J. Eng. Technol. Manage. 32 (2014) 129–146

Contents lists available at ScienceDirect

Journal of Engineering and Technology Management

journal homepage: www.elsevier.com/locate/jengtecman

# Detecting research fronts using different types of weighted citation networks

Katsuhide Fujita [a,*], Yuya Kajikawa [b], Junichiro Mori [c], Ichiro Sakata [d]

[a] Faculty of Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei-shi, Tokyo 184-8588, Japan
[b] Graduate School of Innovation Management, Tokyo Institute of Technology, 3-3-6 Shibaura, Minato-ku, Tokyo 108-0023, Japan
[c] Presidential Endowed Chair for Platinum Society, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[d] School of Engineering, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan

ARTICLE INFO

*Article history:*
Received 1 February 2013
Received in revised form 12 July 2013
Accepted 16 July 2013

*JEL classification:*
O3

*Keywords:*
Research front
Citation network analysis
Bibliometrics
Decision support

ABSTRACT

In this paper, we investigate the performance of different types of weighted citation networks for detecting emerging research fronts by a comparative study. Three citation patterns including direct citation, co-citation and bibliographic coupling, have been tested in three research domains including gallium nitride, complex networks, and nano-carbon. These three patterns of citation networks are constructed for each research domain, and the papers in those domains are divided into clusters to detect the research front. Additionally, we apply some measures to weighted citations like difference in publication years between citing and cited papers and similarities of keywords between them, which are expected to be able to effectively to detect emerging research fronts. To investigate the performance of different types of weighted citation networks for detecting emerging research fields, we evaluate the performance of each approach by using the following measures of extracted research fronts: visibility, speed, and topological and textual relevance.



* Corresponding author. Tel.: +81 42 388 7141; fax: +81 42 388 7141.
*E-mail address:* katfuji@cc.tuat.ac.jp (K. Fujita).

0923-4748/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jengtecman.2013.07.002

## Introduction

Over the past several decades, the number of academic papers has increased exponentially (Price, 1965), and each academic area has become specialized and segmented. Davidson et al. (1998) describe the situation as follows: "For most of history, mankind has suffered from a shortage of information. Now, in just the infancy of the electronic age, we have begun to suffer from information excess". Therefore, it is hard for researchers to perceive their specialized fields as a whole, and segmentation occurs simultaneously with specialization, which brings a severe problem and also opportunity to find crucial knowledge by integrating different domains. Naturally, it is hard for researchers and managers to detect a research front in the early stages by human effort only. Therefore, there is a strong need for computational tools of science mapping and emerging topic detection. Previous studies have established effective algorithms for creating academic landscapes and for detecting emerging topics for certain research fronts.

To support the detection of research fronts and visualization of academic landscapes, methods of science mapping by citation analysis have been proposed and developed (Boyack et al., 2005; Klavans and Boyack, 2009). Researchers have also focused on clustering and visualization (Chen, 1999; Chen et al., 2003; Small, 1999). For example, Leydesdorff (2004) and Leydesdorff and Rafols (2009) made a large-scale investigation of a set of academic papers. Not only creating static academic landscapes, topological and semantic analysis of a citation network also helps us to focus on significant movements in research fronts and emerging research fields in a broad context (Shibata et al., 2008).

The other approach is to detect emerging clusters of densely connected papers. Price (1965) employed the concept of a research front, that is, a research domain under development where papers cite each other densely. Scientists tend to cite the most recently published articles in their papers; therefore, the network belonging in a research front becomes very tight. In a given field, a research front refers to the body of articles that scientists actively cite. Researchers have been studying quantitative methods that can be used to identify and track a research front as it evolves over time. Small and Griffith (1974) showed that activated scientific specialists generate clusters of co-cited papers. Braam et al. (1991a,b) also investigated the topics discussed in co-cited clusters by analyzing the frequency of indexing terms and classification codes occurring in these publications.

On the other hand, different citation patterns between papers offer some ways to detect emerging research domains. Shibata et al. (2009) performed a comparative study to investigate the performance of methods for detecting emerging research fronts between three citation patterns, co-citation, bibliographic coupling, and direct citation. When a paper directly cites another as a reference, it is called a direct citation. In other words, the direct citation is the citing of an earlier paper by a new paper. Co-citation is defined as the edge between two documents cited by the same paper(s) (Small, 1973). Bibliographic coupling is defined as the edge between two documents citing the same paper(s) (Kessler, 1963). Three patterns of citation networks were constructed for each research domain, and the papers in those domains were divided into clusters to detect the research front. Direct citation, which could detect large and young emerging clusters earlier, shows the best performance in detecting a research front, and co-citation shows the worst. Small (2006) proposed a method of tracking and predicting growth areas by co-citation analysis that analyzed co-citation networks generated from the top 1% of highly cited papers. Klavans and Boyack (2006) compared the performance of clustering in journal citation networks created by direct citation and co-citation. Their results suggested that a network of co-citation has higher content similarity. Boyack and Klavans (2010) evaluated textual similarity of papers in clusters extracted by four different citation patterns: bibliographic coupling, direct citation, and a hybrid approach of direct citation and co-citation. Bibliographic coupling slightly outperforms co-citation and direct-citation using textual accuracy measures. In a certain cases, especially for large data set, bibliographic coupling might work better. However, it cannot be applied for research front detections in a specific research domain, because hub papers in a specific domain gather larger citations even when there are less common research topics between citing and cited papers. In fact, this bias would have the less effect on results in large corpuses because citation networks become globally sparse and locally dense.

Despite weighted citation networks can capture important information attributes of papers, most of the existing works focus on the non-weighted citation networks. The purpose of this paper is to

study the characteristics of paper–paper weighted citation networks created by different citation patterns with different weight types. In particular, average publication year, similarities of citation information and similarities of keywords are effective information attributes for detecting research fronts. By introducing them as weights of links to the citation network, it is expected to detect research fronts compared with the non-weighted citation networks effectively.

This paper studies the following three research domains. Gallium nitride (GaN) is widely recognized as a recent prominent innovation in the fields of applied physics and material science. Complex network (CNW) analysis is also recognized as pioneering a new research field after the leading works by physicists has received attention. Nano-carbon (carbon nanotube [CNT]) is also widely recognized as a recent prominent innovation in the fields of applied physics and material science. They are typical examples of recent remarkable innovations having somewhat different characteristics (e.g. breakthrough of the rapid development, material or model-based innovation). These three domains are the same with those selected in our previous study (Shibata et al., 2009). We investigate effectiveness of our weighted citation network approach in these domains to enable comparison with previous results with non-weighted citation network approach. By demonstrating our proposed methodologies, we can show the effectiveness and differences of our proposed method.

We constructed weighted citation networks for each domain and divided the citation networks into clusters to detect research fronts. We evaluated the performance of each method in detecting a research front by comparing visibility, speed, and topological and textual relevance of clustering. Our evaluation strategy is that the best method for detecting the research front is the one that can detect a large, textually and topologically uniform cluster of papers at an earlier stage. Regarding textual similarity, the previous papers deal with text similarity measures for evaluating the effectiveness of citation networks (Boyack and Klavans, 2010; Shibata et al., 2011; Jarneving, 2007; Braam et al., 1991a,b). However, it can only focus on limited aspect of performance that each method should have. By considering the differences, we discuss which type of weight and citation patterns is most suitable for detecting emerging knowledge domains from diverse facets of evaluation.

The remainder of this paper is organized as follows. First, we give an overview of research domains analyzed in our comparative case study. Next, we describe the methodology based on the network clustering and network measures. Then, we present and discuss the performance of the types of weighted citation network for detecting emerging research fronts. Finally, we present our overall conclusions.

## Overview of research domains and core papers

Gallium nitride (GaN), complex network (CNW), and carbon nanotube (CNT) are typical examples of recent remarkable innovations having somewhat different characteristics. Research in GaN has incrementally developed in the field of applied physics. Within a very short period following the mid-1990s, researchers achieved applications of GaN as blue and green light-emitting diodes (LEDs), ultra violet (UV) and blue laser diodes (LDs) (Nakamura, 1991; Nakamura et al., 1994, 1992). Owing to the breakthroughs of overcoming the large acceptor activation energy of GaN by annealing, researchers achieved commercial products. After finding the breakthroughs of overcoming the problems of GaN, the rapid development of GaN research is attracting both researchers and funding. The number of academic papers starts to dramatically increase after 1995. Also, these products are now commercially available, and innovation in this research field has motivated researchers to engage in and open huge new markets.

The second innovation is CNW, which was recently recognized as a new research field. Previously, CNWs have been researched in the following areas: graph theory in mathematics, social network analysis in sociology. Recently, Watts and Barabasi (Barabasi and Albert, 1999; Watts and Strogatz, 1998), whose backgrounds were theoretical and applied mechanics and applied physics revealed the common characteristics of small-world networks (Watts and Strogatz, 1998) and scale-free networks (Barabasi and Albert, 1999). After the leading works by these physicists, studies in this domain have received attention, and a number of papers in this domain have been published. This is probably because their superiority in expressing small-world phenomena by concise quantitative measures assures researchers of future possibility in this research domain.

**Table 1**
Core papers that opened a new research front in three domains.

| Research domain | Core papers |
|---|---|
| Gallium nitride | (A-1) NAKAMURA S, 1991, JPN J APPL PHYS PT 2, V30, P1705<br>(A-2) NAKAMURA S, 1992, JPN J APPL PHYS PT 1, V31, P1258 |
| Complex networks | (B-1) Watts DJ, 1998, NATURE, V393, P440<br>(B-2) Barabasi AL, 1999, SCIENCE, V286, P509 |
| Carbon nanotube | (C) IIJIMA, S, 1991, NATURE, V354, P56 |

The third innovation is CNT, which is useful in nano-science and nanotechnology, due to superior electrical and mechanical properties. A CNT is a nano-sized carbon molecule with morphology like a tube. Fullerenes are also a well-known nano-sized carbon material with morphology like a ball. The existence of fullerenes was known earlier than that of nanotubes (Iijima, 1991). However, after the discovery of the carbon nanotube, the focus of researchers shifted from fullerenes to nanotubes. Therefore, if we could detect research fronts that include papers where the discovery of the nanotube is mentioned, we might expect such a shift of research focus earlier than competitors.

We take these innovative cases because they are typical examples of recent remarkable innovations having somewhat different characteristics. The breakthrough of the rapid development of GaN research is to find the methodologies to overcome the large acceptor activation energy of GaN. On the other hand, the breakthrough of the rapid development of CNT research is to discover the materials (nanotube). Both GaN and CNT are material innovations, however, these breakthroughs are different. The breakthroughs of the rapid development of CNW research are combinations among applied mechanics, sociology and applied physics. Actually, GaN and CNT are material innovations, while CNW is an analytical and model-based innovation. By introducing the CNW, we can demonstrate the effectiveness and differences of our proposed method in some scientific fields.

Core Papers are research papers that receive citations soon after publication, relative to other papers of the same field and age. Generally, papers reach their citation peak two, three, or even four years after publication. However, core papers are recognized very soon after publication, reflected by rapid and significant numbers of citations. These papers are often key researches in their fields. In this paper, core papers are defined as highly cited papers published in the rapid-growth years expected for the review papers using Web of Science, which is a Web-based user interface of the Institute for Scientific Information's (ISI) citation databases. Rapid-growth years in each domain are as follows: Gallium nitride, 1991–1994; CNW, 1998–2001; CNT, 1990–1994.

A list of core papers in each domain, which opened a new research frontier, is shown in Table 1. In GaN, we define the core papers as "(A-1) NAKAMURA S, 1991, JPN J APPL PHYS PT 2, V30, P1705" (Nakamura, 1991) and "(A-2) NAKAMURA S, 1992, JPN J APPL PHYS PT 1, V31, P1258" (Nakamura 1992). In CNW, we define the core paper as "(B-1) Watts DJ, 1998, NATURE, V393, P440" (Watts and Strogatz, 1998) and "(B-2) Barabasi AL, 1999, SCIENCE, V286, P509" (Barabasi and Albert, 1999). In CNT, we define the core paper as "(C) IIJIMA, S, 1991, NATURE, V354, P56" (Iijima, 1991).

## Methodology

The first step is to collect the data of each knowledge domain using Science Citation Index (SCI) and the Social Sciences Citation Index (SSCI) databases. The next step is to create some weighted citation networks. Citation networks are constructed by direct-citation, co-citation and bibliographic-coupling. The weights are frequency of citations, difference of publication years, reference similarity, and keyword similarity. In the third step, maximum connected components were extracted from each network. In the fourth step, we divided the papers in the network into clusters. Finally, we evaluated the visibility, speed, and topological and textual relevance of the clusters to which selected core papers belong.

*Data collection*

First, we collected citation data from the SCI and the SSCI compiled by the Institute for Scientific Information (ISI), which maintains citation databases covering thousands of academic journals and offers bibliographic database services, because SCI and SSCI are two of the best sources for citation data. We used the Web of Science, which is a Web-based user interface of the ISI's citation databases. We searched the papers using the following terms as queries: "*GaN OR gallium nitride*" for the first domain, "*social networks OR social network OR random networks OR random network OR small-world OR scale-free OR complex networks*" for the second domain, and "*carbon AND (nano* OR micro*)*" for the third domain.

In this paper, queries were selected according to the following two steps: (a) the representative keyword, such as gallium nitride and social network, is selected and (b) if the definition of its domain is unclear, more keywords, such as random network, small-world, scale-free, and complex networks, were added. The second step is called as query expansion (Kostoff et al., 1997). Our intention in using so many terms is to retain wide coverage of citation data in order to avoid omission of core papers. For example, we selected the seven search queries in CNW by the query expansion. After selecting the seven queries, we evaluated that these queries retain wide coverage of citation data with avoiding omission of core papers and stopped expanding the queries to eight or more.

The queries for each dataset explained in the previous paragraph are the same as those in the previous paper (Shibata et al., 2009), but retrieved data is not exactly the same because of the data expansion of bibliographic records registered in ISI's databases. The ISI's citation databases enable us to obtain both the attribute data of each paper such as the year published, title, author(s), abstract, author keywords, and citation data.

*Creating weighted citation networks*

After collecting the data including he year published, title, author(s), abstract, author keywords, and citation based on the queries, we create some weighted citation networks. We create citation networks by regarding papers as nodes and three patterns of definitions of citations as edges, as shown in Fig. 1. When a paper directly cites another as a reference, it is called a direct citation. In other words, the direct citation is the citing of an earlier paper by a new paper. Co-citation is defined as the edge between two documents cited by the same paper(s) (Small, 1973). Bibliographic coupling is defined as the edge between two documents citing the same paper(s) (Kessler, 1963). For example, if both papers A and B are cited by C, there is co-citation between A and B; and if both D and E cite C, there is bibliographic coupling between D and E as Fig. 1 shows.

We define the citation graphs $G=(N, E, w)$ comprising a set $N$ of nodes, with each node $N_i$ representing a paper $p_i$ and a set $E$ of edges, with each edge $E_{ij}$ directed from the citing node $N_i$ to

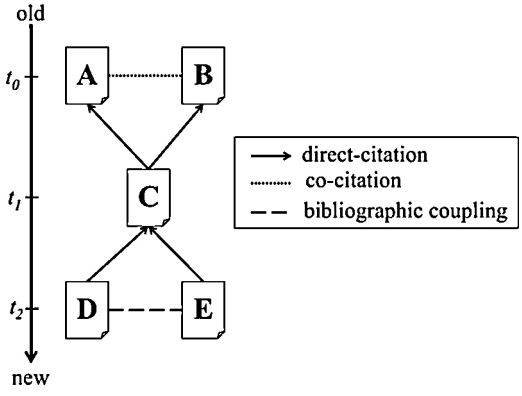


**Fig. 1.** Citation patterns.

the cited node $N_j$, or from the citing node $N_j$ to the cited node $N_i$. $|E_{ij}|$ means the number of citations between $p_i$ and $p_j$. Usually, the number of direct citations is one; however, the number of co-citations and bibliographic-couplings is more than one. In other words, we will build the citation networks defined as a weighted non-directed graph, with each paper representing a node and three patterns of citations representing the edges in the graph. Each node ($N_i$) has several attributes: paper title, author(s), year of publication ($y_i$) and journal name, reference information ($R_i$), and AuthorKeywords and KeyWord Plus ($K_i$). AuthorKeywords are set by the authors when they write the papers. KeyWords Plus is terms which appear in the titles of multiple references.

The network is created in each year, enabling a time-series analysis of citation networks. When we create citation networks on year $y$, we use the data of papers published from 1970 to $y$. In this paper, only the largest-graph component is used because this paper focuses on the relationship among papers, and we should therefore eliminate papers that have no link with the largest-graph component.

We also introduced four types of weights to the citation networks: (i) Frequency of citations, (ii) Publication years, (iii) Reference similarity, (iv) Keyword similarity. The definitions of these weights are as follows:

(i) Frequency of citations: $w(E_{ij})=|E_{ij}|$
(ii) Publication years: $w(E_{ij})=\{(y_i+y_j)/2-1970\}/40$ if $(y_i+y_j)/2<1970$, $w(E_{ij})=0$
(iii) Reference similarity: $w(E_{ij})=Jaccard(R_i, R_j)+1$
(iv) Keyword similarity: $w(E_{ij})=Jaccard(K_i, K_j)+1$

*$Jaccard(x,y)=|x\cap y|/|x\bigcup y|$ (Jaccard similarity is defined by Jaccard, 1912).

By introducing some types of weights based on the attributes, we can detect the research fronts reflecting the important attributes, such as new research fronts growing rapidly.

It is known that normalization has a large effect on clustering results (Jaina et al., 2005). Previous research in citation-based clustering using co-citation and bibliographic coupling has used normalization, because co-citation and bibliographic coupling can give weights to the links as our approach of "(i) Frequency of citations". In this paper, we also adopt min–max normalization technique to some weights. The min–max normalization the process of taking data measured in rates and transforming it to a value between 0.0 and 1.0 is a particular technique of the statistical normalization. The normalized value of this technique is defined as: (*the value* − *the minimum value*)/(*the range of values*). The original functions for weight (ii) in this paper and Jaccard similarity are also based on the min–max normalization technique.

*Topological clustering*

After creating some weighted citation networks, which are constructed by direct-citation, co-citation and bibliographic-coupling, maximum connected components were extracted from each network. A maximum connected component is an "isolated" part of a citation network, which does not have citations to and from another part, and the maximum component is the part that includes most papers in it. By doing this step, non-relevant papers that do not cite papers in the corresponding research domain are removed.

After that, we divided the papers in the network into clusters. For dividing into clusters, a fast-modularity clustering proposed by Newman (2004) is applied in order to discover tightly knit clusters with a high density of within-cluster edges, which enables the creation of a weighted graph consisting of a large number of nodes. The algorithm is based on the idea of modularity $Q$, which is defined as follows:

$$Q=\sum_s (w_{ss}-a_s^2)=Tr(w)-||w||^2$$

where $w_{st}$ is the possibility of the weights of edges in the network that connected nodes in cluster $s$ to those in cluster $t$, and $a_s=\sum_t w_{st}$. In the first part of the equation, $Tr(w)$ represents the sum of density of weights of edges within each cluster. A high value of this parameter means that nodes are densely

connected within each cluster. The second part of the equation, $||w||^2$, represents the sum of density of weights of edges within each cluster when all edges are placed randomly.

In Newman's method, edges that connect clusters sparsely and extract clusters within which nodes are connected densely is cut. A high value of $Q$ represents good community division where only dense edges remain within clusters and sparse edges between clusters are cut off, and $Q=0$ means that a particular division gives no more within-community edges than would be expected by random chance. Then, the algorithm to optimize $Q$ over all possible divisions to find the best structure of clusters is as follows. Starting with a state in which each node is the only member of one of the $n$ clusters, we repeatedly join clusters together in pairs, choosing at each step the joining that results in the greatest increase in $Q$. The change in $Q$ upon joining two clusters is given by

$$\Delta Q = w_{st} + w_{ts} - 2a_s a_t$$

In this paper, we stop joining when $\Delta Q < 0$.

*Topological measures for evaluating citation networks*

By conducting the previous steps written in subsections, we can detect some clusters by dividing some kinds of weighted citation networks. For comparing the tendency of some types of weighted citation networks, visibility, speed, and topological and textual relevance are calculated after clustering for each cluster to which these selected core papers belong. In this paper, we assume that the important front is detected as a larger and more relevant cluster at an earlier stage. When the normalized size of the cluster is larger, we can more easily distinguish the existence of emerging clusters from other clusters. When the average publication year of the cluster is younger, it means that the cluster can be speedily detected at its emerging stage. If there is a time lag for detecting the research fronts, we could fail to find the research fronts in the emerging stage because of the lack of the methodologies' speed. In other words, the lack of speed of emerging detections could fail to grow the seeds of innovations in the industry. Therefore, we consider the speed as the one of the most important measure for evaluating the methodologies. If the cluster is denser, we can check whether clustering is successful for dividing into clusters in the citation networks. If the cluster is more textually relevant, we can detect the textually similar clusters.

The size of a cluster is defined as normalized size to the relative in order to compare certain types of weights:

$$\frac{|N_i \in C|}{|N|}$$

where $|N|$ is the total number of entire nodes $N$ and $|N_i \in C|$ is the number of nodes in cluster $C$.

The density is defined as follows:

$$\frac{|E_i \in C|}{\binom{|N|}{2}}$$

where $|E_i \in C|$ is the number of edges, both of the nodes are in cluster $C$, and $\binom{|N|}{2}$ is the number of combinations from $|N|$ to 2.

The textual similarity between clusters is defined as follows:

$$\frac{\sum_{p_i \in C, p_j \in C (i \neq j)} Sim(p_i, p_j)}{\sum_{C_i \in C_{all}} Sim(C, C_i)}$$

where $p_i \in C$ and $p_j \in C$ are papers in cluster $C$, and $C_i \in C_{all}$ is cluster set. $Sim(p_i, p_j)$ is the similarity measure between paper $p_i$ and paper $p_j$, and $Sim(C, C_i)$ is the similarity measure between cluster $C$ and cluster $C_i$.

$$Sim(p_i, p_j) = \sum_k t\,fid\,f_{p_i}^{(k)} t\,fid\,f_{p_j}^{(k)} \quad \text{where } t\,fid\,f_p^{(k)} = tf_{k,p} \times \log\left(\frac{N}{df_k}\right)$$

$$Sim(C, C_j) = \sum_j t\,fid\,f_c^{(j)} t\,fid\,f_{c_i}^{(j)} \quad \text{where } t\,fid\,f_c^{(j)} = (\sum_{d \in C} tf_{j,d} \times \log\left(\frac{N}{df_j}\right)$$

This textual similarity is based on Jarneving (2007) and is one of the general measures for evaluations in the bibliometrics field.

## Results

### *Basic topologies of the networks*

Fig. 2 shows the time series of $Q_{\max}$ of each research domain. In some years, $Q_{\max}$ in the weight (ii) is the largest in the three patterns of citations. These results are common regardless of the domain and mean that citation network with the weight of the difference of publication years has a "locally dense and globally sparse" structure and can be divided into clusters better than the others. In most of the networks, the $Q_{\max}$ becomes smaller as the domain grows. This suggests that the network becomes random as the domain evolves, partly because it becomes denser not only locally but also globally and cannot be divided well. $Q_{\max}$ becomes higher when extracted clusters do not depend on other clusters. In other words, there are many intra-links but fewer inter-links. The low value of $Q_{\max}$ means that the network is close to a random network that is created by giving all possible ties based on a uniform probability. $Q_{\max}$ becomes 0 in wholly randomized network.

In addition, weight (iii) is almost the same value as weight (iv). This result shows that the "(iii) Reference Similarity" has similar effectiveness to "(iv) Keyword Similarity" in dividing some clusters based on Topological Clustering. The property of paper keywords is similar to that of references at least in our datasets.

### *Performance of each method in detecting emerging domains*

After clustering the networks, we evaluated the performance of the results in each weighted citation network in detecting emerging research domains. The following measures of the cluster, to which selected core papers in each domain belong, were tracked: visibility (as normalized size), speed (as average publication year), text relevance (as text similarity), and topological relevance (as density). The results of the clusters to which core papers belong are shown in Tables 2–6. When the core papers don't form a cluster (e.g. the core paper generates a cluster whose size is 1), the results show "0". When the core paper does not belong in the largest component in generating the citation network, the results show "–".

*Direct citations.* All weights of edges are 1 using weight (i) because a paper cites another paper only once. Therefore, all results of citation networks using weight (i) mean the ones without weights. The normalized size in the citation network using weights (ii)–(iv) is a little smaller than that of weight (i). Usually, the clusters with weighted citation networks form smaller cluster sizes than those with non-weighted citation networks. On the other hand, the density in the citation network using weights (ii)–(iv) is higher than that of weight (i). The weights of citation networks give some effects of dividing into dense and small-sized clusters. In the citation networks with weight (ii), core papers cannot generate the clusters in the expansion stage. In addition, the clusters with weight (ii) become smaller and denser than the ones without the weights. Weight (iii) has similar results to weight (iv): almost the same sizes, average years. This is because the references of the academic paper are tightly related to keywords of the academic papers. On the other hand, the topological relevance of the results with weight (iii) is higher than those with weight (iv), and the textual similarity of the results with weight (iv) is higher than those with weight (iii).

*Co-citations.* In this citation pattern, the results of comparisons between the weights are almost the same as the direct citations. Weight (i) could consider the frequency of citations in the co-citation. Therefore, the density, textual similarity, and average year with weights are better than those without weights. The shortcoming of co-citation is the existence of a time lag in co-citation as pointed out by

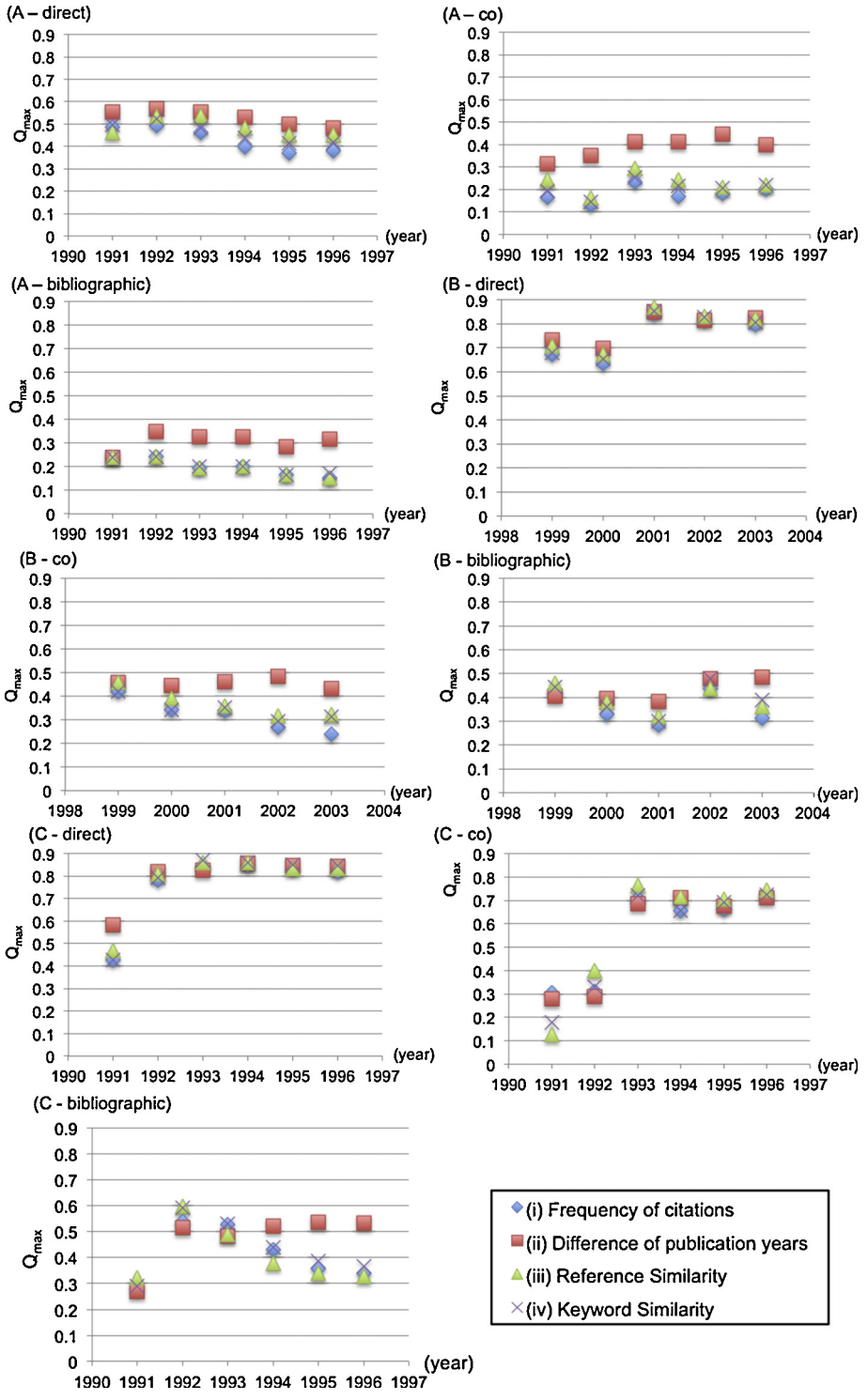


**Fig. 2.** $Q_{max}$ value of each domain: (A) gallium nitride, (B) complex networks, and (C) carbon nanotubes.

**Table 2**
Normalized size, average publication year, density, and textual similarity of the clusters to which core papers belong ((A-1) NAKAMURA S, 1991, JPN J APPL PHYS PT 2, V30, P1705: GaN).

| Year | Direct citation | | | | Co-citation | | | | Bibliographic coupling | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | Density | Text | Avg. year | Size | Density | Text | Avg. year | Size | Density | Text | Avg. year |
| (i) Frequency of citations | | | | | | | | | | | | |
| 1991 | 18 | 0.6923 | 0.2059 | 1989.01 | – | – | – | – | 8 | 14.476 | 0.3783 | 1989.48 |
| 1992 | 24 | 0.6624 | 0.6640 | 1989.96 | – | – | – | – | 39 | 5.4642 | 0.4112 | 1988.74 |
| 1993 | 26 | 0.6547 | 1.0637 | 1991.43 | – | – | – | – | 41 | 7.2106 | 1.4100 | 1992.00 |
| 1994 | 28 | 0.6328 | 1.2181 | 1992.46 | – | – | – | – | 47 | 7.9025 | 1.2855 | 1992.92 |
| 1995 | 37 | 0.364 | 1.0420 | 1993.87 | – | – | – | – | 7 | 12.45 | 0.7682 | 1992.44 |
| 1996 | 32 | 0.2267 | 1.2559 | 1995.00 | – | – | – | – | 30 | 6.6962 | 1.3418 | 1995.19 |
| (ii) Difference of publication years | | | | | | | | | | | | |
| 1991 | 19 | 0.8771 | 0.3411 | 1987.81 | – | – | – | – | 43 | 4.9336 | 0.1567 | 1985.01 |
| 1992 | 23 | 0.8691 | 0.5976 | 1989.82 | – | – | – | – | 22 | 7.3939 | 0.4506 | 1989.42 |
| 1993 | 28 | 0.8402 | 0.9636 | 1991.13 | – | – | – | – | 39 | 7.8362 | 1.1895 | 1991.99 |
| 1994 | 26 | 0.7385 | 1.0179 | 1992.40 | – | – | – | – | 0 | 0 | 0 | 0 |
| 1995 | 30 | 0.4652 | 1.0465 | 1993.87 | – | – | – | – | 0 | 0 | 0 | 0 |
| 1996 | 32 | 0.2661 | 1.2704 | 1994.83 | – | – | – | – | 0 | 0 | 0 | 0 |
| (iii) Reference similarity | | | | | | | | | | | | |
| 1991 | 16 | 0.9343 | 0.5567 | 1988.21 | – | – | – | – | 42 | 8.0237 | 0.3894 | 1984.07 |
| 1992 | 22 | 0.8367 | 0.8651 | 1989.58 | – | – | – | – | 40 | 7.5632 | 0.5884 | 1988.14 |
| 1993 | 27 | 0.8738 | 1.0323 | 1991.13 | – | – | – | – | 38 | 7.7651 | 0.9322 | 1991.99 |
| 1994 | 26 | 0.7536 | 1.1043 | 1991.63 | – | – | – | – | 48 | 7.9772 | 1.3901 | 1992.85 |
| 1995 | 24 | 0.5462 | 1.0356 | 1993.03 | – | – | – | – | 56 | 7.3034 | 1.2639 | 1994.05 |
| 1996 | 32 | 0.4537 | 1.1883 | 1994.69 | – | – | – | – | 49 | 7.6888 | 1.4563 | 1995.22 |
| (iv) Keyword similarity | | | | | | | | | | | | |
| 1991 | 20 | 0.7431 | 0.3774 | 1988.49 | – | – | – | | 9 | 7.8769 | 0.3930 | 1990.13 |
| 1992 | 25 | 0.8022 | 0.5168 | 1989.29 | – | – | – | – | 30 | 5.1136 | 0.9610 | 1991.09 |
| 1993 | 26 | 0.8422 | 0.8463 | 1991.06 | – | – | – | – | 39 | 6.3339 | 1.0416 | 1991.82 |
| 1994 | 19 | 0.9522 | 1.1333 | 1992.35 | – | – | – | – | 45 | 7.9739 | 1.3529 | 1992.85 |
| 1995 | 23 | 0.5164 | 1.2942 | 1993.53 | – | – | – | – | 52 | 7.7374 | 1.3100 | 1994.05 |
| 1996 | 19 | 0.3631 | 1.3555 | 1994.91 | – | – | – | – | 49 | 7.5946 | 1.7370 | 1995.22 |

When the core papers do not belong in the cluster, the results show "0". When the core papers do not belong in the largest component in generating the citation network, the results show "–".

**Table 3**
Normalized size, average publication year, density, and textual similarity of the clusters to which core papers belong ((A-2) NAKAMURA S, 1992, JPN J APPL PHYS PT 1, V31, P1258: GaN).

| Year | Direct citation | | | | Co-citation | | | | Bibliographic coupling | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | Density | Text | Avg. year | Size | Density | Text | Avg. year | Size | Density | Text | Avg.year |
| (i) Frequency of citations | | | | | | | | | | | | |
| 1992 | 24 | 0.6624 | 0.6640 | 1989.96 | 0 | 0 | 0 | 0 | 39 | 5.4642 | 0.4112 | 1988.74 |
| 1993 | 26 | 0.6547 | 1.0637 | 1991.43 | 45 | 5.4538 | 0.3780 | 1988.90 | 41 | 7.2106 | 1.4100 | 1992.00 |
| 1994 | 28 | 0.6328 | 1.2181 | 1992.46 | 53 | 3.8517 | 0.9618 | 1991.37 | 47 | 7.9025 | 1.2855 | 1992.92 |
| 1995 | 37 | 0.364 | 1.0420 | 1993.87 | 55 | 3.2105 | 1.1208 | 1993.03 | 7 | 12.45 | 0.7682 | 1992.44 |
| 1996 | 32 | 0.2267 | 1.2559 | 1995.00 | 49 | 4.1492 | 0.9412 | 1994.28 | 30 | 6.6962 | 1.3418 | 1995.19 |
| (ii) Difference of publication years | | | | | | | | | | | | |
| 1992 | 23 | 0.8691 | 0.5976 | 1989.82 | 25 | 10.909 | 0.0927 | 1984.82 | 49 | 6.3323 | 0.0235 | 1981.76 |
| 1993 | 28 | 0.8402 | 0.9636 | 1991.13 | 36 | 4.6909 | 0.5508 | 1989.17 | 39 | 7.8362 | 1.1895 | 1991.99 |
| 1994 | 26 | 0.7385 | 1.0179 | 1992.40 | 51 | 3.8147 | 1.0043 | 1991.20 | 0 | 0 | 0 | 0 |
| 1995 | 30 | 0.4652 | 1.0465 | 1993.87 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1996 | 32 | 0.2661 | 1.2704 | 1994.83 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| (iii) Reference similarity | | | | | | | | | | | | |
| 1992 | 22 | 0.8367 | 0.5110 | 1989.58 | 5 | 11.538 | 0.1122 | 1984.08 | 30 | 11.538 | 0.1122 | 1990.60 |
| 1993 | 27 | 0.8738 | 0.7945 | 1991.13 | 38 | 5.9194 | 0.5668 | 1989.14 | 42 | 4.9194 | 0.5668 | 1991.08 |
| 1994 | 19 | 0.9782 | 0.9427 | 1992.15 | 51 | 4.77 | 1.0530 | 1991.24 | 40 | 3.77 | 1.0530 | 1992.02 |
| 1995 | 28 | 0.9546 | 0.9384 | 1993.85 | 60 | 3.8787 | 1.1194 | 1992.67 | 32 | 3.0787 | 1.1194 | 1993.24 |
| 1996 | 32 | 1.0091 | 1.1186 | 1994.69 | 56 | 1.8523 | 0.9767 | 1993.64 | 32 | 1.8523 | 0.9767 | 1994.30 |
| (iv) Keyword similarity | | | | | | | | | | | | |
| 1992 | 25 | 0.8022 | 0.1320 | 1989.3 | 37 | 8.181 | 0.16110 | 1979.61 | 26 | 7.2898 | 1.0845 | 1991.09 |
| 1993 | 26 | 0.8422 | 0.8389 | 1991.1 | 36 | 4.0648 | 1.1029 | 1989.56 | 41 | 7.368 | 1.1977 | 1991.82 |
| 1994 | 18 | 0.7823 | 1.1559 | 1992.2 | 52 | 3.86 | 1.2335 | 1991.19 | 45 | 8.3358 | 1.8647 | 1992.86 |
| 1995 | 23 | 0.5164 | 0.9909 | 1993.5 | 63 | 3.2483 | 1.3168 | 1992.86 | 33 | 10.105 | 1.2730 | 1993.74 |
| 1996 | 19 | 0.3631 | 1.3555 | 1994.9 | 68 | 1.2695 | 1.4646 | 1994.21 | 29 | 10.781 | 1.7610 | 1994.62 |

When the core papers do not belong in the cluster, the results show "0". When the core papers do not belong in the largest component in generating the citation network, the results show "–".

**Table 4**
Normalized size, average publication year, density, and textual similarity of the clusters to which core papers belong ((B-1) Watts DJ, 1998, NATURE, V393, P440: Complex Networks).

| Year | Direct citation | | | | Co-citation | | | | Bibliographic coupling | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | Density | Text | Avg. year | Size | Density | Text | Avg. year | Size | Density | Text | Avg. year |
| (i) Frequency of citations | | | | | | | | | | | | |
| 1998 | – | – | – | – | – | – | – | – | – | – | – | |
| 1999 | 1 | 0.474 | 2.6256 | 1998.68 | 3 | 2.1319 | 2.0183 | 1994.04 | – | – | – | – |
| 2000 | 4 | 0.0504 | 2.5901 | 1999.42 | 11 | 0.2744 | 1.9791 | 1994.12 | – | – | – | – |
| 2001 | 9 | 0.0599 | 2.6562 | 2000.51 | 5 | 0.6921 | 2.4099 | 1995.92 | – | – | – | – |
| 2002 | 17 | 0.0648 | 2.8793 | 2001.15 | 7 | 0.8217 | 2.9045 | 1997.99 | – | – | – | – |
| 2003 | 22 | 0.0346 | 2.0593 | 2002.00 | 8 | 1.3668 | 2.9727 | 1999.42 | – | – | – | – |
| (ii) Difference of publication years | | | | | | | | | | | | |
| 1998 | – | – | – | – | – | – | – | – | – | – | – | |
| 1999 | 1 | 0.5018 | 2.8161 | 1996.10 | 0 | 0 | 0 | 0 | – | – | – | – |
| 2000 | 0 | 0 | 0 | 0.00 | 0 | 0 | 0 | 0 | – | – | – | – |
| 2001 | 0 | 0 | 0 | 0.00 | 6 | 0.5476 | 2.4579 | 1996.56 | – | – | – | – |
| 2002 | 0 | 0 | 0 | 0.00 | 0 | 0 | 0 | 0 | – | – | – | – |
| 2003 | 4 | 0.1937 | 3.4409 | 2001.78 | 0 | 0 | 0 | 0 | – | – | – | – |
| (iii) Reference similarity | | | | | | | | | | | | |
| 1998 | – | – | – | – | – | – | – | – | – | – | – | |
| 1999 | 2 | 0.3821 | 3.0023 | 1996.32 | 2 | 2.8506 | 2.1315 | 1994.73 | – | – | – | – |
| 2000 | 3 | 0.346 | 2.7000 | 1996.37 | 9 | 0.8684 | 2.1999 | 1995.12 | – | – | – | – |
| 2001 | 4 | 0.2131 | 2.4088 | 1998.49 | 8 | 0.6348 | 2.1647 | 1996.71 | – | – | – | – |
| 2002 | 4 | 0.2156 | 2.4236 | 2000.37 | 6 | 0.8821 | 2.0359 | 1998.30 | – | – | – | – |
| 2003 | 6 | 0.1904 | 3.1932 | 2001.78 | 10 | 1.389 | 2.4006 | 1998.96 | – | – | – | – |
| (iv) Keyword similarity | | | | | | | | | | | | |
| 1998 | – | – | – | – | – | – | – | – | – | – | – | |
| 1999 | 2 | 0.2838 | 3.0805 | 1996.18 | 2 | 2.5506 | 2.4025 | 1995.10 | – | – | – | – |
| 2000 | 3 | 0.1346 | 2.1593 | 1996.67 | 7 | 0.4641 | 2.2844 | 1995.08 | – | – | – | – |
| 2001 | 3 | 0.1239 | 2.4659 | 1998.51 | 8 | 0.4348 | 2.4937 | 1996.35 | – | – | – | – |
| 2002 | 4 | 0.1568 | 2.6845 | 2000.52 | 8 | 0.6817 | 2.6009 | 1998.17 | – | – | – | – |
| 2003 | 6 | 0.0949 | 3.3899 | 2001.37 | 10 | 0.9152 | 2.8085 | 1999.05 | – | – | – | – |

When the core papers do not belong in the cluster, the results show "0". When the core papers do not belong in the largest component in generating the citation network, the results show "–".

**Table 5**
Normalized size, average publication year, density, and textual similarity of the clusters to which core papers belong ((B-2) Barabasi AL, 1999, SCIENCE, V286, P509: Complex Networks).

| Year | Direct citation | | | | Co-citation | | | | Bibliographic coupling | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | Density | Text | Avg. year | Size | Density | Text | Avg. year | Size | Density | Text | Avg. year |
| (i) Frequency of citations | | | | | | | | | | | | |
| 1999 | 1 | 0.474 | 2.6256 | 1998.68 | – | – | – | – | 19 | 1.1893 | 2.2261 | 1994.62 |
| 2000 | 4 | 0.0504 | 2.5901 | 1999.42 | 11 | 0.2744 | 1.9791 | 1994.12 | 9 | 0.4514 | 2.5450 | 1997.03 |
| 2001 | 9 | 0.0599 | 2.6562 | 2000.51 | 5 | 0.6921 | 2.4099 | 1995.92 | 9 | 0.6509 | 3.0792 | 1998.21 |
| 2002 | 17 | 0.0648 | 2.8793 | 2001.15 | 7 | 0.8217 | 2.9045 | 1997.99 | 7 | 2.8825 | 3.1349 | 2000.16 |
| 2003 | 22 | 0.0346 | 2.0593 | 2002.00 | 8 | 1.3668 | 2.9727 | 1999.42 | 8 | 4.3301 | 3.5362 | 2001.53 |
| (ii) Difference of publication years | | | | | | | | | | | | |
| 1999 | 1 | 0.5018 | 2.8161 | 1996.10 | – | – | – | – | 0 | 0 | 0 | 0 |
| 2000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2001 | 0 | 0 | 0 | 0 | 6 | 0.5476 | 2.4579 | 1996.56 | 0 | 0 | 0 | 0 |
| 2002 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2003 | 4 | 0.1937 | 3.4409 | 2001.78 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| (iii) Reference Similarity | | | | | | | | | | | | |
| 1999 | 2 | 0.5838 | 2.6805 | 1996.18 | – | – | – | – | 19 | 1.4833 | 1.502 | 1994.73 |
| 2000 | 4 | 0.3346 | 2.0059 | 1996.67 | 6 | 0.6666 | 1.2652 | 1999.67 | 6 | 1.6888 | 1.3067 | 1999.80 |
| 2001 | 3 | 0.9666 | 1.4848 | 2000.60 | 7 | 0.6742 | 1.4232 | 2000.32 | 7 | 1.7387 | 1.6797 | 2000.60 |
| 2002 | 6 | 0.5681 | 1.7132 | 2001.42 | 8 | 0.9642 | 1.8284 | 2001.22 | 8 | 1.8487 | 2.0303 | 2001.54 |
| 2003 | 6 | 0.571 | 1.8912 | 2002.06 | 3 | 1.1893 | 1.9705 | 2001.72 | 8 | 1.8411 | 2.1857 | 2002.21 |
| (iv) Keyword similarity | | | | | | | | | | | | |
| 1999 | 2 | 0.2838 | 3.0805 | 1996.18 | – | – | – | – | 19 | 1.2006 | 2.4178 | 1994.86 |
| 2000 | 3 | 0.1346 | 2.1593 | 1996.67 | 7 | 0.4641 | 2.2844 | 1995.08 | 9 | 0.4333 | 2.4869 | 1996.57 |
| 2001 | 3 | 0.1239 | 2.4659 | 1998.51 | 8 | 0.4348 | 2.4937 | 1996.35 | 9 | 0.6702 | 2.9492 | 1997.87 |
| 2002 | 4 | 0.1568 | 2.6845 | 2000.52 | 8 | 0.6817 | 2.6009 | 1998.17 | 8 | 2.2041 | 3.0275 | 1999.89 |
| 2003 | 6 | 0.0949 | 3.3899 | 2001.37 | 10 | 0.9152 | 2.8085 | 1999.05 | 8 | 4.036 | 3.3062 | 2001.50 |

When the core papers do not belong in the cluster, the results show "0". When the core papers do not belong in the largest component in generating the citation network, the results show "–".

**Table 6**
Normalized size, average publication year, density, and textual similarity of the clusters to which core papers belong ((C) IIJIMA, S, 1991, NATURE, V354, P56: Carbon Nano Tube).

| Year | Direct citation | | | | Co-citation | | | | Bibliographic coupling | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Size | Density | Text | Avg. year | Size | Density | Text | Avg. year | Size | Density | Text | Avg. year |
| (i) Frequency of citations | | | | | | | | | | | | |
| 1991 | 1 | 2.8571 | 1.14347 | 1989.43 | – | – | – | – | 23 | 1.8826 | 0.6802 | 1989.11 |
| 1992 | 8 | 0.6057 | 1.76990 | 1991.60 | 10 | 3.746 | 1.8493 | 1991 | 12 | 4.2622 | 1.7598 | 1991.69 |
| 1993 | 10 | 0.3358 | 2.27078 | 1992.12 | 8 | 4.2261 | 1.4992 | 1991.69 | 10 | 13.3736 | 2.0069 | 1992.60 |
| 1994 | 9 | 0.2635 | 2.35245 | 1993.09 | 10 | 2.5471 | 2.3955 | 1992.42 | 10 | 0.7754 | 2.4486 | 1993.00 |
| 1995 | 9 | 0.2071 | 2.54667 | 1993.78 | 0 | 0 | 0 | 0 | 3 | 2.2466 | 2.3989 | 1993.27 |
| (ii) Difference of publication years | | | | | | | | | | | | |
| 1991 | 6 | 1.1 | 0.1597 | 1981.84 | – | – | – | – | 23 | 1.874 | 0.7016 | 1989.13 |
| 1992 | 8 | 0.592 | 1.7599 | 1991.61 | 10 | 3.746 | 1.4139 | 1991.00 | 12 | 4.2622 | 1.6821 | 1991.69 |
| 1993 | 7 | 0.3419 | 2.3222 | 1992.15 | 14 | 2.3881 | 1.6314 | 1991.63 | 16 | 6.5004 | 2.0291 | 1992.51 |
| 1994 | 9 | 0.2707 | 2.3434 | 1993.07 | 17 | 1.1417 | 1.9159 | 1991.59 | 2 | 4.2778 | 1.9473 | 1992.86 |
| 1995 | 9 | 0.2108 | 2.6583 | 1993.77 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| (iii) Reference similarity | | | | | | | | | | | | |
| 1991 | 1 | 2.8571 | 1.20748 | 1989.43 | – | – | – | – | 19 | 2.3076 | 0.6013 | 1988.83 |
| 1992 | 8 | 0.792 | 1.61483 | 1991.61 | 10 | 3.746 | 1.19537 | 1991.00 | 12 | 4.2622 | 1.6865 | 1991.69 |
| 1993 | 7 | 0.4545 | 1.9181 | 1992.24 | 14 | 2.8296 | 1.58136 | 1991.65 | 15 | 7.0554 | 1.9091 | 1992.52 |
| 1994 | 9 | 0.3731 | 1.8792 | 1993.07 | 11 | 2.2262 | 1.8784 | 1992.13 | 16 | 4.7103 | 2.1298 | 1993.18 |
| 1995 | 9 | 0.3126 | 1.8257 | 1993.78 | 12 | 2.2716 | 1.7338 | 1992.97 | 10 | 12.9409 | 2.0442 | 1994.06 |
| (iv) Keyword similarity | | | | | | | | | | | | |
| 1991 | 1 | 2.8571 | 1.14929 | 1989.43 | – | – | – | – | 17 | 2.6271 | 0.6244 | 1989.15 |
| 1992 | 8 | 0.5789 | 1.7573 | 1991.55 | 10 | 3.746 | 1.8968 | 1991.55 | 16 | 2.9074 | 1.6613 | 1991.70 |
| 1993 | 7 | 0.3419 | 2.3358 | 1992.15 | 13 | 2.535 | 1.6791 | 1991.84 | 16 | 6.6648 | 2.0540 | 1992.52 |
| 1994 | 9 | 0.2695 | 2.4406 | 1993.07 | 15 | 2.5579 | 2.2917 | 1992.27 | 11 | 8.9546 | 2.3839 | 1993.40 |
| 1995 | 9 | 0.2083 | 2.73252 | 1993.76 | 13 | 2.6353 | 2.3983 | 1993.02 | 17 | 4.8751 | 2.0085 | 1994.06 |

When the core papers do not belong in the cluster, the results show "0". When the core papers do not belong in the largest component in generating the citation network, the results show "–".

Hopcroft et al. (2004) and Shibata et al. (2009), and therefore co-citation is not suitable for research front detection. In addition, regarding "(A-2) NAKAMURA S, 1992, JPN J APPL PHYS PT 1, V31, P1258," "(B-1) Watts DJ, 1998, NATURE, V393, P440," "(B-2) Barabasi AL, 1999, SCIENCE, V286, P509," and "(C) IIJIMA, S, 1991, NATURE, V354, P56", the core papers were not involved in the largest component in the birth year.

*Bibliographic couplings*. In this citation pattern, the results of comparisons between the weights are almost the same results as the co-citations. The bibliographic coupling could be expected to be better than the co-citation because it could potentially detect more edges earlier than the co-citations. However, the results of bibliographic coupling are slightly worse than direct citation when introducing the weighted citation networks. In addition, in the case of "(B-2) Watts DJ, 1998, NATURE, V393, P440," the core paper was not involved in the largest component. Core papers with the bibliographic coupling cannot generate more clusters than others in introducing weight (ii). This is because that the citation network with bibliographic coupling is concentrated on the specific famous papers regardless of its contents.

*Differences between domains*. In this paper, we evaluate some methodologies to three domains (GaN, CNW, and CNT). They are typical examples of recent remarkable innovations having somewhat different characteristics in focusing on the way of the breakthroughs of the rapid developments. The results of sizes, average publication year, density, and textual similarity among in three domains are almost same. On the other hand, there are some characteristic differences among three domains. Comparing GaN with CNW, the results of GaN are the larger and denser than the ones of CNW, however, the textual similarity of CNW is better than the one of GaN. Comparing GaN with CNT, the results of GaN are the larger and denser than the ones of CNT, however, the textual similarity of CNT is better than the one of GaN. The main reasons of these results are the breakthroughs of the rapid developments. Fuller study for the generalizability connecting the methodologies and domains lies outside the scope of this paper. On the other hand, we conducted the comparative studies of effectiveness among some weighted citation networks in many kinds of characteristic domains.

## Discussions

A summary of comparisons of the results is shown in Table 7. Weight (i) in co-citation and bibliographic coupling has better textual similarity and higher density clusters compared with weight (i) in the direct citation. This means that the frequency of citations is effective for generating the textual and topologically relevant clusters. Weight (ii) generates small-sized clusters, and is the worst in the speed to detect research fronts. By introducing the difference in publication years for placing the weights to new edges, we can detect the small clusters formed by the important academic papers only in the early stage. However, it is not in the very early stage as Table 7 shows. In addition, the core-papers acting as the important aspects in the academic fields could be missed in the expansion stage because the weights of edges between a paper and the core-paper become weaker. Theoretically, weight (ii) is expected to efficiently detect research fronts because it gives a large weight on the link where both citing and cited papers are young, however, it does not work well actually. Core papers cannot generate the clusters when the year has passed since the core paper was published.

Weights (iii) and (iv) have almost the same tendency compared with the others. The reason for this is that both the reference similarities and the keyword similarities represent the contents of papers. On the other hand, when the topologically relevant clusters are required, weight (iii) is especially

**Table 7**
Brief result of comparison of four types of weights.

| | Visibility | Topological relevance | Text relevance | Speed |
|---|---|---|---|---|
| Direct citation | (i)>(ii)=(iii)=(iv) | (iii)>(ii)=(iv)>(i) | (iv)>(ii)=(iii)>(i) | (i)>(ii)=(iii)=(iv) |
| Co-citation | (i)=(iii)=(iv)>(ii) | (iii)>(iv)=(i)>(ii) | (iv)>(i)=(iii)>(ii) | (i)=(iii)>(iv)>(ii) |
| Bibliographic coupling | (i)=(ii)=(iii)=(iv) | (iii)>(i)=(ii)=(iv) | (iv)>(i)>(ii)=(iii) | (i)=(iii)=(iv)>(ii) |

*Note*. (i) Frequency of citations; (ii) difference of publication years; (iii) reference similarity; (iv) keyword similarity.

effective. When the textually relevant clusters are required, weight (iv) is especially effective. In fact, both the references and keywords of academic papers are important information for judging the academic area and contents of academic papers. On the other hand, the discussions of relevance of clustering between citation information and keyword information for detecting the emerging front have been conducted for a long time in the bibliometric field (Chen, 2006; Kostoff et al., 2001; Losiewicz et al., 2000).

We present a comparative study to investigate the performance of methods for detecting emerging research fronts among weighted citation networks, which include the frequency of citations, the difference of publication years, similarities of the references and keywords. In addition, we evaluated the weighted citation networks to some kinds of citation networks (Direct citation, Co-citation, Bibliographic coupling). As a result of the comparative studies in the previous section, we could produce two conclusive results.

First, we showed that the citation networks with the weights are more useful in detecting the characteristic research fronts (emerging fronts) than those without the weights. Weighted citation networks can capture important information attributes of papers compared with non-weighted citation networks. In the comparative studies, the frequency of citations in co-citation and bibliographic coupling has better textual similarity and higher density clusters compared with the non-weighted method in the direct citation. However, most of the existing works focus on the non-weighted citation networks. By this comparative studies, we can expand the capability of the citation network analysis for emerging detections by introducing and evaluating the weights, practically and scholarly.

Next, we showed the relationships between the types of weights and topological measures for evaluating the emerging fronts. We demonstrated that different weight strategy is effective for different purpose. When the topologically relevant clusters are required, weight of the reference similarity is especially effective. When the textually relevant clusters are required, weight of the textual similarity is especially effective. Based on these results, we can select the weight strategies based on the characteristics of the task of users. When weighting strategy to adjust the purpose is introduced, users can detect the research fronts effectively and/or reliably.

In all of the above cases, earlier and more accurate detection of research fronts is essential information for both researchers and research and development (R&D) managers in planning their research focus and strategy. For applying the weighted networks to detect the research fronts for managerial purpose, selecting which weights to use by adjusting the purpose of analysis is important. We cannot select the best method without purpose. Through the comparative studies in this paper, the characteristics of every weight become clear. Therefore, it is important to select the effective weights based on the definition of emerging fronts or purpose of analysis. Typically, the frequency of citations is effective for detecting the emerging fronts, which are defined as large, young, and relevant clusters when the co-citations and bibliographic coupling are used. On the other hand, the network with co-citations has time-lag issues, and the bibliographic coupling was slightly worse than direct citations in the non-weighted citation network analysis.

It is plausible that co-citation and bibliographic coupling were slightly worse because of the process of creating networks. In our method, only the citations among papers within the dataset that were collected by the queries were used to create each citation network. By performing an additional examination to create a network using the data of additional "one path" from the collected papers, the results of co-citation and bibliographic coupling networks show superior performance by considering this additional link (Boyack and Klavans, 2010). We cannot decline a possibility that the co-citation and bibliographic coupling weighted networks can extract better emerging clusters. It is the one of the important and possible future directions in more detailed comparative studies.

Another important aspect of effective weighted citation network analysis is the way to decide the mathematical representations in the functions of weights. In this paper, the functions of weights are decided based on the ad-hoc heuristics for gallium nitride (GaN), complex network (CNW), and carbon nanotube (CNT) domains. However, the automatic identifications of optimal functions of weights are necessary for effective analysis. By introducing machine learning technology, we could achieve them. Fuller study of this topic lies outside the scope of this paper.

## Conclusions

This paper presents a comparative study to investigate the performance of methods for detecting emerging research fronts among weighted citation networks, which include the frequency of citations, the difference of publication years, similarities of the references and keywords. A case study was performed in three research domains, gallium nitride, complex networks, and carbon nanotubes. After some types of weighted citation networks were constructed, papers in each research domain were divided into clusters using topological clustering. We evaluated the visibility, speed, and topological and textual relevance of the clusters to which selected core papers belonged.

By using the weight based on the frequency of citations, young and dense clusters are detected. By using the weight based on the difference of publication years, clustering techniques generate small clusters. Using the weight based on keywords and reference information has almost the same tendency. In addition, the weight with the references shows more topologically relevant contents than that with author keywords, and the weight based on paper keywords shows more textually relevant contents than that with author keywords as expected.

## References


Braam, R.R., Moed, H.F., Raan, A.F.J., 1991a. Mapping of science by combined co-citation and word analysis, I. Structural aspects. Journal of the American Society for Information Science 42 (4) 233–251.

Barabasi, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439) 509–512.

Boyack, K.W., Klavans, R., 2010. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? Journal of the American Society for Information Science and Technology 61 (12) 2389–2404.

Boyack, K.W., Klavans, R., Börner, K., 2005. Mapping the backbone of science. Scientometrics 4 (3) 351–374.

Braam, R.R., Moed, H.F., Raan, A.F.J. van, 1991b. Mapping of science by combined co-citation and word analysis. I. Structural aspects. Journal of the American Society for Information Science 42, 233–251.

Chen, C., 1999. Visualizing semantics paces and author co-citation networks in digital libraries. Information Processing and Management 35 (2) 401–420.

Chen, C., Cribbin, T., Macredie, R., Morar, S., 2003. Visualizing and tracking the growth of competing paradigms: two case studies. Journal of the American Society for Information Science and Technology 53, 678–689.

Chen, C., 2006. 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) 359–377.

Davidson, G.S., Hendrickson, B., Johnson, D.K., Meyers, C.E., Wylie, B.N., 1998. Knowledge mining with VxInsight: discovery through interaction. Journal of Intelligent Information Systems 11, 259–285.

Hopcroft, J., Khan, O., Kulis, B., Selman, B., 2004. Tracking evolving communities in large linked networks. Proceedings of the National Academy of Sciences 101 (Suppl. 1) 5249–5253.

Iijima, S., 1991. Helical microtubules of graphitic carbon. Nature 354, 56–58.

Jaccard, P., 1912. The distribution of the flora in the alpine zone. New Phytologist 11 (2) 37–50.

Jarneving, B., 2007. Bibliographic coupling and its application to research-front and other core documents. Journal of Informatics 1 (4) 287–307.

Jaina, A., Nandakumara, K., Ross, A., 2005. Score normalization in multimodal biometric systems. Pattern Recognition 38 (12) 2270–2285.

Kessler, M.M., 1963. Bibliographic coupling between scientific papers. American Documentation 14, 10–25.

Klavans, R., Boyack, K.W., 2009. Toward a consensus map of science. Journal of the American Society for Information Science and Technology 60 (3) 455–476.

Klavans, R., Boyack, K.W., 2006. Identifying a better measure of relatedness for mapping science. Journal of the American Society for Information Science and Technology 57, 251–263.

Kostoff, R.N., Eberhart, H.J., Toothman, D.R., 1997. Database tomography for information retrieval. Journal of Information Science 23, 301–311.

Kostoff, R.N., del Río, J.A., Humenik, J.A., García, E.O., Ramírez, A.M., 2001. Citation mining: integrating text mining and bibliometrics for research user profiling. Journal of the American Society for Information Science and Technology 52, 1148–1156.

Leydesdorff, L., 2004. Clusters and maps of science journals based on bi-connected graphs in. Journal of Documentation 60 (4) 371–427.

Leydesdorff, L., Rafols, I., 2009. A global map of science based on the ISI subject categories. Journal of the American Society for Information Science and Technology 60 (2) 348–362.

Losiewicz, P., Oard, D.W., Kostoff, R.N., 2000. Textual data mining to support science and technology management. Journal of Intelligent Information Systems 15 (2) 99–119.

Nakamura, S., 1991. GaN growth using GaN buffer layer. Japanese Journal of Applied Physics 30 (10A) 1705–1707.

Nakamura, S., Mukai, T., Senoh, M., 1994. Candela-class high-brightness InGaN/AlGaN double-heterostructure blue-light-emitting diodes. Applied Physics Letters 64 (13) 1687–1689.

Nakamura, S., Mukai, T., Senoh, M., 1992. Si- and Ge-doped GaN films grown with GaN buffer layers. Japanese Journal of Applied Physics 31 (9R) 2883.

Newman, M.E.J., 2004. Fast algorithm for detecting community structure in networks. Physical Review E 69, 066133.

Price, D.J., 1965. Networks of scientific papers. Science 149, 510–515.
Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K., 2008. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. Technovation 28 (11) 758–775.
Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K., 2009. Comparative study on methods of detecting research fronts using different types of citation. Journal of the American Society for Information Science and Technology 60 (3) 571–580.
Shibata, N., Kajikawa, Y., Sakata, I., 2011. Measuring relatedness between communities in a citation network. Journal of the American Society for Information Science and Technology 62 (7) 1360–1369.
Small, H., 1999. Visualizing science by citation mapping. Journal of the American Society for Information Science 50 (9) 799–813.
Small, H., 2006. Tracking and predicting growth areas in science. Scientometrics 68 (3) 595–610.
Small, H., 1973. Co-citation in the scientific literature: a new measure of the relationship between two documents. Journal of the American Society for Information Science 24, 265–269.
Small, H.G., Griffith, B.C., 1974. The structure of scientific literatures. I. identifying and graphing specialties. Science Studies 4, 17–40.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of "small-world" networks. Nature 393, 440–442.