Jointly published by Akadémiai Kiadó, Budapest
and Springer, Dordrecht

Scientometrics, Vol. 78, No. 3 (2009) 543–558
DOI: 10.1007/s11192-007-2012-5

Optics: a bibliometric approach to detect emerging research domains and intellectual bases

YOSHIYUKI TAKEDA, YUYA KAJIKAWA

Institute of Engineering Innovation, School of Engineering, University of Tokyo, 2-11-16 Yayoi,
Bunkyo-ku, Tokyo 113-8656, Japan

Optics is an important research domain both for its scientific interest and industrial applications. In this paper, we constructed a citation network of papers and performed topological clustering method to investigate the structure of research and to detect emerging research domains in optics. We found that optics consists of main five subclusters, optical communication, quantum optics, optical data processing, optical analysis and lasers. Then, we further investigated the detailed sub-cluster structures in it. By doing so, we detected some emerging research domains such as nonlinearity in photonic crystal fiber, broad band parametric amplifier, and in-vivo imaging techniques. We also discuss the distinction between research front and intellectual base in optics.

Introduction

A prominent feature of modern activities is the creation, dissemination, and application of scientific knowledge. 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 these scientific activities. While recently the transition to a knowledge-based society has been much publicized as a key concept for further economic development in the 21st century, the 20th century was already based on scientific knowledge. Electronics are typical industrial category based on scientific knowledge.

Received May 27, 2008

Address for correspondence:
YOSHIYUKI TAKEDA
E-mail: takeda@sogo.t.u-tokyo.ac.jp

0138–9130/US $ 20.00
Copyright © 2008 Akadémiai Kiadó, Budapest
All rights reserved

Due to this important industrial-academic relationships, the amount of scientific knowledge measured by the number of academic publications has been rapidly increasing since this century reflecting high capital inputs, which was first mentioned by de Solla Price [1]. There are a number of reports and also many opportunities to create, disseminate, and apply knowledge to real world in an active research area. But the speed and scope of development in such an area make it critical for researchers, engineers, and policy makers to notice information published across different research domains and different institutions.

In this situation, bibliometrics is expected to work as a powerful tool to overview scientific activities in the manner that individuals cannot handle. Progress in computational speed enables us to treat huge amount of data by bibliometric methods. There are various motivations to conduct bibliometric works; to evaluate research output [2–4], to grasp overall structure of research [5–8], and to detect emerging research domains [7–9]. For example, Boyack et al. visualized academic landscape of applied physics especially focusing on microsystems [7]. Small explored the possibility of using co-citation clusters over three time periods to track the emergence and growth of research areas, and predict their near-term changes [9]. In these studies, publications and patents provide the primary raw material for building and developing an R&D indicator. Bibliometric information, i.e., scientific publications, patents and citations to these publications and patents, constitute an adequate information source for the mapping of fields or subfields of scientific and technological enquiry as well as a means of assessing the performance of the major actors in those fields and subfields. Scientific publications and patents constitute a generally accepted, though not always perfect, output indicator of scientific and technological activity [10].

The aim of this study is to investigate academic landscape of optics research and to detect emerging research domains there. We focus on academic publications and analyze the structure of citation networks, because, in this technological domain, R & D expenditure generally precedes business investments and sales and profit earnings [11], and therefore investigation on research trend in science may lead to the findings in business opportunities. In the previous works, bibliometric studies on semiconductor and related industries were already published. In these works, statistical measures such as the number of publications in journals [12, 13], by authors [12, 14, 15], institutions [12, 16], countries [17, 18], and paper impacts [19] were calculated and reported.

While the number of publications and their importance in semiconductor industry are still high, optics has raised their importance in these days. Figure 1 is an example showing such a trend. It shows the value of production in both electronics and optoelectronics industries in Japan. The value for electronics is decreasing with fluctuations, while that for optoelectronics is increasing. But bibliometric study on optics is scarce, except for the report where Trimble investigated citation frequencies of papers studying large optical telescopes [20]. Our primary focus is to investigate the structure of research in optics by citation network approach. By this approach, we also detect emerging subdomains in optics.

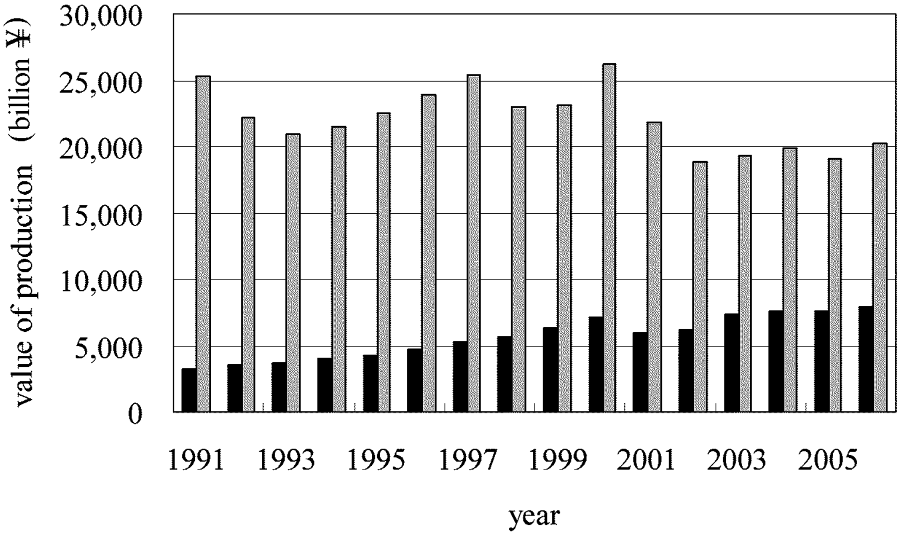

Figure 1. Value of production in electronics and optoelectronics. Data on electronics and optoelectronics are based on Japan Electronics and Information Technology Industries Association and Optoelectronic Industry and Technology Development Association (OITDA) in Japan, respectively.
Data on optoelectronics at 2006 is prospective

Data and methods

Data

We collected citation data of optics-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. We collected citation data by two manners. One is journal based, and another is topic based approach. For the journal based approach, 56 journals classified by the SCI under the category of OPTICS at 2006 were used as the query. For the topic-based approach, we used Ei Compendex Thesaurus. We collected narrower or related terms of optics in the thesaurus, and eliminated general terms such as focusing and light from them, because these have other meanings. As a result, the following set of terms were used as the query; acoustooptical devices, electric lamps, electron lenses, holography, light modulation, light modulators, mirrors, optical communication, optical design, optical devices, optical engineering, optical films, optical instruments, optical materials, optical properties, optical resolving power, optical testing, optical transfer function, optical variables control, optical variables measurement, photography, adaptive optics, aspherics, atmospheric optics, beam propagation method, chromogenics, diffractive optics, fiber optics, fourier optics, geometrical optics, gradient index optics, industrial optics, integrated optics, laser optics, microoptics, nonlinear optics, optical constants, optical losses, particle optics, physical optics, quantum optics, space optics, statistical optics, and X ray optics.

By the above manner, bibliographic records of 281,404 papers were collected. Top 25 journals in the number of publications are shown in Table 1. The data includes all ISI records, i.e. articles, letters, reviews, editorials, meeting abstracts, and so on. Some of these papers may be not relevant to optics research even though they are published as optics-related research, because we collected data simply by journal titles as a clue.

Table 1. Journal characteristics in and out of the network

		Whithin network			Out of network		
Rank	Journal name	#papers	year_{ave}	TC_{all}	#papers	year_{ave}	TC_{all}
1	*Phys Rev A*	42342	1992.2	21.5	4614	1989.5	12.9
2	*Appl Opt*	30524	1988.9	12.4	4882	1977.8	3.1
3	*Opt Commun*	17906	1993.5	9.2	1423	1985.0	3.2
4	*Opt Lett*	14760	1996.3	17.4	501	1991.3	3.2
5	*J Phys-B-At Mol Opt Phys*	13464	1991.6	15.4	4514	1985.4	13.8
6	*Ieee Photonic Tech L*	9647	1999.3	8.3	334	1998.6	2.2
7	*Opt Eng*	7517	1995.3	6.6	2759	1992.0	1.4
8	*J Lightwave Technol*	6662	1996.7	11.4	286	1993.4	1.5
9	*J Opt Soc Am B-Opt Physics*	6411	1995.8	15.8	1244	1990.3	9.2
10	*Opt Express*	5231	2004.6	5.4	211	2003.8	1.1
11	*J Opt Soc Am A-Opt Image Sci*	4719	1996.2	12.7	4022	1990.0	9.0
12	*J Lumin*	3978	1994.8	10.5	3517	1993.3	3.8
13	*J Mod Opt*	3848	1997.3	7.3	306	1999.2	1.0
14	*Appl Phys*	3522	2001.3	8.0	304	2001.0	2.8
15	*Optik*	3266	1986.0	5.8	1690	1972.2	2.6
16	*Opt Quantum Elect*	2339	1992.6	6.4	366	1991.3	1.7
17	*Opt Spectrosc*	1989	2002.1	1.6	1021	2001.7	0.7
18	*Laser Phys*	1975	2001.1	2.6	720	1999.9	0.8
19	*Ieee J Sel Top Quant*	1796	2000.9	11.2	202	2000.6	4.5
20	*Opt Mater*	1615	2002.3	5.6	768	2002.4	2.8
21	*Opt Laser Technol*	1509	1998.3	3.1	1121	1994.7	0.5
22	*Microelectron Eng*	1480	2001.3	5.1	4474	1999.6	2.5
23	*J Opt A-Pure Appl Op*	1338	2003.3	3.1	146	2002.9	1.4
24	*Opt Laser Eng*	1271	1998.6	3.7	357	1996.8	1.0
25	*Microwave Opt Technol Lett*	1259	2002.0	1.4	6799	2000.1	1.8
All journal		203203	1997.3	8.5	78197	1994.1	3.5

Therefore, we focused on the maximum connected component. The retrieved data were converted into a non-weighted, non-directed network. The obtained network currently has 203,203 papers (72.21% of the retrieved data). In other words, we regarded papers not citing other papers in the component as digressional.

In Table 1, we also show the number of publications (#papers), average publication year (year$_{\text{ave}}$), and times cited registered by ISI (TC$_{\text{all}}$) are shown for top 25 journals in #papers. Both papers within the network and those out of the network are counted and shown. As shown in the Table 1, papers not included in the network are usually old and less cited, and therefore, we consider that elimination of these papers has little influence on the results analyzed in the following.

Method

Subsequently, the network was divided into clusters using the topological clustering method [21, 22]. Traditionally, co-citation has been used to analyze a citation network. However, because co-citation is accompanied by a time lag to create a link, and analysis of intercitation is more relevant in the similarity of pairs of documents than co-citation [23], we used intercitation as a link. The clustering algorithm is based on modularity Q, which is defined as follows [21, 22]:

$$Q = \sum_{s=1}^{N_m} \left[\frac{l_s}{l} - \left(\frac{d_s}{2l} \right)^2 \right] \tag{1}$$

where N_m is the number of clusters, l_s is the number of links between nodes in cluster s, and d_s is the sum of the degrees of the nodes in cluster s. In other words, Q is the fraction of links that fall within clusters, minus the expected value of the same quantity if the links fall at random without regard for the clustered structure. Since a high value of Q represents a good division, we stopped joining when ΔQ became minus. 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 heuristically characterized each cluster by the titles and abstracts of papers that are frequently cited by the other papers in the same cluster. It does not mean that all papers in the cluster study the same topics as covered in these frequently cited papers. In fact, each paper studies its own topics, and each paper has its own unique focus. However, as a first approach, it is reasonable to treat these inter-cited papers as a cluster to investigate the brief structure of a research domain and to consider the frequently cited papers in the cluster as representative of the same. But it is not a trivial task for us to characterize the unique content of the cluster, especially when the cluster covers a variety of topics. Therefore, our strategy to characterize and name the cluster is a bottom-up process. Instead giving a name of the cluster at first, we determine the subclusters names of it, and then name the cluster by aggregating such subclusters names. The clustered network is visualized by using a large graph layout (LGL) [24], which is based on a spring layout algorithm where links play the role of spring connecting nodes. Thanks to such layout, papers that cite each other and form a group can be located in closer proximity.

Results

The citation network of optics can be divided into 825 clusters by topological clustering method, where the number of nodes in each cluster varies from 2 (the smallest clusters) to 50,725 (the biggest cluster, #1). Papers in each cluster are strongly coupled with the within-cluster citations. Cluster size, i.e., the number of nodes in each cluster, steeply decreases until the 4th cluster, and after the 5th cluster they become negligible. Therefore, in the following, we focus on the top 5 clusters. They cover almost 90% of papers in the network.

Table 2 summarizes the top 5 clusters and their subclusters. We call these subclusters as 2nd level clusters. 2nd level clusters were obtained by dividing 1st level clusters by topological clustering method again. In Table 2, 2nd level clusters with their sizes larger than 1,000 nodes are shown. The largest 1st level cluster, cluster #1, is on optical communication, whose main subclusters are waveguide (cluster #11), converter (cluster #12), generation (cluster #13), modulators and imaging (cluster #14). In optical communication system, light generation, transmittance in wave guides, conversion and modulators are the key steps and components, and it is also used in imaging as the key application. Cluster #1 is the youngest among top 5 clusters. The second largest cluster is quantum optics (cluster #2), where electric structures and exited states (cluster #21), squeezed light and quantum optics (cluster #22), dynamics in strong laser fields (cluster #23), electron scattering (cluster #24), quantum computation and mechanics (cluster #25), and polarization and instability of lasers (cluster #26) are included as subclusters. Cluster #25 is typically young. Cluster #2 studies fundamental and theoritical aspects of optics. Cluster # 3 is optical data processing. Its subclusters are image processing (cluster #31), optical computing (cluster #32), laser intensity direction and ranging (cluster #33), holographic-interferometry (cluster #34), and image reconstruction (cluster #35). While cluster #1 has main emphasis on hardware, cluster #3 focuses on software. Cluster #4 is optical analysis, and includes various clusters such as optical microscopy and analysis (cluster #41), ionization (cluster #42), phase transition and fractal growth (cluster #43), optical spectroscopy (cluster #44), interaction with materials (cluster #45), photodiodes (cluster #46), and quantum mechanics (cluster #47). Cluster #4 is the oldest among the top 4 clusters which probably reflects their fundamental and common characteristics over optics research. In cluster #4, fabrication process of various low dimensional structures and their interaction with light are also investigated in cluster #43 and #44. The topics in cluster 4 seem to focus on analytical science and materials science. Reflecting the variety of such structures, the clusters splits into relatively large number of subclusters considering the size of the cluster. Cluster #5 is dedicated to laser research, and it includes semiconductor lasers (cluster #41), pumping and electron state transition (cluster #42), and dye lasers (cluster #43) as subclusters.

Table 2. Taxonomic structure of optics. 1st level and 2nd level clusters are shown.
At each level, we show cluster information according to the following format: cluster name (number of papers; average publication year).
The publication year was expressed by the decadal system

Cluster name (#papers, year$_{ave}$)	Subcluster name (#papers, year$_{ave}$) >1000 nodes
#1 optical communication (50,725; 1998.8)	#11 waveguide (15,053;1998.5) #12 converter (13,288; 1999.6) #13 generation (12,153; 1998.0) #14 modulators and imaging (8,088; 1998.9)
#2 quantum optics (44,855; 1994.0)	#21 electric structures and exited states (14,254; 1991.1) #22 squeezed light and quantum optics (10,952; 1996.5) #23 dynamics in strong laser fields (9,403; 1995.5) #24 electron scattering (3828; 1989.2) #25 quantum computation and mechanics (2,981; 2001.9) #26 polarization and instability of lasers (1,462; 1990.1)
#3 optical data processing (43,419; 1993.8)	#31 image processing (12,350; 1993.9) #32 optical computing (9,540; 1994.8) #33 laser intensity direction and ranging (9,360; 1992.4) #34 holographic-interferometry (5,561; 1993.0) #35 image reconstruction (3,777; 1993.0)
#4 optical analysis (28,909; 1991.0)	#41 optical microscopy and analysis (6,397; 1994.1) #42 ionization (5,041; 1989.0) #43 phase transition and fractal growth (4,611; 1986.5) #44 optical spectroscopy (3,413; 1991.2) #45 interaction with materials (2,336; 1996.3) #46 photodiodes (1,165; 1983.2) #47 quantum mechanics (1,052; 1993.7)
#5 lasers (12,636; 1995.5)	#41 semiconductor lasers (2101; 1999.1) #42 pumping and electron state transition (1,987;1997.5) #43 dye lasers (1,420; 1993.9)

Table 3 shows the publication trend of countries in each cluster. Top 5 countries for each cluster are shown. In all of these clusters, USA has the largest publications and TC$_{all}$, which means the leadership and the presence of USA. Japan, Germany, and France, and China have also large number of publications in these clusters. Japan has relative advantage in cluster #5, which focuses on lasers. Japan also has a large number of publications in cluster #1 and #3, while the citedness of Japanese papers are relatively week. But the citedness of them are high in cluster #5. Germany has high citedness and is more focused on cluster #2 and #4. German seems to focus on basic research rather on applied research. France has a large number of publications in cluster

#4 and #5 but not strong position in cluster #2. China has large number of publications especially in cluster #1 and #3, while they are not cited frequently as shown by low TC_{all}.

Table 3. The top five countries in cluster #1–#5

	Cluster #1			Cluster #2			Cluster #3		
Rank	Country	#papers	TC_{all}	Country	#papers	TC_{all}	Country	#papers	TC_{all}
1	USA	16242	13.2	USA	16628	19.6	USA	14813	11.3
2	China	5588	3.5	Germany	4717	14.1	China	3277	2.9
3	Japan	5138	8	Italy	3576	10.8	Japan	3088	6.1
4	Germany	3845	11	China	3105	7.9	Germany	2269	8
5	France	3725	10	Japan	2786	9.4	France	2055	6.8

	Cluster #4			Cluster #5		
Rank	Country	#papers	TC_{all}	Country	#papers	TC_{all}
1	USA	10960	20.9	USA	3495	10.9
2	Germany	2535	14	Japan	1456	8.2
3	France	2374	17.1	France	1130	8.1
4	Japan	2029	9.5	China	1052	3.7
5	China	1335	4.1	Germany	1043	9.7

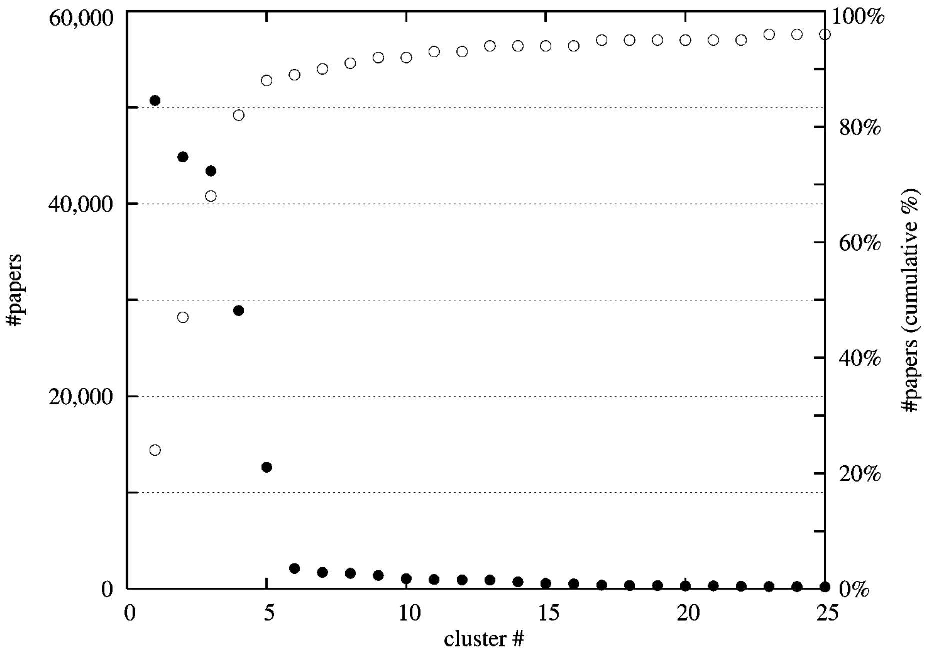

Figure. 2. Cluster size. Closed dots are the number of nodes included in each cluster. Open dots are the cumulative probability of the number of nodes

Figure 3 shows the publication trend of each cluster. It is clearly shown that they exhibit different types of curves. Cluster # 1 has the lowest publication volume until 1980, but after the mid of 1980s it steeply rises. During the late 90s, it levels off but since 2002, it shows tremendous growth. Currently we have about 5,000 annual publications in cluster #1. Cluster #2 was the largest from the mid 70s to 1990, and is currently under developing, but the increasing speed of publication is not high as cluster #1. Cluster #3 has a similar trend with cluster #2, but it exhibited a fluctuated pattern after 1990. Cluster #4 shows a characteristic publication pattern. Already in 70s and 80s, it shows large volume but during the 90s it decreases and saturates. But recently it increases again, which might be driven by the development of other clusters, especially cluster #1. Cluster #5 shows similar trend wit cluster #2. Recently publication amounts are increasing in all of these five clusters. Among these, cluster #1 is the youngest among top 4 clusters, and its subclusters are also young compared to those in the subclusters of the other 1st level clusters. Therefore, in order to detect emerging research domains in optics, we investigate the detailed structures of cluster #1.

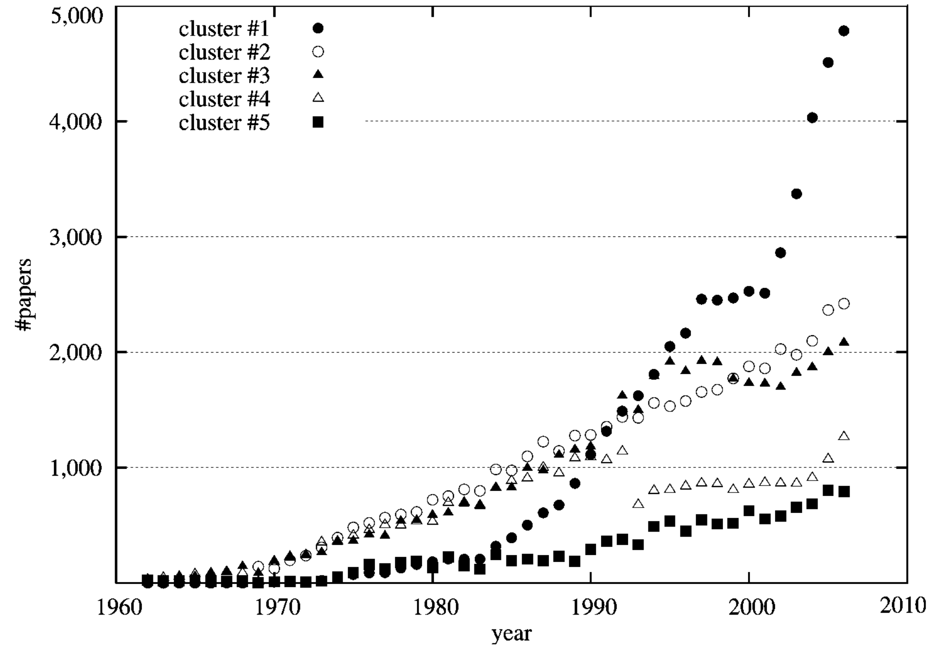

Figure 3. Number of publications in cluster #1-#5

Table 4-7 is 3rd and 4th level subclusters of cluster #1. In these tables, clusters with their size larger than 1,000 for 3rd level clusters and 300 for 4th level clusters are shown. By judging from year$_{ave}$, we can detect emerging subdomains of optics research. For example, in cluster #11, waveguide, photonic crystal optical waveguides (cluster #1121; #node = 923; year$_{ave}$ = 2003.6) and ring and disk resonator (#1123; 704; 2003.0) can be detected as especially emerging clusters. In these two emerging clusters, design, fabrication, and evaluation of new optical devices are investigated.

Table 4. 3rd and 4th level subclusters of the cluster #11 wave guide (15,053;1998.5)

Cluster name (#papers, year$_{ave}$)		Subcluster name (#papers, year$_{ave}$) >300 nodes
#111	beam propagation in fiber (3627; 1993.8)	#1111 refractive-index profiles (939;1994,5) #1112 beam propagation (803; 1995.9) #1113 coupled mode theory (765; 1994.1) #1114 analysis of waveguides (710; 1991.4)
#112	wave guide (3,352; 2001.7)	#1121 photonic crystal optical waveguides (923; 2003.6) #1122 multimode interference and self-imaging(920; 2000.4) #1123 ring and disk resonator (704; 2003.0) #1124 Raman amplification (422; 2001.3)
#113	fiber gratings (3202; 2001.0)	#1131 fiber Bragg gratings (810; 2001.5) #1132 optical fiber grating and sensors (713; 2000.2) #1133 multiwavelength generation in Er-doped fiber (678; 2001.8) #1134 fabrication and characterization of fiber (558; 1999.0)
#114	other fiber components (2971; 1999.6)	#1141 Er-doped fiber amplifier (891; 1998.7) #1142 N x N multiplexer (826; 2000.9) #1143 fiber laser (400;1999.8)

Table 5. 3rd and 4th level subclusters of the cluster #12 converter (13,288; 1999.6)

Cluster name (#papers, year$_{ave}$)		Subcluster name (#papers, year$_{ave}$) >300 nodes
#121	transmittance and amplifiers (3,803; 1999.4)	#1211 polarization dispersion in optical fiber (1,233;1998.2) #1212 transmission and noise control (1,194; 2000.6) #1213 soliton transmission control (890; 1998.9)
#122	nonlinear phenomena (3,464; 1999.5)	#1221 nonlinearity in photonic crystal fiber (1,140; 2003.9) #1222 soliton self-frequency shift (1033; 1993.8) #1223 ring soliton laser and mode-locking (933; 1999.1)
#123	modulators (2,902; 2001.0)	#1231 wavelength converter (640; 1999.3) #1232 frequency modulators and amplifier (638;2001.2) #1233 clock recovery and amplifier (523; 2000.4) #1234 packet switching (376; 2002.9) #1235 broadband parametric amplifier (349; 2003.3)
#124	nonlinear phenomena (2,536; 1998.4)	#1241 spatial solitons (648; 1998.9) #1242 non-linear interface (621; 1994.8) #1243 non-linear phase modulation (563; 1998.0) #1244 nonlinearity in periodic medium (379; 2001.4)

Table 6. 3rd and 4th level subclusters of the cluster #13 generation (12,153; 1998.0)

Cluster name (#papers, year$_{\text{ave}}$)	Subcluster name (#papers, year$_{\text{ave}}$) >300 nodes
#131 ultrashort pulse generation (3,682; 1998.5)	#1311 pulse shaping (1181; 2000.4)
	#1312 self-mode-locking (1066; 1998.8)
	#1313 pulse compression (845; 1994.9)
	#1314 phase control (394; 2001.7)
#132 laser-induced nonlinear processes (3680; 1996.7)	#1321 multi-photon ionization (1,270;1992.5)
	#1322 high-order harmonic-generation (1,011; 1999.9)
	#1323 dissociative ionization (822; 1998.2)
	#1324 electron-atom collision and scattering (430; 1998.9)
#133 property of light (2,847; 2000.1)	#1331 angular-momentum and particle trapping (931; 2001.9)
	#1332 Bessel waves (706; 2000.0)
	#1333 spontaneous optical-pattern formation (692; 1997.0)
	#1334 paraxial wave and focusing (363; 2001.5)

Table 7. 3rd and 4th level subclusters of the cluster #14 modulators and imaging (8,088; 1998.9)

Cluster name (#papers, year$_{\text{ave}}$)	Subcluster name (#papers, year$_{\text{ave}}$) >300 nodes
#141 fiber laser and amplifier (2,306; 1999.3)	#1411 diode laser pumped Nd-YAG laser (557; 1997.1)
	#1412 resonators for solid-state lasers (516; 1997.5)
	#1413 Q-switching and pumping (510; 2001.1)
	#1413 optical parametric oscillator (441; 2000.4)
#142 receivers and modulators (1,885; 1998.2)	#1421 wavelength division multiplexing (508; 1999.1)
	#1422 optical modulators and detectors (366; 2000.1)
	#1423 optical receivers (365; 1993.1)
#143 optical imaging (1,539; 2000.2)	#1421 optical coherence tomography (589; 1999.5)
	#1422 Brillouin scattering in optical fibers (420; 1998.5)
	#1423 in-vivo imaging techniques (314; 2003.9)
#144 laser and gas detection (1,216; 1998.4)	

Among the 4th level clusters in cluster #11, multimode interference and self-imaging(#1122; 920; 2000.4), Raman amplification (#1124; 422; 2001.3), fiber Bragg gratings (#1131; 810; 2001.5), optical fiber grating and sensors (#1132; 713; 2000.2), multiwavelength generation in Er-doped fiber (#1133; 678; 2001.8), and N x N multiplexer (#1142; 826; 2000.9) are also emerging.

In the other 2nd level clusters of cluster #1, the following clusters are emerging. In converter (cluster #12), transmission and noise control (cluster #1212; 1,194; 2000.6), nonlinearity in photonic crystal fiber (#1221; 1,140; 2003.9), frequency modulators and amplifier (#1232; 638; 2001.2), clock recovery and amplifier (#1233; 523; 2000.4), packet switching (#1234; 376; 2002.9), broadband parametric amplifier (#1235; 349; 2003.3), and nonlinearity in periodic medium (#1244; 379; 2001.4) are under rapid developing.

In generation (cluster #13), two of subclusters about ultrashort pulse generation are emerging; pulse shaping (#1311; 1,181; 2000.4); phase control (#1314; 394; 2001.7).

Some 4th level clusters relating property of light are also emerging, i.e., angular-momentum and particle trapping (#1331; 931; 2001.9), Bessel waves (#1332; 706; 2000.0), and paraxial wave and focusing (#1334; 363; 2001.5).

In modulators and imaging (cluster #14), two fiber laser-related subclusters are emerging, i.e., Q-switching and pumping (#1413; 510; 2001.1), optical parametric oscillator (#1413; 441; 2000.4). Another hot topic is optical modulators and detectors (#1422; 366; 2000.1). In cluster #14, usage of lasers for medical applications are also hot topic, i.e., microscopy and tomography for in-vivo imaging techniques (#1423; 314; 2003.9).

Discussion

In this paper, we performed topological clustering method to investigate the structure of research and to detect emerging research domains in optics. In his classical paper, de Solla Price [25] originally introduced the concept of a research front, research domains under developing where papers cites each other densely. According to Price, there seems to be a tendency for scientists to cite the most recently published articles. The research front builds on recent work, and the network there becomes very tight. In a given field, a research front refers to the body of articles that scientists actively cite. Researchers have studied quantitative methods that can be used to identify and track the research front as it evolves over time. Small and Griffith represented currently activated scientific specialties as clusters of co-cited articles [26]. But the term, research front, implies that the research area is hot and currently under rapid developing. Dense network of published articles does not necessarily research front. Therefore, to detect research front, in our words, emerging research domains, the usage of average publication year in the cluster or time slices of the networks [9, 26] are effective. In our analysis, we can detect emerging subdomains of optics research by using average publication year.

We must note that detected research domain does not correspond with a research topic on one-on-one level. For example, in addition to cluster #123 (modulators), cluster #142 (#142 receivers and modulators) also appears and studies modulators. This might be because the researches on optical modulators are under reconfigurations and the citation network is not closely correlated and intermixed. Therefore, it seems for us to keep in mind the limitation of the citation network analysis to detect hot research areas. The clusters derived by citation network approach are determined by individual citations and researcher's motivations behind them, which is obvious but sometimes neglected.

But except for some cases, the analysis is fairly good to detect the emerging research domains as shown in the previous section. It is also worth to point out that citation network is useful approach not only to detect the emerging domains but also to overview and understand the structure of research. Figure 4 visualizes the structures of the citation network for cluster #1–#5. In Figure 4, we assign the same gray-scale color to intra-cluster links for each cluster. The visualization is spring-layout algorithm and therefore, when the structure of a cluster in Figure 4 is compact and round, it means that papers in the cluster has strong tendency to cite other papers in the same cluster. Conversely when a cluster is stretched and spiky, the cluster is closely related with other cluster in the direction. When two clusters are at near positions, it means that the papers in these two papers cite each other.

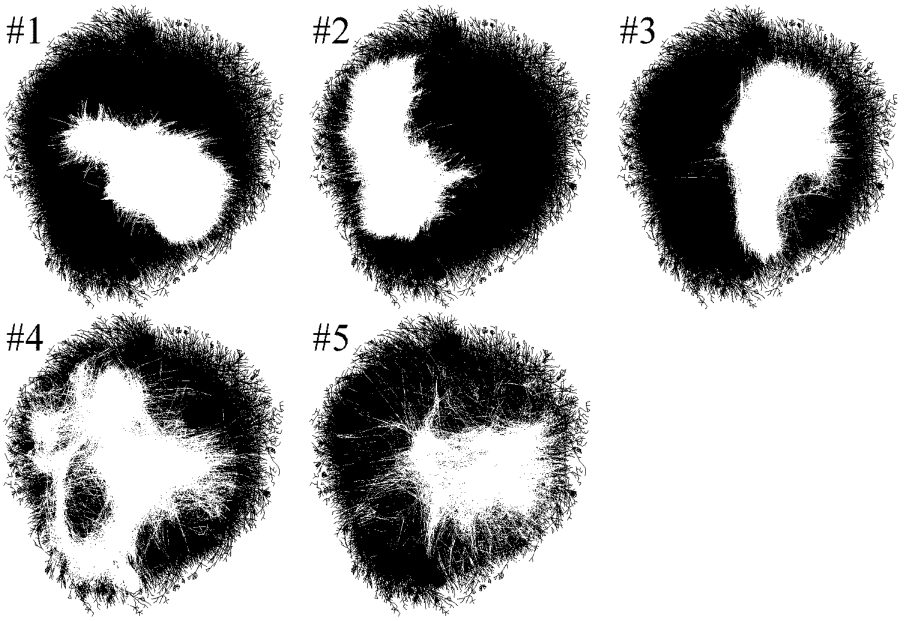

Figure 4. Visualization of the citation networks

As shown in Figure 4, cluster #1, #2, and #3 are compact despite their large network sizes. On the other hand, cluster #4 and #5 are stretched and not compact despite their relatively small sizes, which mean that these clusters have outer links to the other clusters. As already shown in Table 2, cluster #1, #2, and #3 are the central clusters having > 40,000 papers. Cluster #1 and #3 focus on rather applied research for optical communication and data processing, and cluster #2 is basic research. Cluster #4 (optical analysis) and #5 (lasers) offer instruments for these applied and basic research, and therefore, the out come of these latter two clusters are utilized the former big three clusters. Therefore, the visualized network structure of cluster #4 and #5 are stretched, while the other three clusters are compact. Persson [29] made a distinction between a research front and an intellectual base. According to Persson, research front consists of the citing articles and intellectual base consists of the cited articles. Expanding to this definition into the relationships with clusters, we can say that cited old clusters offer intellectual bases for citing emerging clusters. Considering the content of the cluster (Table 2) and their topological positions (Figure 4) with their historical data (Figure 3), we can speculate that optical analysis (cluster #3) and lasers (cluster #5) offer instruments to perform research for applied (cluster #1 and #3) and basic (cluster #2) optics research, and therefore work as intellectual bases.

Conclusion

A prominent feature of modern activities is in the creation, dissemination and application of scientific knowledge. Optics is important research domain both for scientific interest and industrial applications. In this paper, we constructed a citation network of papers and performed topological clustering method to investigate the structure of research and to detect emerging research domains in optics.

We found that optics consist of main five subclusters, optical communication, quantum optics, optical data processing, optical analysis, and lasers. In all of these clusters, USA has the largest publications and receives citations, which means the leadership and the presence of USA. China has large number of publications, but not citations. The average publication year of papers in each cluster indicated that the largest cluster, optical communication, has the lowest publications before 1990 but recently it steeply increases. Then, we further investigated the detailed sub-cluster structures in optical communication cluster. By doing so, we detected some emerging research domains such as nonlinearity in photonic crystal fiber and in-vivo imaging techniques. Finally, we discuss the distinction between research front and intellectual base in optics. The visualization of the network for the top five clusters showed that three large clusters are compact but two relatively small clusters were stretched. This implies that the later works as an intellectual base for the former. Topological clustering method can be used not only for the detection of emerging research field and entire structure of research such as interrelationships among subdomains of the research field.

*

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Grant-in-Aid for Young Scientists (B), 18700240, 2006.

References

1. D. J. D. PRICE, *Little Science, Big Science*, Columbia University Press, New York, 1963.
2. J. KING, A Review of bibliometric and other science indicators and their role in research evaluation, *Journal of Information Science*, 13 (5) (1987) 261–276.
3. E. C. M. NOYONS, H. F. MOED, M. LUWEL, Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study, *Journal of the American Society for Information Science and Technology*, 50 (2) (1999) 115–131.
4. R. ROUSSEAU, Journal evaluation: Technical and practical issues, *Library Trends*, 50 (3) (2002) 418–439.
5. K. BÖRNER, C. CHEN, K. W. BOYACK, Visualizing knowledge domains, *Annual Review of Information Science and Technology*, 37 (2003) 179–255.
6. R. N. KOSTOFF, D. R. TOOTHMAN, H. J. EBERHART, J. A. HUMENIK, Text mining using database tomography and bibliometrics: a review, *Technology Forecasting and Social Change*, 68 (3) (2001) 223–253.
7. K. W. BOYACK, B. N. WYLIE, G. S. DAVIDSON, Domain visualization using VxInsight for science and technology management, *Journal of the American Society for Information Science and Technology*, 53 (2002) 764–774.
8. C. CHEN, T. CRIBBIN, R. MACREDIE, S. MORAR, Visualizing and tracking the growth of competing paradigms: two case studies, *Journal of the American Society for Information Science and Technology*, 53 (2002) 678–689.
9. H. SMALL, Tracking and predicting growth areas in science, *Scientometrics*, 68 (3) (2006) 595–610.
10. A. VERBEEK, K. DEBACKERE, M. LUWEL, E. ZIMMERMANN, Measuring progress and evolution in science and technology. I: The multiple uses of bibliometric indicators, *International Journal of Management Reviews*, 4 (2002) 179–311.
11. A. KAMEOKA, Evaluating research projects at TOSHIBA. designing a conceptual framework of evaluating research and technology development (RTD) programs, *Scientometrics*, 34 (1995) 427–439.
12. M.-Y. TSAY, S.-J. JOU, S.-S. MA, A bibliometric study of semiconductor literature, *Scientometrics*, 49 (2000) 491–509.
13. M.-Y. TSAY, H. XU, C.-W. WU, Journal co-citation analysis of semiconductor literature, *Scientometrics*, 57 (2007) 7–25.
14. M.-Y. TSAY, H. XU, C.-W. WU, Author co-citation analysis of semiconductor literature, *Scientometrics*, 58 (2003) 529–545.
15. R. FURUKAWA, A. GOTO, Core scientists and innovation in Japanese electronics companies, *Scientometrics*, 68 (2006) 227–240.
16. M.-H. HUANG, L.-Y. CHIANG, D.-Z. CHEN, Constructing a patent citation map using bibliographic coupling: A study of Taiwan's high-tech companies, *Scientometrics*, 58 (2003) 489–506.
17. R. ROJO, I. GÓMEZ, Analysis of the Spanish scientific and technological output in the ICT sector, *Scientometrics*, 66 (2006) 101–121.
18. J. GUAN, N. MA, A bibliometric study of China’s semiconductor literature compared with other major asian countries, *Scientometrics*, 70 (2007) 107–124.
19. J. GUAN, Y. HE, Comparison and evaluation of domestic and international outputs in Information Science & Technology research of China, *Scientometrics*, 65 (2006) 215–244.
20. V. TRIMBLE, Productivity and impact of large optical telescopes, *Scientometrics*, 36 (2) (1996) 237–246.
21. M. E. J. NEWMAN, Fast algorithm for detecting community structure in networks, *Physical Review*, E 69 (2004) 066133.
22. M. E. J. NEWMAN, M. GIRVAN, Finding and evaluating community structure in networks, *Physical Review*, E 69 (2004) 026113.
23. R. KLAVANS, K. W. BOYACK, Identifying a better measure of relatedness for mapping science, *Journal of the American Society for Information Science and Technology*, 57 (2006) 251–263.
24. A. T. ADAI, S. V. DATE, S. WIELAND, E. M. MARCOTTE, LGL: Creating a map of protein function with an algorithm for visualizing very large biological networks, *Journal of Molecular Biology*, 340 (1) (2004) 179–190.

25. D. J. D. PRICE, Networks of scientific papers, *Science*, 149 (1965) 510–515.
26. H. G. SMALL, B. C. GRIFFITH, The structure of scientific literatures: I. Identifying and graphing specialties, *Science Studies*, 4 (1974) 17–40.
27. S. A. MORRIS, G. YEN, Z. WU, B. ASNAKE, Timeline visualization of research fronts, *Journal of the American Society for Information Science and Technology*, 54 (2003) 413–422.
28. C. CHEN, CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature, *Journal of the American Society for Information Science and Technology*, 57 (3) (2006) 359–377.
29. O. PERSSON, The intellectual base and research fronts of JASIS 1986-1990, *Journal of the American Society for Information Science*, 45 (1) (1994) 31–38.