subdomains, we use a set of queries. We use simple queries for fuel cell and solar cell research, i.e., fuel cell* and solar cell*, respectively. The data include 15,600 records for the fuel cell and 16,199 records for the solar cell.

Method

Although there is general agreement to analyze the citation patterns in order to detect emerging research domains, the definition of "citation" differs among the researches. There are three types definition of citation: direct-citation, co-citation, and biblio-citation (bibliographic coupling). Direct-citation is one we normally call as citations. Co-citation is defined as the edge between two documents cited together. Biblio-citation is defined as the edge between two documents cites together. If both paper A and B are cited by C, there exists a co-citation between A and B. And if both D and E cite F, there exists a biblio-citation between D and E. Traditionally, co-citation and biblio-citation have been used as a strategy to analyze citation networks. But, recently, Shibata et al. first showed that direct-citation has a superior performance to detect emerging knowledge domain [13]. Therefore, we should use direct-citation for that purpose.

After obtaining the above data, the citation network is converted into a non-weighted, non-directed network. Finally, the network is divided into clusters using the topological clustering method. The clustering is not fuzzy. A good partition of a network into clusters means there are many within-cluster links and minimal between-cluster links. After clustering the network, we analyzed the characteristics of each cluster by the titles and abstracts of papers that are frequently cited by the other papers in the cluster, as well as the journals in which the papers in the cluster were published. We named each cluster and also listed the keywords for each cluster from the titles and abstracts of the top twenty most cited papers in the cluster. The number of papers in each cluster was plotted along the time line to know the technological trend. Average publication paper in the cluster ($year_{ave}$) was also calculated. Detailed procedures are shown in elsewhere [2,5].

2.2. Results and Discussion

Figure 2 shows structure of citation network of energy research. Publication trends of the top nine clusters in energy research. Brief positions of top 10 clusters in citation network are shown in Fig. 2. Each cluster number is in the order of cluster size (i.e., number of papers in the cluster). Number of papers, average age of cluster, keywords in the cluster are also shown below the cluster name. Fig. 3.

Combustion (Cluster E_1) is the largest cluster among them. The reaction mechanism of a flame in turbulent flow is the main topic discussed in Cluster E_1. As shown in Fig. 1, the number of publications monotonically increases after the 1970s, while some dips are observed after 2000. Coal (Cluster E_2) is the second largest, and the liquefaction and gasification of coal and char are principally studies, but have a low growth rate in publication. The oldest average year of publication of this cluster (1994.1) reflects this fact. Cluster E_3 (Battery) is relatively new and emerging among the top ten clusters, and its average publication year is 1997.9. Cluster E_4 (Petroleum) is as old as Cluster E_2 (Coal). Cluster E_5 (Fuel cell) shows remarkably different growth curves. In sharp contrast to the above clusters, publications drastically increase after around 2000. Cluster E_6 (Wastewater) also shows discontinuous increases after 1991. But this cluster seems to be noisy from the perspective of sustainable energy. In E_6, treatment of wastewater such as textile dye is mainly discussed, while only a small fraction of papers study sustainable energy, e.g., biomass. This inclusion of a noisy cluster is attributed to our selection of corpus, i.e., we simply collected papers from journal categories of ISI to know the global trend of energy research but not form queries. Among the rest of the clusters, Cluster E_9 (Solar cell) is the youngest and has the largest number of publications after 1990. Cluster E_{10} are power system. Other clusters include geology of petroleum and oil-recovery process, but occupy a small fraction of the network. From the growth curve in Fig. 3, Cluster E_2 (Fuel cell) and E_6 (Solar cell) seem to be emerging research domains. In particular, the Fuel cell cluster is the youngest among the top nine clusters.

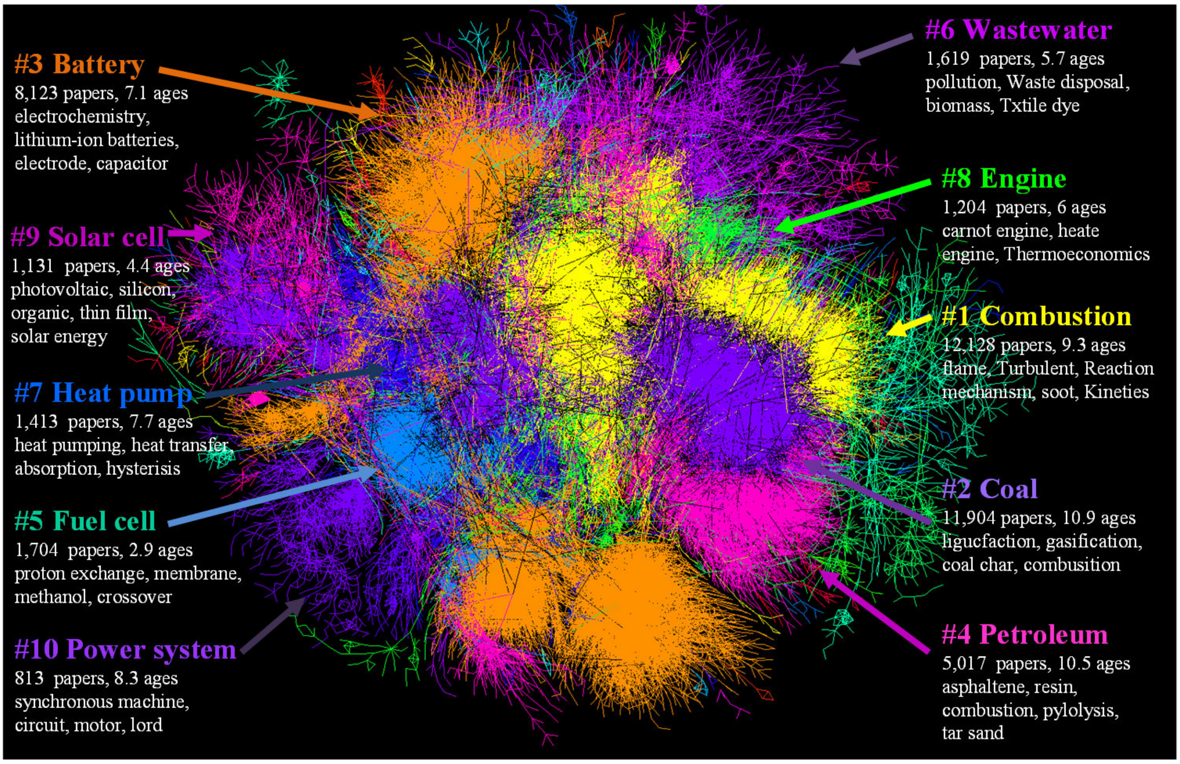

Fig. 2. Academic landscape of energy research.

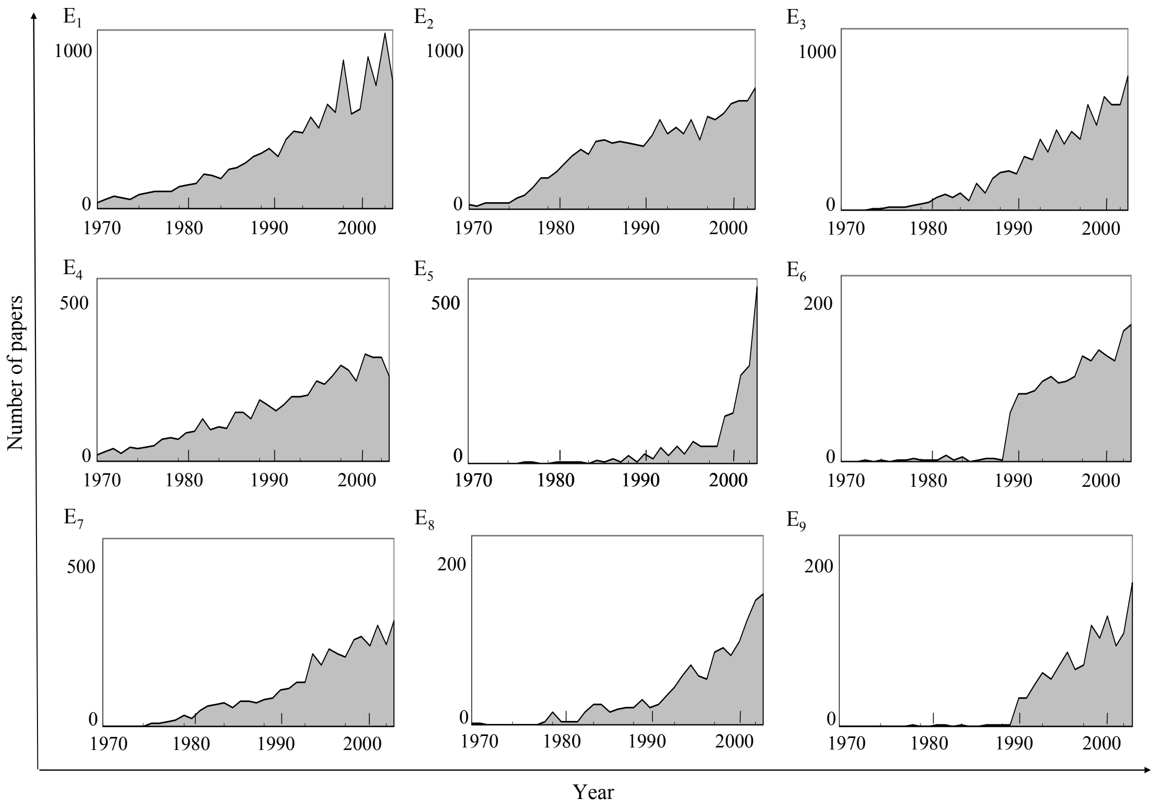

Fig. 3. Publication trend of energy research.

Table 1 shows the characteristics of the top four clusters in fuel cell research. For the fuel cell, the citation network can mainly be divided into four clusters, i.e., PEFC (polymer electrolyte fuel cell), SOFC (solid oxide fuel cell), DMFC (direct methanol fuel cell), and MCFC (molten carbonate fuel cell). The type of electrolyte material seems to be foremost responsible for this clustering result. PAFC (Phosphoric-acid fuel cell) does not appear which might reflect the maturity of this technology. While Cluster F_4 (MCFC) is relatively old, the clusters show similar growth curves and can be regarded as young, emerging domains (Fig. 4). We performed recursive clustering for top 2 clusters.

Table 1. Top four clusters in fuel cell research.

Cluster ID	Cluster name	# papers	$year_{ave}$
F_1	PEFC	3,834	2003.0
F_2	SOFC	3,598	2001.5
F_3	DMFC	3,213	2002.4
F_4	MCFC	1,046	1998.6

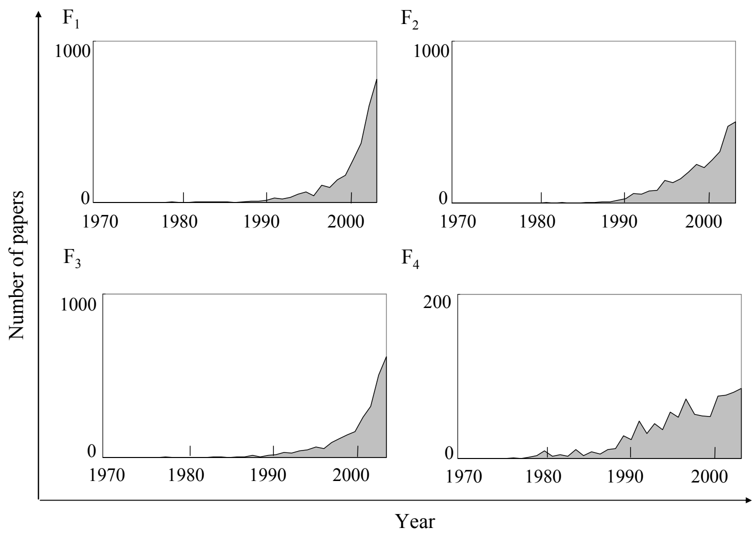

Fig. 4. Publication trend of fuel cell research.

The results are shown in Table 2. F_1 (PEFC) can be mainly divided into three clusters, i.e., Crossover, Modeling, and Proton conductivity. Each cluster can be further divided into sub-clusters. For example, the cluster Crossover can be divided into DMFC and Crossover, Membrane and methanol permeation, Microbial fuel cell, and Micro devices. The application PEFC to micro devices are typically emerging one whose average publication year is 2004.4. We can also detect emerging domains in the subclusters of Modeling and Proton conductivity. For F_2 (SOFC), We found emerging clusters such as Direct oxidation, Interconnects, and Vapor deposition.

Table 2. Subclusters of fuel cell research. Clusters whose average publication year is larger than 2003 are shown in red.

Level 1	Level 2	Level 3
PEFC 3,834; 2003.0	Crossover 1,303; 2003.0	DMFC and crossover 514; 2003.0
		Membrane and methanol permeation 350; 2003.8
		Microbial fuel cells 251; 2000.9
		Micro devices 164; 2004.4
	Modeling 1,275; 2002.7	Catalyst layer of polymer electrolyte 501; 2001.4
		Mathematical model 486; 2003.6
		Measurement of current distribution and water transport 244; 2003.6
	Proton conductivity 1,204; 2003.3	Proton conductivity and water uptake 429; 2002.5
		High-temperature operation 361; 2003.3
		New polymer membrane 356; 2004.2
SOFC 3,598; 2001.5	Electrolytes 1,077; 2001.7	Doped lanthanum gallate electrolyte 300; 2001.7
		Ceria-based electrolytes 267; 2001.5
		Thin electrolytes 250; 2002.1
		YSZ and Al_2O_3 composites 116; 2001.8
		Hydrogen sulfide fuel cell 95; 2001.2
	Anode and modeling SOFC 1,046; 2002.2	Ni-YSZ cermet 301; 2001.4
		Modeling 299; 2002.5
		Direct oxidation 297; 2003.1
	Cathode, electrodes, Interconnects 85; 2001.0	Processing and sintering 424; 1999.4
		(La, Sr)MnO_3 and 265; 2001.5
		Interconnects 181; 2003.2
		Vapor deposition 85; 2003.0

Table 3 shows the characteristics of the top four clusters in solar cell research. The citation network of solar cell research can be divided according to the material used in the cell. This is similar to the clustering result of fuel cell research. The growth curve seems to be classified into three types. One is the smooth increase as shown in S_1 and S_2. The second is the rapid increase after the late 90s as seen in S_3 and S_4. The third is a constant rate of publication (S_5). It is worth noting that there is small variance, for example, the number of publications in S_1 drops from the late 1980s and increases again after the 1990s. Clusters S_3 and S_4 are definitely young and rapidly developing among these clusters Therefore, we regard Clusters S_3 and S_4 as emerging domains and performed further clustering of these clusters (Table 3). Cluster S_5 is the oldest and has not expanded recently. Therefore, we did not perform further clustering.

Table 3. Top five clusters in solar cell research.

Cluster ID	Cluster name	# papers	$year_{ave}$
S_1	Silicon	4,634	1995.2
S_2	Compounds	3,481	1998.6
S_3	Dye-sensitized	2,267	2003.3
S_4	Organics	1,390	2002.3
S_5	GaAs	989	1991.7

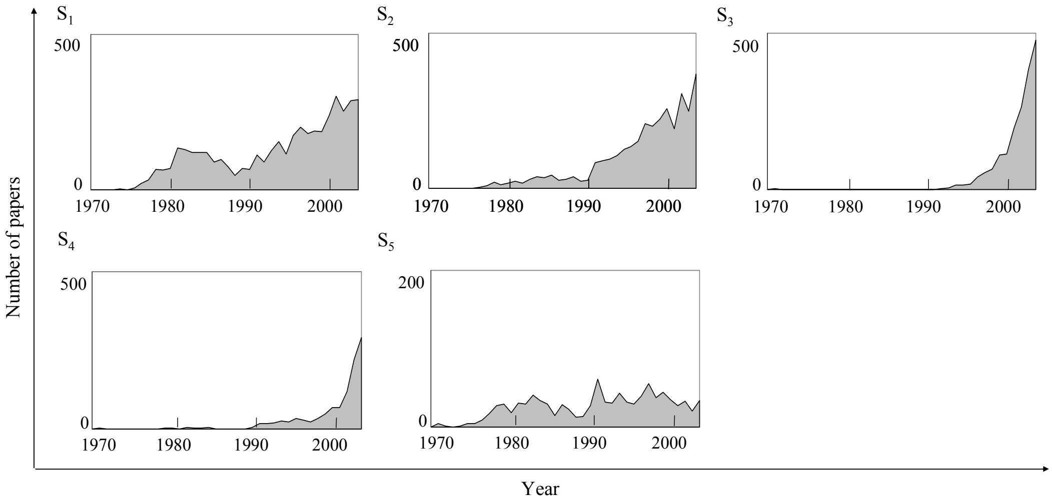

Fig. 5. Publication trend of solar cell research.

Table 3. Subclusters of solar cell research. Clusters whose average publication year is larger than 2003 are shown in red.

Level 1	Level 2	Level 3
Dye-sensitized 2,267; 2003.3	Photosensitizer 737; 2002.4	Charge injection, transport, and recombination 240; 2002.6
		New Ru-based dyes 236; 2002.5
		New organic dyes 185; 2001.7
	Electrolyte 715; 2004.2	Gel electrolyte 237; 2004.0
		Nanostructure control 220; 2004.4
		Solid-state dye-sensitized device 166; 2004.1
	Modeling 498; 2003.4	Photocurrent spectroscopy 172; 2003.4
		Mechanism and modeling 172; 2002.7
		Core-shell nanoparticles 113; 2004.5
	Fabrication 205; 2003.5	
Organics 1,390; 2002.3	Plastic solar cell 448; 2004.6	General 126; 2004.3
		Cell efficiency 108; 2004.9
		Transport and recombination in the membrane 97; 2004.6
		Morphology and performance 79; 2005.1
	Heterojunction 373; 2002.9	Modeling of heterojunction cells 100; 2004.2
		Intermixed structure 80 2002.3
		Doping 74; 2002.2
	Cyanine 328; 1997.0	
	Conjugated polymer 120; 2004.0	

2.3. Application

We compare the above results with the Technology Strategic Roadmap (TSR) was developed at the initiative of New Energy and Industrial Technology Development Organization (NEDO) in Japan, which includes a fuel cell and solar cell roadmap constructed by expert-based approach organized by the. S&T roadmap is a represented figure to be constructed for clarifying the direction of research and development (R&D) and sharing future visions on technologies, and promote interdisciplinary collaborations among different participants both in industry and academia. S&T roadmaps are expected to offer a means of communicating visions, attracting resources from business and government, stimulating investigations, and monitoring progress [14].

NEDO is responsible for R&D project planning and formation, project management. The fuel cell roadmap in TSR was published in 2006 and is divided into three categories, i.e., PEFC, DMFC, and SOFC. These three research domains for the fuel cell match the results of our research (Table 2). Our citation network analysis reveals that PEFC, DMFC, and SOFC are the main clusters in which the number of papers is rapidly increasing. Therefore, the roadmap by NEDO is reasonably to focus on these main emerging domains. As for a solar cell roadmap, NEDO established the Photovoltaics (PV) Roadmap toward 2030 (PV2030) as a long-term strategy for PV R&D in 2004. In PV2030, the target is the crystalline Si solar cell, which has the highest market share of PV, and also the high-efficiency GaAs-based or $CuInSe_2$ (CIS) solar cell. Dye-sensitized cells were the subject of discussion as seed research after 2010. But according to our results, the organic solar cell is another emerging domain, which was missed in TSR. Our analysis successfully can detect the organic solar cell as an emerging research domain. Therefore, we consider that the computer-based approach can be utilized to offer supplemental information to construct a roadmap using an expert-based approach by retaining global data awareness. The citation network approach is a powerful tool to support experts to construct roadmaps in domains where the number and speed of publications is higher than can be handled such as energy research.

3. Industrial structure in regional cluster

In the last decades, there has been a widespread resurgence of interest in the economics of industrial locations, particularly in the issue of regional clusters. In the context of regional cluster, innovation is closely associated with places where relevant resources are easily accessed by firms in close proximity. Some regions have competitive advantages owing to superior innovative capabilities, knowledge, relationships, motivation that distant rivals cannot match, while companies in a global economy can source capital, goods, information and technology from around the world. Silicon Valley and the Route 128 zone of Boston, Cambridge, Baden-Württemberg and "Third Italy" are typical example of such distinguished regions.

Regional clusters can offer more opportunities for innovation than scattered locations, which is typically driven by reduced transaction cost, access to venture capitalists, local labor market pooling, entrepreneurial activity within the region, enhancement of knowledge diffusion, and localized learning. Regional clusters are distinguished from pure agglomerations by their interconnected nature, i.e. clusters are characterized as collaborative networks and concentrations of collaboration and competition, which offer significant opportunities and stimulate economic development. Another characteristic of regional clusters is the diversity of actors contained within; an industrial cluster includes suppliers, consumers, peripheral industries, governments, and supporting institutions such as universities. In sum, the network among actors is the key to understanding the performance of regional clusters.

Networks are especially important for small and medium-sized firms, since they lack their own resources to compete effectively with other firms. To overcome these deficiencies they must either depend on resource transfers from large enterprises or be linked to a community of small firms in which productive resources are jointly procured, developed, and utilized. Dense networks can reinforce trust building. Trusting behavior affects the persistence of interfirm networks and improves the quality of information flows critical to innovation. The connection to market leaders or highly regarded firms that can give a reputation or legitimacy to the young firm. In this way small firms can enjoy many of the advantages possessed by large firms, and consequently offer jobs of comparable quality. Especially for regional clusters consisting of medium and small firms, networking activity and the resulting network structure should play an important role.

Therefore, understanding the network structure in the focused region is an inevitable step to grasping the current status of regional industrial structure and effective policy development. The aim of this section is to investigate network structures in eighteen regional clusters and to discuss the route to enhance regional networking. We examine the interfirm networks of eighteen regional clusters in Japan. Next, we illustrate our research methodology.

3.1. Methodology

The term "network" refers to a set of nodes and the relationships that connect them. A social network can be defined as 'a set of nodes (e.g. persons, organizations) linked by a set of social relationships (e.g. friendship, transfer of funds, overlapping membership). Because regional clusters are distinguished from pure agglomerations by their interconnected nature, the level of our analysis is interfirm network and firm modules. In this work, we regard customer-supplier relationships as links in the network among various relationships between multiple firms, because these relationships are known as the best source of information for Japanese firms. Researchers also increasingly regard interaction within customer-supplier relationships as key to the successful management of innovation, as customer and supplier relationships play a critical role in knowledge development, resource mobilization and co-ordination. The key characteristic of customer-supplier relationships in Japan is the fact that the relationships with customers are more dedicated and long-term than in other countries. In the following, we explained our data and analyzing schema.

Data

We select eighteen clusters as shown in Table 1. We listed firms corresponding with the industrial categories located in each region, using a database provided by NTT. This NTT compiled database includes the addresses and industrial category of the firms registered by themselves. We define business transactions between firms as links. The data by the Teikoku Data Bank (TDB) has up to five suppliers and customers for each firm, meaning each firm can link up to a maximum of ten other firms. Because business transactions include a range of traded volume, this restriction on the number of links enables us to extract not the entire business network in the region but just its essential features. These datasets were collected in the year 2007 for cluster B (Chukyo), D (Hiroshima-Okayama), I (Niigata), K (Kyoto), N (Hamamatsu), and R (Okinawa), and 2005 for the rest of the regions. We integrated these two databases by NTT and TDB into a single dataset on networks consisting of nodes and links. The networks are non-weighted and non-directed. Subsequently, we extract the maximum connected component of each network. The resulting networks have approximately 500 to 9,000 nodes, and 2,000 to 43,000 links for each cluster (Table 1).

Table 4. Basic characteristics of 18 regional clusters

#	Region	Main industrial category	*n*	*K*
A	Osaka	Manufacturing	8,834	43,092
B	Chukyo	Manufacturing	7,914	34,162
C	Kinki	Pharmaceutical & Medical	5,437	25,310
D	Hiroshima-Okayama	Manufacturing	3,553	13,772
E	North-Kyushu	Manufacturing	3,275	13,420
F	Hukuoka	Environment	3,272	14,226
G	Hokkaido	Pharmaceutical & Agriculture	2,038	7,740
H	Nagano	Manufacturing	1,933	10,018
I	Niigata	Manufacturing	1,898	8,426
J	Sapporo	Pharmaceutical & Agriculture	1,871	6,086
K	Kyoto	Manufacturing	1,798	7,362
L	Toyama	Pharmaceutical	1,397	5,364
M	Sapporo	Information Technology	1,113	3,820
N	Hamamatsu	Manufacturing	1,049	4,080
O	Hukuoka	Medical Device	931	2,702
P	Aomori	Agriculture	673	2,164
Q	Yamagata	Manufacturing	625	2,078
R	Okinawa	Food	527	2,594

Method

We investigated the structure of interfirm networks by using density and inter-module coordination. Density is defined as the number of links per node. When evaluating density, we treat the network uniformly. However, interfirm network is often neither uniformly dense nor sparse. The structure is uneven, composed of regions that are more or less filled with relationships. A group of firms extensively sharing partners have dense relationships with certain partner groups and sparse or no relationships with others. Therefore, we also evaluate inter-module coordination, which is defined as the reverse of modularity. Details of analysis are shown in Ref [15,16]. Evaluation of density and inter-module coordination is essential to evaluate network performance. Our further analysis and detailed implications for each region are seen in Ref. [17-22]

3.2. Results

Figure 6 shows the relationships between cluster size, i.e., number of nodes in each region and network structures, i.e., density and inter-module coordination. As shown in Fig. 6, those measures improve as node size on the network (n), in other words, number of firms in each region, increases especially in the region where $n > 4{,}000$. On the other hand, below $n \sim 4{,}000$, there is a marked variance both in APR and IMC. The variance is large especially among regions with small n, and as n increases network structure improves. These results indicate that activities to support regional networking are necessary especially for small clusters. In those clusters, there is a room to improve networking

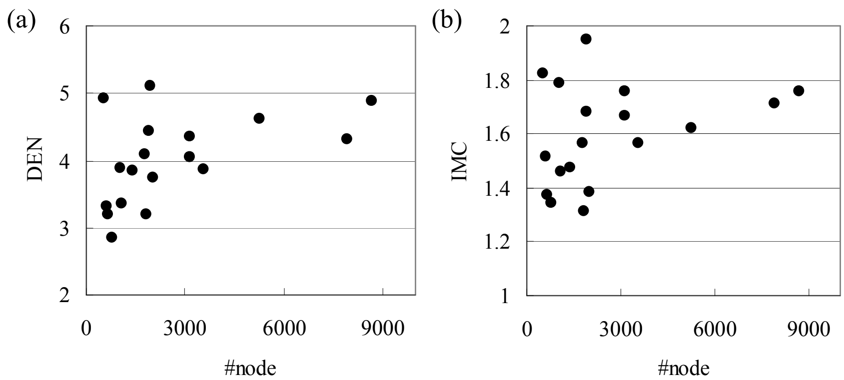

Fig. 6. Effect of cluster size. (a) density (DEN) and (b) inter module coordination (IMC).

But how can we assist regional networking? To answer it, further analysis is performed based on a comparative study. Figure 7 is a rearranged plot of Fig. 6 according to the rank of each region in their size. It is clear that density and inter-module coordination are correlated. And cluster Hamamatsu and Nagano have superior network structures among relatively small clusters. On the other hand, cluster Hiroshima-Okayama and Hokkaido has inferior structure among larger cluster, and Sapporo, Hukuoka, and Aomori have low value of density and inter-module coordination among small clusters. Therefore, we compare the detailed structures of interfirm networks at cluster Hamamatsu and Nagano with those of cluster Hiroshima-Okayama and Sapporo.

In the following, we discuss the role of hub firms in Nagano and Hamamatsu are typical regions having superior network structures, and Hiroshima-Okayama and Sapporo

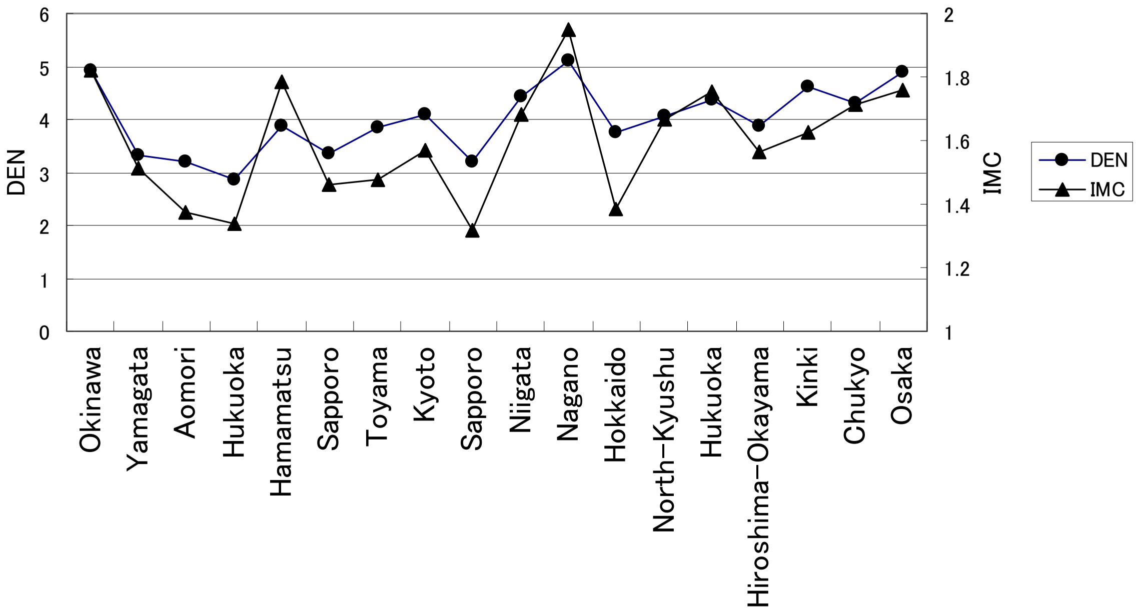

Fig. 7. Network performance of each region.

Figure 8 are visualizations of modular network structures coordinated by spring layout algorithm. We use graph drawing tool, Pajek, to visualize them. The size of node is proportional to the number of firms in each module obtained by topological clustering. The width of lines is proportional to the number of links between two modules. In those figures, hub firms of each module is also shown. When the firm name is surrounded by a rectangle, it means that the headquarter of the firm locate in the region. When the firm name is not surrounded, the headquarter is outside the region.

In Nagano, each module is strongly connected, which apparently contribute to the superior network structures. It can be also seen that most of hub firms have their headquarters at the region. For example, it is well know the Seiko-Epson is the leading firm of the region (Nagano prefecture). Other companies such as Shinko-Denki, Nissei-Jushi, and Tamagawa-Seiki has their original products and services having competitive advantage. Some firms whose headquarters are outside the region such as Fujitsu, Mitsubishi-electronics, Yuasa-Shoji are also seen but are minor components.

Hamamatsu also has a number of firms whose headquarters are in the region. Examples are Yamaha-motor, Suzuki, Yamaha, Hamamatsu-photonics. Honda has a root in Hamamatsu but the current headquarter of Honda is Iwata city that is the neighbor city of Hamamatsu.

In Nagano and Hamamatsu, regional firms locating their headquarters at the focal region are hub firms of the main modules located in central positions in the networks.

On the contrary, in Hiroshima-Okayama and Sapporo, modules which regional firms belong to located at the peripheral positions in the network. For example, in the modules located at the central positions of cluster H, hub firms are JFE-steel, IHI, Mitsubishi-chemical, Mitsubishi-heavy-industry and so on. In these regions, we can see agglomeration of firms not but well-networked regional clusters.

According to these results, we conclude that the variance in network structure among regional clusters is derived from the location of headquarters of hub firms [23]. This hypothesis drawn from our case comparative case study suggested the difficulty to promote industrial policy based on the concept of regional cluster or regional networking especially at the region where there is less headquarters in the region. In these regions, policy to promote the relocation of headquarters from the metropolitan area to the region or the development of regional firms whose headquarters are at the region are critical factor to enhance the regional networking.

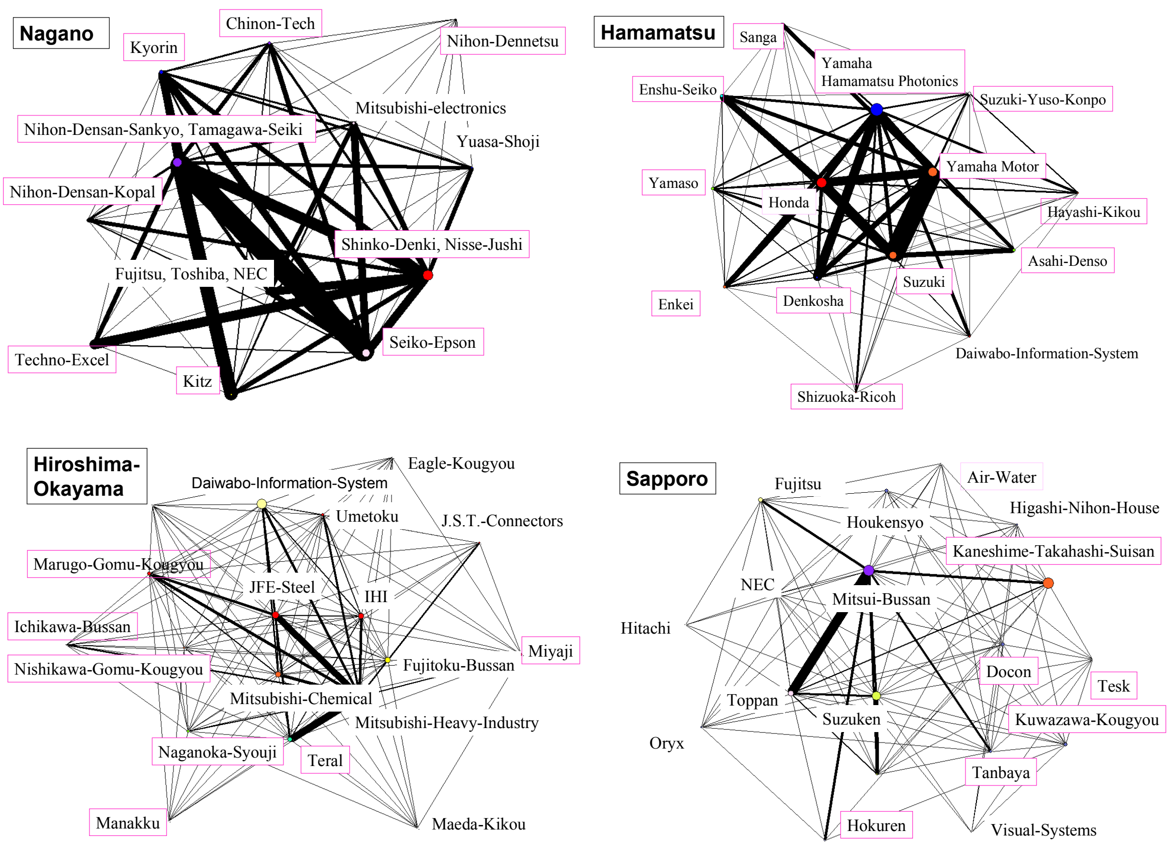

Fig. 8. Modular network structures in the regions

4. Community structure in the Web

In the previous two section, we focus on the relationships between knowledge (scientific papers) and also organization (firms). In this section, we illustrated our tools extracting community structure from the web. Community structure extraction is essential to understand social process because social network can work as a conduit of knowledge and resource and promote coordination as discussed in the previous section. People conduct communications and share information through social relations with others such as friends, family, colleagues, collaborators, and business partners. Our lives are profoundly influenced by social networks without our knowledge of the implications. Direct applications of social networks in information systems include viral marketing through social networks and e-mail filtering based on social networks.

Relations among people are important for many intelligent systems. In my research, a social network, especially a collaboration network of researchers, is extracted automatically from the Web. The algorithm distinguishes the relations into co-authorship, co-participation to the same project, co-affiliation, and co-attendance to the same conferences. We have developed social network mining system called POLYPHONET [24-26]. POLYPHONET is a social network mining system that serves for promoting communication among researchers. It is operated in The Japanese Society for Artificial Intelligence (JSAI) annual conferences and a couple of international conferences. In the following, we introduce the example of POLYPHONET operated at JSAI conference.

4.1 Methodology

A social network is extracted through two steps. First we set nodes, then we add edges. In our approach, nodes in a social network are given. In other words, a list of persons is given beforehand. We collect authors and co-authors who have presented works at past JSAI conferences; we posit them as nodes. Next, edges between nodes are added using a search engine. We estimates the strength of their relation by co-occurrence of their two names in the results of the retrieval. We add an edge between the two corresponding nodes if the strength of relations is greater than a certain threshold. Then, we extracted semantic relationships among them. Through POLYPHONET, we target the relations in a researcher community. Among them, four kinds of relation are picked up because of the feasibility of identifying them and their importance in reality:

- Co-author: co-authors of a technical paper,
- Lab: members of the same laboratory or research institute,
- Proj.: members of the same project or committee, and
- Conf.: participants in the same conference or workshop.

Each edge might have multiple labels. For example, X and Y have both “Co-author” and “Lab.”

relations. By our methods, we can extract semantic relationships with high precision (70-92%) and recall (67-98%), which depend on the kind of the relationship. The details of the procedure are shown in our previous papers [24-26].

4.2 Results

We show the extracted social network among JSAI researchers in Fig. 9, which is used in for navigation of research presentation and retrieval of researchers at the annual conferences. The network includes 266 nodes and 690 edges. The nodes are selected from among 1560 researchers who participated JSAI annual conferences in 2003 and the past 3 years; we exclude those whose hit count is less than 30, and those who do not have any edges recognized (i.e. isolated nodes). A social network of participants is displayed in POLYPHONET to illustrate a community overview. Various types of retrieval are possible on the social network: researchers can be sought by name, affiliation, keyword, and research field; related researchers to a retrieved researcher are listed; and a search for the shortest path between two researchers can be made.

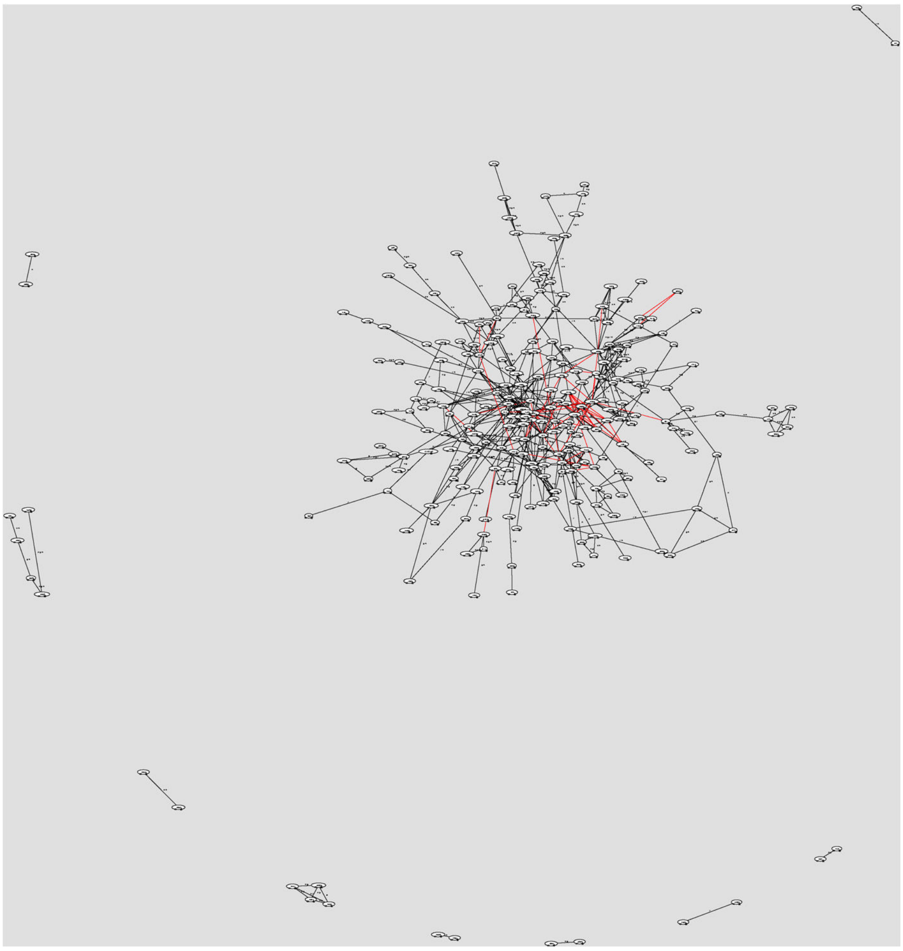

Fig. 9. A social network of JSAI researchers.

Fig. 10 is a portal page that is tailored to an individual user, called my page. The user's presentations, bookmarks of the presentations, and registered acquaintances are shown along with the social network extracted from the Web. Fig. 11 shows the shortest obtained path between two persons on a social network. More than 200 users used the system during each 3-day conference. Comments were almost entirely positive; they enjoyed using the system. POLYPHONET provides an area for future work in which feedback is collected through the system. That information is useful as training data if a user registers another user as an acquaintance. In this way, the system would improve over time in the style of active learning, which can be sought in the future.

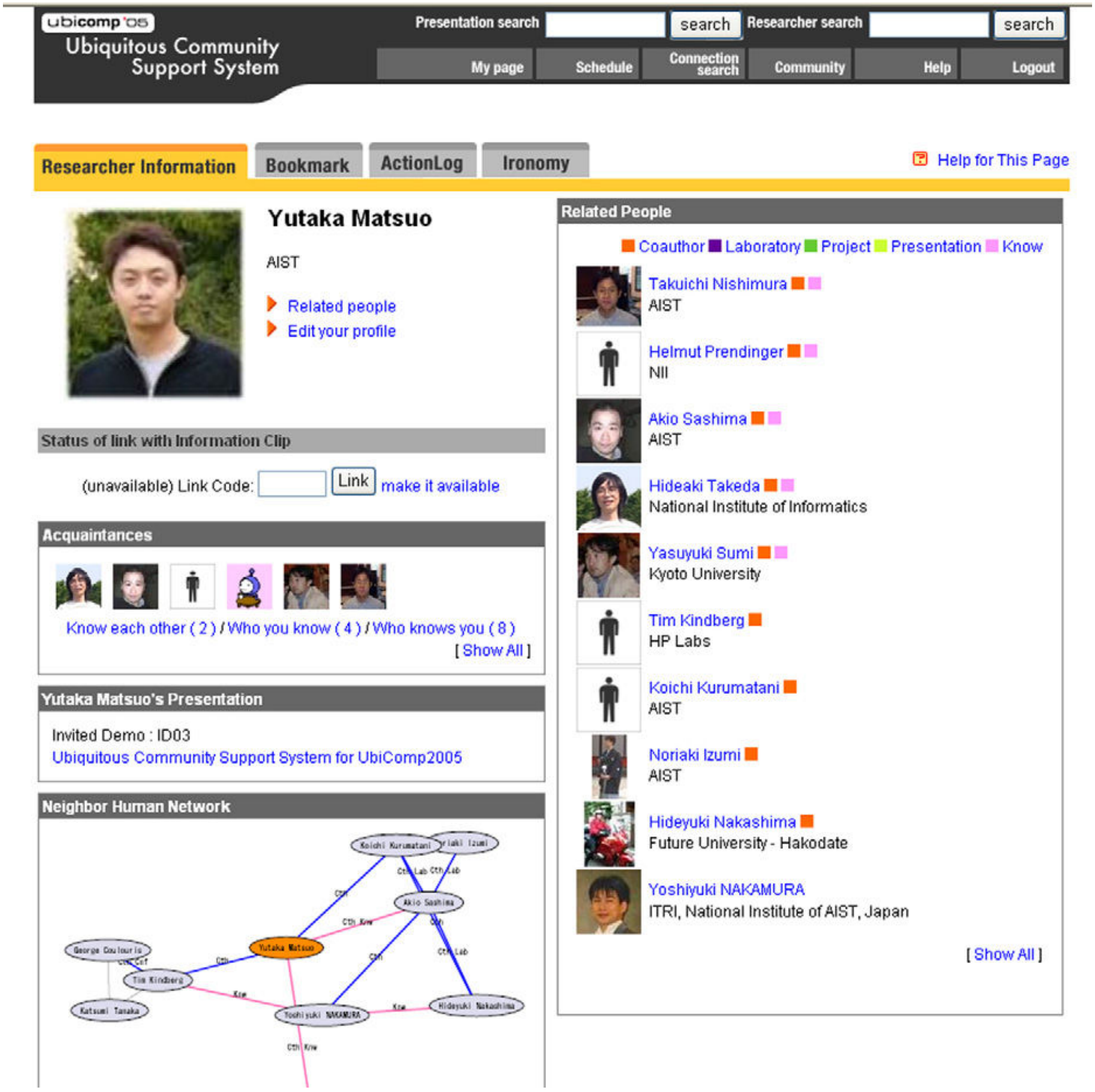

Fig. 10. My page on POLYPHONET.

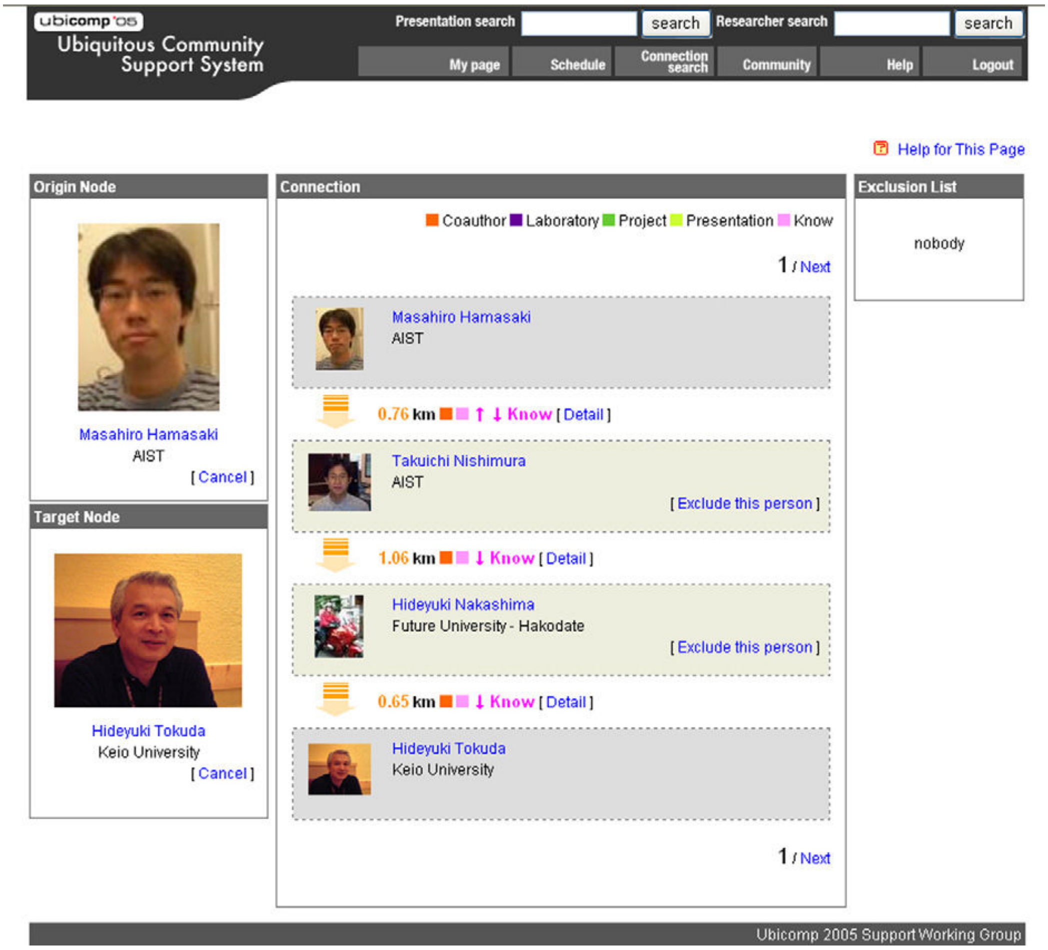

Fig. 11. Shortest path from a person to a person on POLYPHONET.

5. Conclusion

In this presentation, we introduced tools developed by Innovation Policy Research Center at the University of Tokyo to assist the planner of scientific and technological research to grasp the broader coverage of research, and make decisions on effective investment in promising and emerging technologies especially under circumstances of limited resources. We demonstrated the effectiveness of our tools by taking energy research, regional cluster, and web community as examples. These computer-based approaches are expected to supplement expert-based approach.

At first, we illustrated citation network analysis as a technological forecasting methodology and to describe technological trends of energy research. We tracked emerging research domains in energy research by using citation network analysis. Our analysis confirmed that the fuel cell and solar cell are rapidly growing domains in energy research. We further investigated the detailed structure of fuel cell, solar cell, and biomass research. In fuel cell research, the number of publications in PEFC and DMFC are under rapid developing. In solar cell research, dye-sensitized and organic fuel cells are emerging research domains. In biomass research, bio-energy has a sharp increase in publication recently. By using citation network analysis, we can detect and track emerging research domains among a pile of publications efficiently and effectively. Clustering citation network is useful approach to investigate the detailed structures of a research domain. We should use the results as an intellectual basis for constructing a roadmap.

Secondly, we introduced our work analyzing the interfirm network structure of regional clusters. There is an increasing interest in interfirm networks which is expected to work as a source of innovation by circulating resources and knowledge within the network. We analyzed network structures of eighteen regional clusters by average path length ratio and inter-module coordination. We found that these network properties among regional clusters show marked differences. These differences seem to be controlled by the number of firms in the region and the location of the headquarters of hub firms. Because the variance in network structure is large in relatively small clusters, policy for networking can work effectively for small clusters. To develop dense networks, rearing regional firm or attracting headquarters from the other regions especially Tokyo and incubating regional firm as hubs for the region are necessary.

And finally, we extracted social networks from the web. In the above two analysis, i.e., academic landscape and regional cluster, data for network analysis and semantic relationships are given. But in this final section, we extracted such network data from numerous unstructured texts in the Web. We also showed the example we prepared for JSAI conference. POLYPHONET can successfully navigate the researchers participated in the conference.

Merging the vast amount of information scattered in academic database, industrial database, and the Web and producing higher-level information might contribute to many knowledge-based systems We intend to apply our approach in the future to extract much structural knowledge for the development of effective innovation policy in a efficient way.

References

[1] Y. Kajikawa, K. Abe, and S. Noda, Filling the gap between researchers studying different materials, different methods: A proposal of structured keywords, Journal of Information Science 32 (2006) 511-524.

[2] Y. Kajikawa, J. Ohno, Y. Takeda, K. Matsushima, and H. Komiyama, Creating an academic landscape of sustainability science: an analysis of the citation network, Sustainability Science 2 (2007) 221-231.

[3] Y. Takeda, S. Mae, Y. Kajikawa, and K. Matsushima, Nanobiotechnology as an emerging

research domain from nanotechnology: A bibliometric approach, Scientometrics in press.
[4] Y. Takeda and Y. Kajikawa, A bibliometric approach on optics to detect emerging research domain and intellectual base, Scientometrics in press.
[5] Y. Kajikawa, J. Yoshikawa, Y. Takeda, and K. Matsushima, Tracking emerging technologies in energy research: toward a roadmap for sustainable energy, Technological Forecasting and Social Change in press.
[6] Y. Kajikawa and Y. Takeda, Structure of research on biomass and bio-fuels: A citation-based approach, Technological Forecasting and Social Change in press.
[7] Y. Kajikawa, Y. Takeda, and K. Matsushima, Computer-Assisted Roadmapping: A Case Study in Energy Research, Portland International Conference on Management Engineering and Technology 2008 (PICMET'08) in Cape Town, South Africa, (July 27 - 31, 2008)
[8] N. Shibata, Y. Kajikawa, and K. Matsushima, Topological analysis of citation networks to discover the future core papers, Journal of the American Society for Information Science and Technology 58 (2007) 872-882.
[9] Y. Kajikawa, Y. Sugiyama, H. Mima, and K. Matsushima, Causal knowledge extraction by natural language processing in material science: a case study in chemical vapor deposition, Data Science Journal 5 (2006) 108-118.
[10] M. Yarime, Y. Takeda, and Y. Kajikawa, Patterns of collaboration in emerging fields of trans-disciplinary science: the case of sustainability science, 25th DRUID Celebration Conference 2008 on Entrepreneurship and Innovation: Organizations, Institutions, Systems and Regions, Copenhagen, Denmark, June 17-20 (2008).
[11] M. Yarime, Y. Takeda, and Y. Kajikawa, Patterns of collaboration in emerging fields of trans-disciplinary science: the case of sustainability science, Portland International Conference on Management Engineering and Technology 2008 (PICMET'08) in Cape Town, South Africa, (July 27 - 31, 2008).
[12] M. Yarime, Y. Takeda, and Y. Kajikawa, Towards Institutional Analysis of Sustainability Science: An Examination of the Patterns of Research Collaboration, 14th Annual International Sustainable Development Research Conference, New Delhi, India, (September 21-23, 2008).
[13] N. Shibata, Y. Kajikawa, Y. Takeda, and K. Matsushima, Comparative study on methods of detecting research fronts using different types of citation, Journal of the American Society for Information Science and Technology, in press.
[14] Y. Kajikawa, O. Usui, K. Hakata, Y. Yasunaga, and K. Matsushima, Structure of knowledge in science and technology roadmaps, Technological Forecasting and Social Change 75 (2008) 1-11.
[15] Y. Kajikawa, Y. Takeda, I. Sakata, and K. Matsushima, Structure of Interfirm networks in regional clusters, Portland International Conference on Management Engineering and Technology 2008 (PICMET'08) in Cape Town, South Africa, (July 27 - 31, 2008).

[16] Y. Kajikawa, Y. Takeda, I. Sakata, and K. Matsushima, Interfirm networks of regional clusters in Japan, IEEE International Technology Management Conference (TMC-Japan 2008) Japan, Tokyo, (12-13 October, 2008).

[17] I. Sakata, Y. Kajikawa, Y. Takeda, M. Hashimoto, K. Matsushima, An Analysis of industrial networks in a regional cluster: A case study of the food and tourist industries at Okinawa prefecture, Development Engineering 14 (2008) 23-36.

[18] M. Hashimoto, Y. Kajikawa, Y. Takeda, N. Shibata, I. Sakata, K. Matsushima, A study on the Roles and Functions for the Japanese Research Universities in Regional Innovation Networks, Nihon Chizai Gakkaishi 5 (2008) 27-51.

[19] I. Sakata, Y. Kajikawa, Y. Takeda, M. Hashimoto, N. Shibata, K. Matsushima, Network dynamics in the twelve regional clusters, RIETI Discussion Paper Series (2007) 07-J -023.

[20] I. Sakata, Y. Kajikawa, Y. Takeda, N. Shibata, M. Hashimoto, K. Matsushima, Structural Analysis of the Hokkaido Agri-Bio Network, Development Engineering 13 (2007) 31-43.

[21] I. Sakata, Y. Kajikawa, Y. Takeda, N. Shibata, M. Hashimoto, K. Matsushima, Regional knowledge networks of corporations, universities and public institutions, RIETI Discussion Paper Series (2006) 06-J-055.

[22] I. Sakata, N. Shibata, T. Kojima, Y. Kajikawa, and K. Matsushima, Optimal structure of social network for the growth of regional economy: comparative analysis of four regional clusters based on small-world networks theory, Hitotsubashi Business Review 53 (2005) 182-195.

[23] Y. Takeda, Y. Kajikawa, I. Sakata, and K. Matsushima, An analysis of geographical agglomeration and modularized industrial networks in a regional cluster: A case study at Yamagata prefecture in Japan, Technovation 28 (2008) 531-539.

[24] Y. Matsuo, M. Hamasaki, Y. Nakamura, T. Nishimura, K. Hasida, H. Takeda, J. Mori, D. Bollegara, and M. Ishizuka, Spinning Multiple Social Network for Semantic Web, Proc. 21st National Conference on Artificial Intelligence (AAAI-06) , (2006)

[25] Y. Matsuo, J. Mori, M. Hamasaki, K. Ishida, T. Nishimura, H.Takeda, K. Hasida, and M. Ishizuka, POLYPHONET: An Advanced Social Network Extraction System, Proc. 15th International World Wide Web Conference (WWW2006), (2006).

[26] Y. Matsuo, J. Mori, M. Hamasaki, T. Nishimura, H.Takeda, K. Hasida, and M. Ishizuka, POLYPHONET: An advanced social network extraction system from the Web, *Journal of Web Semantics* **5** (2007) 262–278.