JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Iwami, Shino; Ojala, Arto; Watanabe, Chihiro; Neittaanmäki, Pekka

Title: A bibliometric approach to finding fields that co-evolved with information technology

Year: 2020

Version: Published version

Copyright: © 2019 The Authors

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Iwami, S., Ojala, A., Watanabe, C., & Neittaanmäki, P. (2020). A bibliometric approach to finding fields that co-evolved with information technology. Scientometrics, 122(1), 3-21. https://doi.org/10.1007/s11192-019-03284-9

Scientometrics
https://doi.org/10.1007/s11192-019-03284-9

A bibliometric approach to finding fields that co-evolved with information technology

Shino Iwami[1,2] · Arto Ojala[2] · Chihiro Watanabe[2,3] · Pekka Neittaanmäki[2]

Received: 23 November 2017
© The Author(s) 2019

Abstract
Among the declining industries, for example music industry, some have been revived by information technology (IT). At the same time, in academic fields, some have expected co-evolutions between IT and other fields to cause the resurgence of either field. In this research, the clustering of citation networks with 14,438 academic papers resulted in the identification of 28 academic fields in the areas "Computer Science" or "Information Science and Library Science." Co-evolutions between these 28 fields and citing fields to the 28 fields were evaluated by an investigation of contents; a methodology to search co-evolutions was also proposed. This paper proposes that pairs of academic fields (with both high correlation and high dissimilarity) co-evolve, and some co-evolving pairs of academic fields were found. This research contributes to the discovery of the co-evolution between academic fields.

Keywords Co-evolution · Academic landscape · Big data · Network analysis · Citation analysis · Clustering · Unsupervised classification · Cosine similarity · Horizon scanning · Exploratory scanning

Introduction

With the advancement of information technology (IT), new industries have been developed, and old industries have been replaced. For example, the music industry has shifted from physical music recordings to downloading or streaming; this was due to IT advancements. In the music industry, sales of old media have declined. Furthermore, users have begun to participate in new attitudes toward music activities such as streaming music and crowdsourcing by using IT; sales related to live music have started to rise again after declining (IFPI 2017; Naveed et al. 2017). In the music industry, the music rankings like Billboard began in America at the end of the 19th century, and in 1990 it became the ranking of

✉ Shino Iwami
iwami@ipr-ctr.t.u-tokyo.ac.jp

1 Osaka University, Suita, Japan

2 Faculty of Information Technology, University of Jyväskylä, P.O. Box 35 (Agora), 40014 Jyvaskyla, Finland

3 International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Published online: 09 November 2019

sales based on the number of plays (Broven 2009). Also, in 1979, Sony in Japan launched Walkman with the concept of "carrying music and feel free to enjoy." At that time, it was music mobile with cassette tape, but now products evolved into digital products such as iPod (Haire 2009).

The evolution of IT also has revived the gaming, printing, and movie industries through new types of digital platforms that have enabled the delivery of content to customers through various pieces of digital equipment. For instance, Ojala (2016) demonstrated how the co-evolution of IT and the gaming industry has enabled new ways of bringing video games to the market—developing digital gaming platforms that radically change traditional business models in the industry. In the field of robotics, which is related to IT, robot manipulation of electronic engineering was mainstream in the beginning of the 21st century. However, in BigDog (Boston Dynamics 2018) and surgery robot (Tadano and Kawashima 2006), traditional mechanical engineering technologies (such as hydraulic pressure and air pressure) are being used to improve small robot manipulation. In an interview about soft robots, which one of the authors attended in 2015 (Center for Research and Development Strategy 2016), a story of experience that traditional technologies overcame the issue of small robot manipulations was obtained. In these ways, the resurgence of a declining industry can be observed due to the co-evolution of IT with that industry.

Therefore, by discovering co-evolution and resurgence of science, this research intends to contribute to industrial redevelopment. The purpose of this research is to establish a methodology for finding academic fields that have co-evolved with other fields by the means of bibliometric methodology. In this research, information technology is selected as the comprehensively explored scientific field. Although previous research (Naveed et al. 2017) showed a resurgence of the music industry by co-evolution in well-known pairs of music and IT on the basis of revenue data, this research aims to find unknown field combinations with co-evolutions and resurgences.

Relevant literature

Regarding bibliometric methodologies, Garfield (1955) began the use of citation analysis (such as journal impact factors) to evaluate research in the fields of bibliometrics and scientometrics. The Web of Science provided by Clarivate Analytics took over the Science Citation Index (SCI), which was produced by the Institute for Scientific Information (ISI); it provides most major indicators such as times cited and impact factor among researchers. The number of scientific publications in each field is increasing, and experts need to roughly grasp their fields. Conventionally, experts have struggled to read all scientific publications to grasp their fields, but now it becomes a difficult way. Therefore, bibliometrics and scientometrics have provided methodologies for overviewing academic fields and identifying remarkable research and researchers (Börner et al. 2003; Leydesdorff and Rafols 2009; Rafols et al. 2010; Iwami et al. 2014). Some functions are provided as tools for performing complex analyses easily with web interfaces (Kajikawa et al. 2007; Innovation Policy Research Center 2013). These methodologies and tools have been developed to contribute to evidence-based policy, foresight-based policy, and horizon scanning, which are desired by governments (U.S. Department of Health and Human Services 2016; European Commission 2016; European Commission 2017).

In a previous study, which applied bibliometrics to co-evolution, Murray (2002) explored the co-evolution of science and technology by using the citation analysis of patents and academic papers in tissue engineering. She showed that co-evolution and

spillovers can be measured by the cross-citations of papers in patents, simultaneous publication and patenting by companies, and co-publishing across academic and business institutions. Kimmerle et al. (2010) visualized the co-evolution of information in Wikipedia with link analysis. Co-evolution was identified by using the co-occurrence of IPC (International Patent Classification) codes and keywords in a patent on aerospace; this was done by van der Pol and Rameshkoumar (2018). In the way of using similarity of sentences, the methodologies to match social issues and scientific solutions for solving social issues (Ittipanuvat et al. 2014), to match fields of a domain and strong fields of a country for increasing country's strengths (Iwami et al. 2015b), and to match cutting-edge fields and all scientific fields for developing new collaborations (Iwami et al. 2015a).

While citations, co-occurrences of keywords, collaborations between authors or institutions by co-authoring are available for the measure of co-evolution, this research focuses on co-evolution between sciences (which are industrial seeds) with citations. To identify resurgence, the rise and fall of the field time-series chart was used.

Significance

In the recent years, one mission of bibliometrics or scientometrics becomes to provide evidence for decision making about science, technology, research, development, and innovation management (U.S. Department of Health and Human Services 2016; European Commission 2016; European Commission 2017). This research tries to expand the utilization of bibliometric or scientometric approaches. Traditionally, experts such as researchers or funders have sought for investment targets by the means of qualitative investigation like interview, workshop and literature review. Meanwhile, it is difficult for experts to catch up with everything even in their specialty because a large amount of knowledge has been stored every year. Thus, experts need to be supported by the computational methodologies in order to overview their specialty and capture the cutting-edge science and technology. Because the principles and guidelines for evidence-based policy-making have been enacted (European Commission 2002) and data science has been required in technology and innovation management, the demand for evidence for policy and management is high.

This research explores co-evolved fields with IT fields in a half-open attitude, which means that the analysis does not restrict the number of fields on one side of the pairs of co-evolved fields though the fields on the other side is fixed in the IT fields. A previous study (Naveed et al. 2017) has limited the fields to those within the author's knowledge; this limitation is rational as it removes noise and completes analyses. However, finding unexpected pairs is an advantage of big data analysis, and unexpected pairs will be able to further stimulate industrial development. This research takes the risk of not finding good co-evolved pairs, but exploratory-seeking good co-evolved pairs in all fields is worth a challenge.

As shown in Fig. 1, this research uses three types of relations: citation, correlation, and dissimilarity. Citations are consciously connected by authors, indicating causality. Unlike qualitative analysis (such as interviews), citation analysis shows co-evolutions objectively, since causality is strengthened by many citations. Co-evolution requires two fields to be growing simultaneously. Resurgence is judged by a rise after a fall. Thus, correlations measure concurrency of rises and falls of timelines. Dissimilarities indicate that two fields are different.

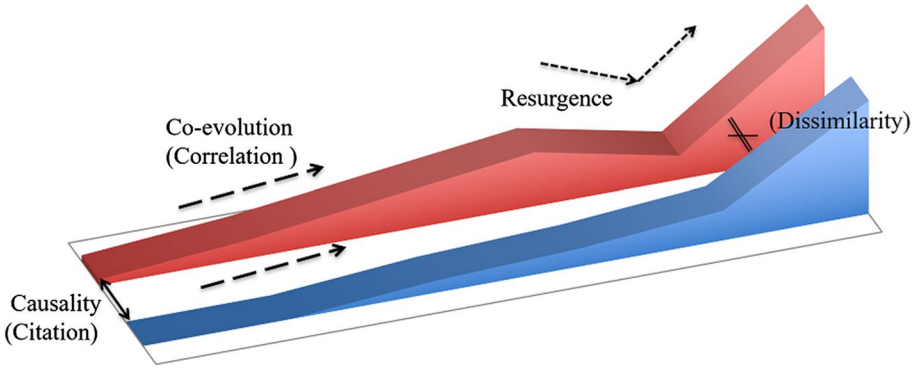


Fig. 1 The concept of co-evolution and resurgence with citation, dissimilarity, and correlation

The rest of this article is organized as follows: "Methodology" section describes the steps of this research and the technological factors. "Results and discussion" section indicates the results and provides a discussion. "Conclusion" section summarizes this research.

Methodology

As shown in Fig. 2, this research was performed in the following steps:

1. Fields of the top 1% of academic papers in "Computer Science" or "Information Science and Library Science" were identified by the means of clustering citation analysis.
2. Fields of the academic papers citing the top 1% of academic papers were identified through citation analysis.

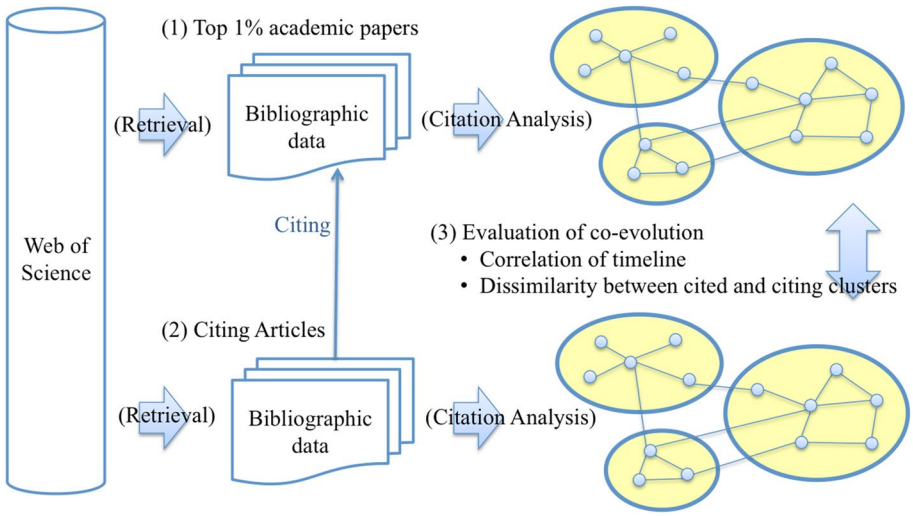


Fig. 2 Steps of this research

3. Co-evolutions between citing fields and cited fields were evaluated based on dissimilarity and correlation.

Data

To identify the IT fields, the bibliographic data of 14,438 academic papers were retrieved from the Web of Science; the data were provided by Clarivate Analytics (as of October 10, 2017) and are shown in Fig. 3. The academic papers were those of the top 1% of academic papers—based on their number of citations in the research areas of "Computer Science" or "Information Science and Library Science." Here, the top 1% of academic papers was defined as the sum of each top 1% of academic papers in each year since 1960. This was because older papers obtain more citations; thus, only old papers would have been included in the top 1% of academic papers for the entire period if this method had not been used. Table 1 in the "Appendix" shows the data of the top 1% of academic papers and the citing academic papers.

On the Web of Science, citing academic papers is called "Citing Articles." Citing Articles (CAs) is retrieved based on citations to the top 1% of academic papers, as shown in Fig. 4. In Fig. 4, squares represent datasets, and blue circles indicate clusters. While the methodology (Iwami 2017) was proposed to retrieve CAs after cited articles were analyzed, the retrieval of CAs' bibliographic data was not performed for the second layer of CAs (CAs of CAs) in this research. Consequently, 1,512,545 academic papers were retrieved.

Citation analysis

Types of citation

Academic papers have citations, so networks can be constructed by connecting citations. There are three types of citations: direct citations, co-citations, and bibliographic couplings, as shown in Fig. 5 (Fujita et al. 2012).

In order to build a network of the top 1% academic papers, direct citations, co-citations and bibliographic couplings were used. The reason for using these three types of citations simultaneously was to compensate for the number of direct citations of only 1% of academic papers.

However, in order to build networks of CAs, direct citation and co-citation were used. The types of citations were selected to complete the analyses in a reasonable computation time, incorporating as many citations as possible, depending on the server capacity. The number

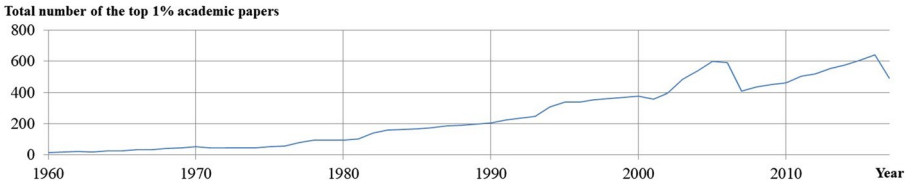


Fig. 3 The trend in the top 1% of academic papers in "Computer Science" or "Information Science and Library Science"

of CAs became too large to be computed as a result of the actual collection of CAs' bibliographic data. For example, when a small size of CAs (including around 10,000 papers) was computed actually to know the scale, the number of edges according to direct citation and co-citation was about five times greater than the number of direct citations; the number of edges according to the three types of citations was about 100 times greater than the number of direct citations and co-citations. The gap of 100 times the number of edges led to three times the computation time and one-fifth of the number of clusters. Thus, the usage of bibliographic coupling (based on papers citing CAs) was discarded, and direct citation and co-citation were used for CAs, as shown in Table 1 in the "Appendix". Since there have been no previous cases in which over 300,000 records were analyzed with three types of citations, the CAs were analyzed without bibliographic coupling to avoid a server crash and irrational time consumptions.

Clustering

The citation network was divided into clusters by using the academic landscape system (Kajikawa et al. 2007; Innovation Policy Research Center 2013; Kajikawa et al. 2014) with an unsupervised classification, the Newman method (Newman 2004). The system was built based on the methodology of Kajikawa et al. (2007), and it was implemented on a website (Innovation Policy Research Center 2013). The utilization of citations has the merit that the issue of homonyms and synonyms is solved in the form of citations, which use the collective knowledge of experts.

After that, the topics of the clusters are manually identified with keywords, titles, abstracts, and journal names. The merit of clustering of citation networks is that analysts do not need inclusive knowledge of the field's topics. Therefore, classification can be executed before analysts understand the contents of large amounts of data. The merit of the Newman method (Newman 2004) is that the number of fields is decided automatically by computation in order to maximize modularity; Newman defined "modularity" as an indicator that occur at about 0.3 or more in communities, for example citation network. In the actual utilization of results after clustering (since several minor clusters with less than 10 papers are generated), the tool's users determine the threshold of the cluster nodes for the adopted clusters—depending on the purpose.

Dissimilarity

The academic landscape system provides a feature keyword list of respective clusters, which can be computed by the means of a mutant "TF–ICF (Term Frequency–Inverse Cluster

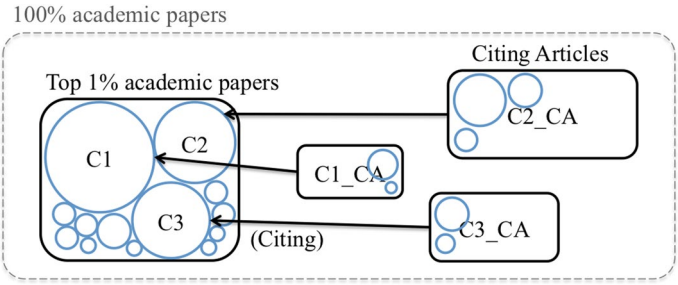


Fig. 4 Relations between the top 1% of papers and CAs

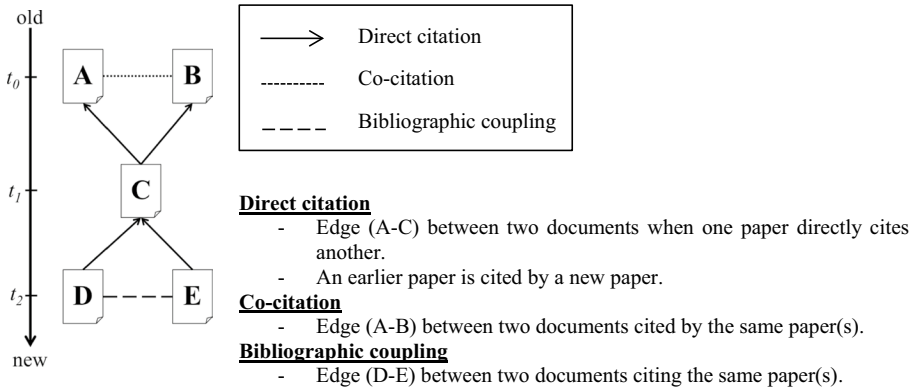


Fig. 5 Three types of citations

Frequency)" of TF–IDF (Term Frequency–Inverse Document Frequency). The equation is the following:

$$\begin{aligned}\mathrm{TFICF} &= \mathrm{TF} \times \mathrm{ICF} \\ &= \frac{[\text{Number of occurences of one word in one cluster}]}{[\text{Number of occurences of all words in one cluster}]} \\ &\quad \times \log \frac{[\text{Number of all clusters}]}{[\text{Number of clusters for one word}]}\end{aligned}$$

TF is a score that shows that high-frequency words are important. ICF in a cluster and IDF in a document are scores that show that high-frequency words in other clusters and documents are not important. ICF uses a comparison of words between clusters in a dataset (instead of a comparison of words between documents). Here, clusters mean one type of academic fields, which are different from the definition by Web of Science. The reason to use ICF instead of IDF is to target the mass such as clusters for observing the rise and fall of each specialty. The bibliographic data of some academic papers do not include abstract, though they have semantic connections via citations. Old academic papers do not include abstract in its bibliographic data. These academic papers without abstract cannot contribute to this analysis when IDF analyze individual academic papers. However, ICF enables academic papers without abstract to be counted in time-series trends. Figure 6 shows the concept of TF–ICF, which is compared with TF–IDF. With the keywords based on scores of TF–ICF, the computation to identify topics of clusters is provided.

Here, dissimilarity is defined as a reciprocal of cosine similarity. Cosine similarity is based on the Ochiai index or the Ochiai coefficient, which have both been suggested for measuring the geographical distribution of fish (Ochiai 1957). The cosine similarity between a cluster A and another cluster B was computed as shown in Eq. (1) with a vector $\boldsymbol{x}$ of the top 30 keywords in A and a vector $\boldsymbol{y}$ of the top 30 keywords in B. The scores of TF–ICF were used to determine the top 30 keywords.

$$\frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\| \, \|\boldsymbol{y}\|} \tag{1}$$

The scores of similarities were relative values in the dataset because word diversity in each field was different. The author's source code of similarity was provided in https://github.com/shinoiwami/a_JYU2017_06/blob/master/110_keyword_similarity.py.

Correlation

The numbers of the top 1% of papers can be counted in each cluster according to each year. Correlations between citing papers and cited papers can be computed with the time-series numbers. The equation of correlation from the Microsoft Excel function is expressed as the follow:

$$\mathrm{Correl}(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Pairs of (x, y) are composed of number of cited papers and citing papers in the same year.

Dissimilarity enabled the exclusion of the same fields between citing fields and cited fields for the purpose of finding co-evolutions. Thus, mappings of dissimilarity and correlation give the following interpretations:

- High dissimilarity and strong correlation: co-evolution
- Low dissimilarity and strong correlation: the same fields between citing fields and cited fields
- Weak correlation: non-co-evolution.

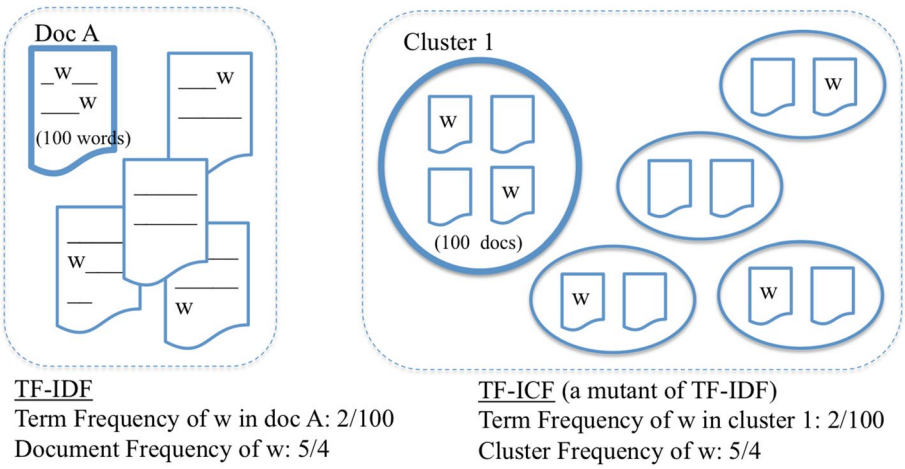


Fig. 6 The concept of TF–ICF compared with TF–ICF

Results and discussion

Academic fields in computer science and information technology

Figure 7 illustrates the fields of the top 1% of academic papers based on the number of citations in the research areas of "Computer Science" and "Information Science and Library Science." However, clusters with less than 100 papers were excluded. The average published year of the horizontal axis represents the average publication years in each field, and the circle size is associated with the number of academic papers. Academic fields were manually categorized into the network, IT system, applied science, and fundamental theory. These four categories roughly align from implementation to theory on the vertical axis. Circle density indicates the rate of academic papers related to the Internet—so the internet-applied fields have expanded (like the red background). Numbers for the average published year, circle size, and circle density are tabulated in Table 2 in the "Appendix". This research explored multiple fields to find clusters that matched the theory. However, because analysis using citations requires a huge amount of data, this research focused on the amount of data in the top 1%. In the later years, when excellent academic papers are generated or new fields appear, the boundaries of clusters will change (Shibata et al. 2008, 2011). When someone uses this methodology in practice, the person will identify and analyze only declining fields that will be the remaining fields after the cause of resurgence becomes independent. Therefore, the person can use 100% articles and other databases instead of the top 1%.

The IT fields on the network band do not necessarily include the word "internet," but general-purpose technologies related to local area network (LAN) and wide area network (WAN) have appeared. The recent IT fields shown in Fig. 7 have been used for business—even in applied science and fundamental theory, such as "genome analysis" (McCarthy 2012; Google 2014), "fuzzy control" (Chiou et al. 2009), "motion capture" (Vlasic et al. 2007), "face recognition" (Abate et al. 2007), and "classification,

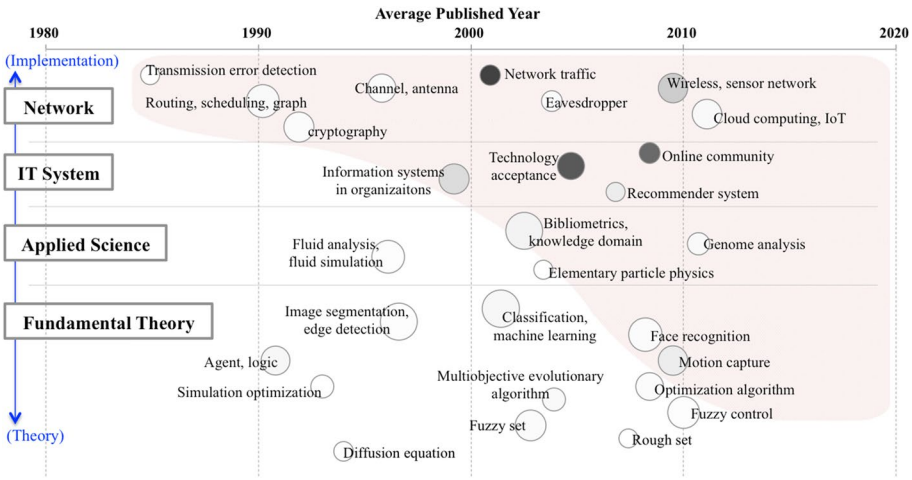


Fig. 7 IT-related fields based on the top 1% of academic papers

machine learning." The IT fields related to "optimization" and "fuzzy" appear twice before and after 2005.

This research focused on resurgences within co-evolutions. Since "fuzzy set" and "simulation optimization" had a closely related field of "optimization algorithm" and "fuzzy control" in the later years, we confirmed that there was resurgence by creating a graph of time series data of "fuzzy set" and "simulation optimization" and their CAs, similar to Figs. 8, 9 and 10. Thus, "fuzzy set" and "simulation optimization" were selected to subdivide the fields into subfields and identify the contents in "Case 1: fuzzy set" and "Case 2: simulation optimization" sections; this was due to the existence of similar fields in the top 1% of academic papers. Additionally, "channel, antenna" was chosen because its timeline presented a small resurgence in the number of academic papers in "Case 3: channel, antenna" section. "A proposed methodology to judge co-evolutions" section provides an overview of how to identify co-evolution across multiple fields including cases.

Case 1: fuzzy set

In Fig. 7, if the IT field "fuzzy set" had subsequent IT fields of "rough set" and "fuzzy control," it was expected to have resurgence in addition to co-evolutions with citing fields. According to Fig. 8a, the cited IT field had two peaks (in 1996 and 2016), and the latter peak followed the increases of CA3, CA4, and CA6. CA2 started to grow with the cited IT field.

As shown in Fig. 8b, CA1–CA10 are divided into two groups by dissimilarity: fields related to decision-making (CA1–CA3, CA5–CA7) and medical fields (CA8 and CA9). The former group had a lower dissimilarity (0.531–0.764) than the latter group (1.000), so the former group contained the successors of the cited IT field. The latter addressed blood vessels in medicine and the applications of the cited IT field. CA10 was unique, belonging to computer science and mathematics, but having high dissimilarity (1.000).

Thus, CA3, CA5, CA6, and CA8 highly co-evolved with the IT field; in particular, a co-evolution between the cited IT field and CA8 fits the background of heterogeneous co-evolutions like publications about fuzzy set applied to images of vessel tracking and segmentation (Grisan et al. 2004; Ayala et al. 2005). Additionally, CA3, CA5, and CA6 contributed to the resurgence of the cited IT field.

Case 2: simulation optimization

The IT field of "simulation optimization" has latent co-evolutions—though it has low correlations with CAs. At step 1 of the methodology, the "simulation optimization" has a high potential for resurgence; this is because its average published year is 1993.0, and the "optimization algorithm" appears in the latter half of the 2010s. Both "simulation optimization" and "optimization algorithm" include the word "optimization." Figure 9a illustrates trends in the cited IT field of "simulation optimization" and the citing fields as CAs. The peak in the 2000s co-occurred with the trend in CA6 of "model order reduction," and the peak in 1990 matched the trend in CA9 "electron nucleus" (Huang et al. 1990).

According to Fig. 9b, CA6 and CA9 have weak correlations (0.200 and 0.219), which were relatively higher than the others and had high dissimilarities (0.936 and 1.000) with the IT field "simulation optimization." Thus, CA6 and CA9 relatively co-evolved with the

cited IT field within a dataset in Fig. 9b, but the score does not indicate co-evolution in Fig. 11. This is a limitation due to differences in the data. Apart from the other cases, the IT field "simulation optimization" was cited by non-IT fields (such as particle physics and chemistry), and it included interdisciplinary collaborations. It was considered that interdisciplinary collaboration has a high potential for heterogeneous co-evolution, but correlations became relatively low.

Case 3: channel, antenna

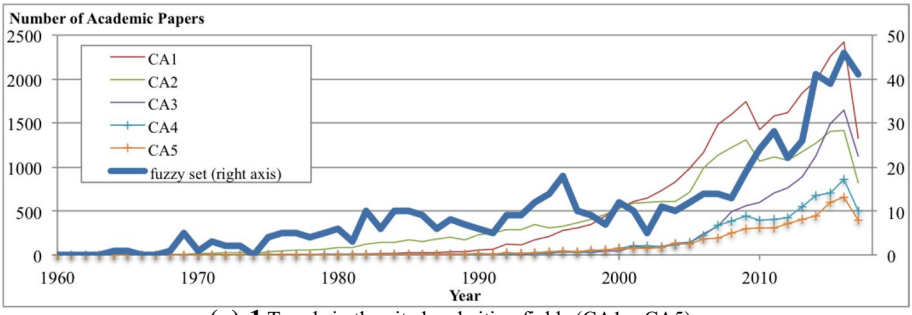
The IT field of "channel, antenna" co-evolved with other fields. According to Fig. 10a, "channel, antenna" had peaks in 2003 and 2005, and CA8, CA10 and CA11 seem like to associate with IT field. As shown in Fig. 10b, CA1 "space time" and CA2 "multiple-input multiple-output" were direct successors because of the lowest dissimilarity (0.413 and 0.407) among CAs; in other words, they have the highest similarity and were almost the same fields of the cited IT field "channel, antenna." Consequently, the co-evolving CAs with high dissimilarity and high correlation were CA7, CA8, CA9, CA10 and CA11, which belonged to the field of IT.

Observing the history of antenna, Yagi antenna was invented in 1926 (Yagi and Uda 1926; Wikipedia 2018), and it has reduced following the end of analog broadcast. The parabolic antenna is for receiving radio waves from the satellite regardless of analog broadcasting and digital broadcasting (Tuli et al. 2006). Personal Handy-phone System (PHS) has provided to the average people since 1990s, and many countries finished the service in 2010s (NTT DoCoMo Inc. 2014). The whip antenna of PHS was outside of the device, and its form was taken over to a mobile phone. For these wireless devices, low power was an important factor. The research in CA8 (low power) increased since 1999, and it reached the peak in 2005–2006. The 3rd Generation (3G) service enabled videophone in 2001 (Charny 2001), while number of publications in CA10 (video) has a peak in 2004. Smartphones with an invisible antenna started to be sold in 2004, and iPhone since 2008 made smartphones a kind of daily necessities. Multipliers have been studied as shown in CA11 of Fig. 10a, and the efficiency of the circuit including antenna was improved. In this way, tracing the history of the devices involved in the antenna confirms that CA8, CA10 and CA11 co-evolved for the resurgence of the antenna.

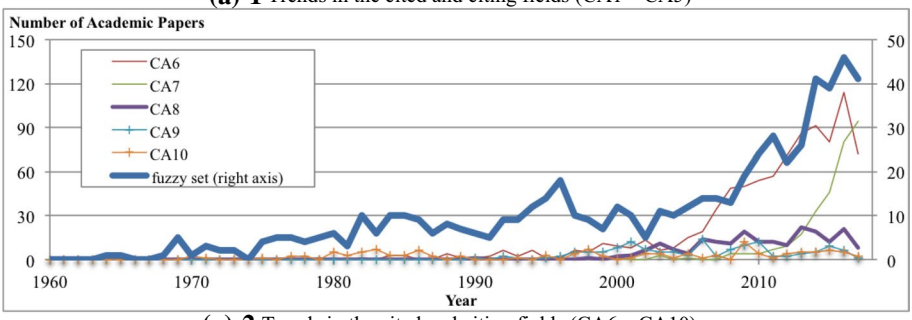
A proposed methodology to judge co-evolutions

In the case of CA4 in the "Case 1: fuzzy set" section, CA1 in the "Case 2: simulation optimization" section, and CA1 and CA2 in the "Case 3: channel, antenna" section, the CAs that have strong correlations with their citing IT fields mean the same fields with the citing fields—but not the co-evolved fields. Some cases in this research had successful pairs of co-evolution, but most of the pairs were not co-evolved. In Fig. 11, clusters until the average year 2002 (as shown in Table 1 in the "Appendix") are plotted, and the mapping indicates relations between dissimilarity and correlation. However, CAs of C1-1, C1-2, and C2-1 were excluded due to memory shortage and the failure of accesses to the Web of Science.

In summary, co-evolutions were indicated as shown in the upper-right and lower-right part of Fig. 11. The thresholds were decided based on the relations in the set, so they

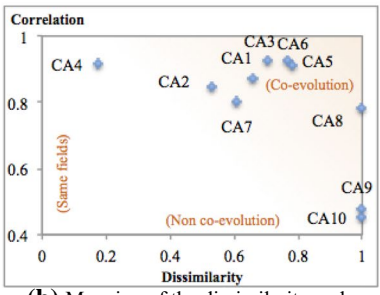


(a)-1 Trends in the cited and citing fields (CA1 – CA5)



(a)-2 Trends in the cited and citing fields (CA6 – CA10)

[Fields citing to the "fuzzy set"]
CA1: clustering
CA2: fuzzy number
CA3: decision making
CA4: intuitionistic fuzzy
CA5: cognitive map
CA6: metric space
CA7: neutrosophic set
CA8: vessel tracking and segmentation
CA9: atrial fibrillation
CA10: nonlinear system
(*Bold underlined fields co-evolved with the cited IT field.)



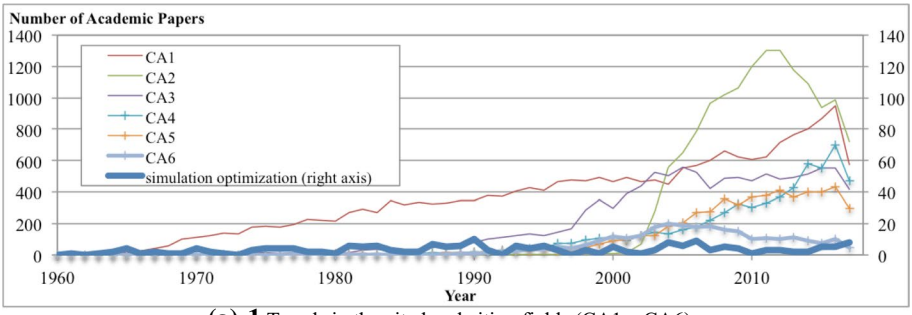
(b) Mapping of the dissimilarity and correlation

Fig. 8 Discovery of co-evolution with the "Fuzzy Set" (C5-2)

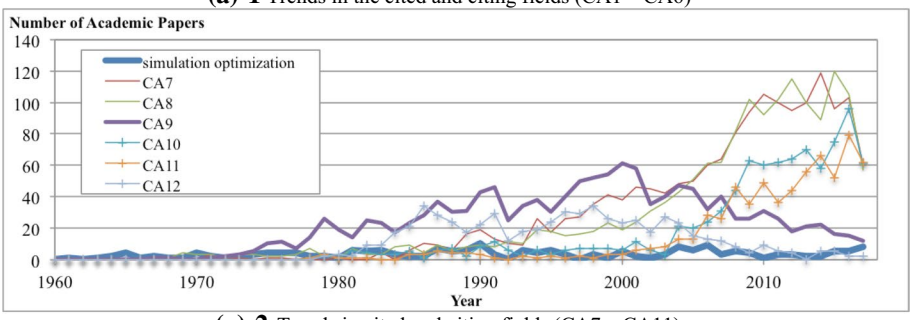
should be studied in the future. Here, the following seven pairs (among 146 pairs when the fields of co-evolution were checked) definitely co-evolved:

- Channel, antenna ↔ low power
- Channel, antenna ↔ video
- Channel, antenna ↔ multiplier
- Fuzzy set ↔ cognitive map
- Fuzzy set ↔ metric space
- Fuzzy set ↔ vessel tracking and segmentation
- Fluid analysis, fluid simulation ↔ quantum molecular dynamics.

The three co-evolutions from "channel, antenna" and the three co-evolutions from "fuzzy set" in Fig. 11, as listed above, were also co-evolutions in the timelines of Figs. 8 and 10.

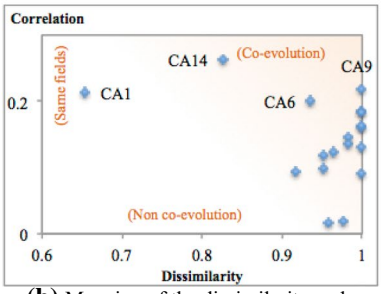


(a)-1 Trends in the cited and citing fields (CA1 – CA6)



(a)-2 Trends in cited and citing fields (CA7 – CA11)

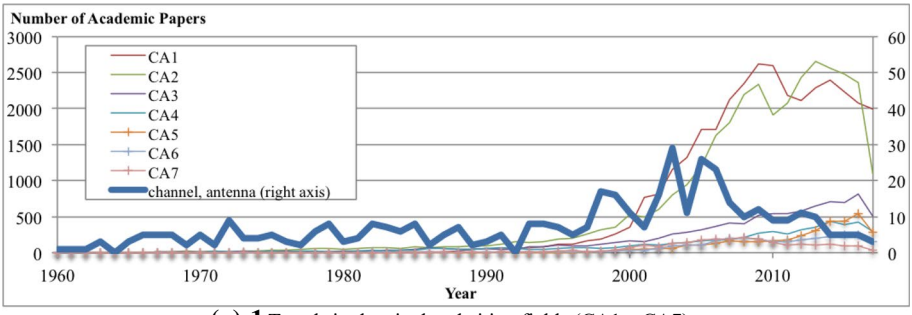
[Fields citing to "simulation optimization"]
CA1: function minimization
CA2: phylogenetics, Bayesian evolutionary analysis
CA3: biomolecules
CA4: global optimization, scheduling
CA5: simulation of chemical reaction
CA6: model order reduction
CA7: heat exchanger network
CA8: QSPR (Quantitative Structure Property Relationship)
CA9: electron nucleus
CA10: Petri Net
CA11: particle physics, neutrino
CA12: pharmacokinetics
CA13: copositive programming, noisy channel
CA14: bilevel programming
CA15: learning automata
CA16: location management
CA17: stochastic programming, dual dynamic programming
CA18: schedulability
CA19: conservation and spatial planning for fishery and marine
CA20: clustering for software modularization
(*Bold underlined fields co-evolved with the cited IT field.)



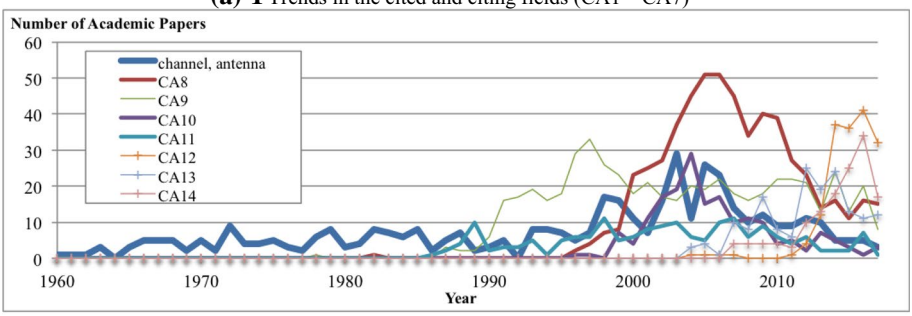
(b) Mapping of the dissimilarity and correlation

Fig. 9 Discovery of co-evolution with "Simulation Optimization" (C3-5)

In particular, with regard to "channel, antenna" having co-evolutions along the well-known history about mobile phones as described in the "Case 3: channel, antenna" section, co-evolutions can be extracted definitely even among the other fields.

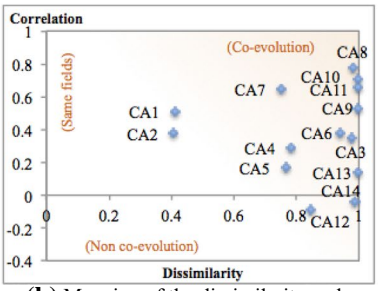


(a)-1 Trends in the cited and citing fields (CA1 – CA7)



(a)-2 Trends in the cited and citing fields (CA8 – CA14)

[Fields citing to "channel, antenna"]
CA1: space time
CA2: multiple-input multiple-output
CA3: protein database
CA4: quantum channel
CA5: energy harvesting
CA6: orthogonal frequency division multiplexing
CA7: low-density parity-check
CA8: low power
CA9: model-based diagnosis
CA10: video
CA11: multiplier
CA12: molecular communication
CA13: Kalman filter
CA14: steganography
(*Bold underlined fields co-evolved with the cited IT field.)



(b) Mapping of the dissimilarity and correlation

Fig. 10 Discovery of co-evolution with "Channel, Antenna" (C4-3)

Conclusion

This research extracted pairs of co-evolving academic fields by using a large amount of bibliographic data related to information technology. After the 28 IT fields in "Computer Science" and "Information Science and Library Science" were identified with a citation analysis for the top 1% of academic papers, academic papers citing 28 IT fields were retrieved. The dissimilarity and correlation between the citing fields and cited fields were computed. Some pairs of IT fields and citing fields were scrutinized, and there is an issue that pairs with high correlation included both co-evolutions and the same fields between the citing fields and cited fields. Thus, this research regarded pairs with both high dissimilarity and

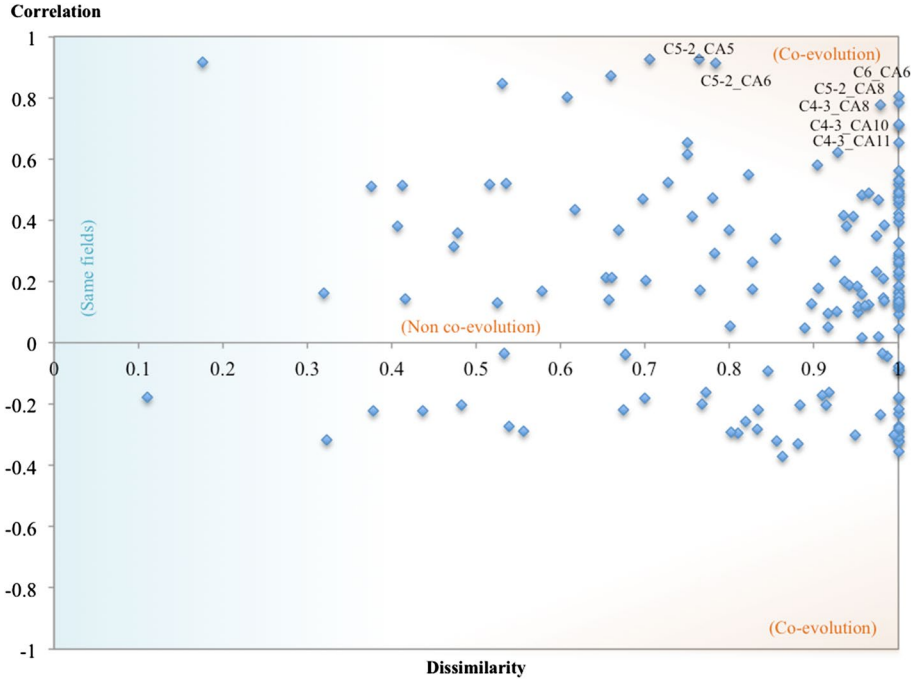


Fig. 11 Mapping of dissimilarity and correlation for the discovery of co-evolution

high correlation to be co-evolving pairs. Furthermore, using the time series charts, the background of the co-evolving academic field contributing to resurgence in IT field was confirmed by literature survey.

The findings from the results were as follows: (1) The IT fields "fuzzy set," "simulation optimization," and "channel, antenna" experienced a resurgence, and their co-evolved academic fields were identified from their trend charts. (2) Some pairs, such as "fuzzy set" and "vessel tracking and segmentation," with high correlation and high dissimilarity were extracted as co-evolutions between different fields.

The contributions of this research were as follows: (1) to provide evidence of co-evolution and resurgence in the past and (2) to propose a methodology to find co-evolution. To know unexpected collaboration between heterogeneous fields, it is better that co-evolving fields are combinations of IT field and other than IT field. In fact, there was a phenomenon that co-evolution by insiders and an outsider of IT field support the resurgence in the field of "channel, antenna." At the resurgence of "fuzzy set," "vessel tracking and segmentation" became one of co-evolved fields. It is found that co-evolutions contributing to resurgence were identified.

For the precautions and limitations when the methodology was used, (1) more pairs of co-evolution were found when the lower threshold of correlation and dissimilarity was set. (2) Although Fig. 11 was overviewed in multiple fields, thresholds should be designedly set in individual fields as shown in the "Case 2: simulation optimization" section. (3) Causalities were guaranteed with the usage of citation—even if their rise and fall in time-series trends did not occur simultaneously. (4) The gap of timelines to find co-evolution was not considered in this research.

This research provided a methodology that used correlation, dissimilarity, and citation to find co-evolutions, but it still required checking time-series trends to find resurgence.

Future works should address a methodology to find past resurgence and forecast resurgence candidates. In addition, in order to use co-evolution effectively, the durability or scale of co-evolutions had better to be measured. The discovery of co-evolution and resurgence will provide evidence for reviving industries through IT.

Acknowledgements Open access funding provided by University of Jyväskylä (JYU). The research leading to these results has received funding from the Strategic Research Council at the Academy of Finland under Grant Agreement No 293446—Platform Value Now: Value capturing in fast-emerging platform ecosystems. The server resources are supported by the research project of Prof. Ichiro Sakata, Prof. Junichiro Mori's laboratory in the University of Tokyo, and Prof. Yuya Kajikawa in Tokyo Institute of Technology.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Tables 1 and 2.

Table 1 Data specifications

Task	Records	Time	Citation
Clustering of Top 1% papers	14438	43 m	DC + CC + BC
Sub-clustering of C1	(4198)	12 m	DC + CC + BC
Sub-clustering of C2	(2664)	18 m	DC + CC + BC
Sub-clustering of C3	(2649)	22 m	DC + CC + BC
Sub-clustering of C4	(2076)	26 m	DC + CC + BC
Sub-clustering of C5	(1514)	19 m	DC + CC + BC
Clustering of citing articles to C2-2	94553	6 h 05 m	DC + CC
Clustering of citing articles to C3-1	175616	10 h 55 m	DC + CC
Clustering of citing articles to C3-2	113810	5 h 43 m	DC + CC
Clustering of citing articles to C3-3	7562	14 m	DC + CC
Clustering of citing articles to C3-5	77605	3 h 58 m	DC + CC
Clustering of citing articles to C4-3	100413	6 h 44 m	DC + CC
Clustering of citing articles to C4-5	32921	1 h 13 m	DC + CC
Clustering of citing articles to C4-6	29375	1 h 06 m	DC + CC
Clustering of citing articles to C5-2	79062	7 h 08 m	DC + CC
Clustering of citing articles to C6	178381	17 h 49 m	DC + CC
Clustering of citing articles to C7	24647	1 h 03 m	DC + CC
Total	1614690	184 h 57 m	–

Due to the eventual shortage of computer performance, the cosine similarities of the two strike-through records cannot be calculated after citing articles to C1-1 and C1-2 are clustered

The above times do not include the time to try and error for identifying suitable size and category of datasets before practically analyzing. In the column of the dataset, the numbers of C (cluster) correspond to C# (cluster ID) in Table 2. In the column of the citation type, DC + CC + BC uses direct citations, co-citations, and bibliographic couplings. DC + CC uses direct citations and co-citations. It takes time to retrieve data from the service within our access right without a DoS (Denial of Service) attack

Table 2 Fields of the top 1% of academic papers

C#	Ave. year	# of nodes	# Including "Internet"	Topic
1	2002.1	4198	46	Image recognition, machine learning
1-1	2001.4	1436	20	Classification, machine learning
1-2	1996.6	1374	5	Image segmentation, edge detection
1-3	2008.2	858	5	Face recognition
1-4	2009.5	497	15	Motion capture
2	2003.1	2664	177	Technology acceptance
2-1	2002.5	1358	26	Bibliometrics, knowledge domain
2-2	1999.2	544	31	Information systems in organizations
2-3	2004.7	356	86	Technology acceptance
2-4	2010.7	172	2	Genome analysis
2-5	2008.4	137	30	Online community
3	1994.6	2649	22	Optimization, genetic algorithm
3-1	1990.2	798	8	Routing, scheduling, graph
3-2	1991.9	543	3	Cryptography
3-3	1990.8	457	6	Agent, logic
3-4	2008.4	387	3	Optimization algorithm
3-5	1993.0	193	0	Simulation optimization
3-6	2003.9	192	2	Multiobjective evolutionary algorithm
4	2004.6	2076	233	Network, wireless, cloud, sensor
4-1	2011.1	503	136	Cloud computing, IoT
4-2	2009.5	494	39	Wireless, sensor network
4-3	1995.8	418	3	Channel, antenna
4-4	2003.8	136	1	Eavesdropper
4-5	2000.9	128	34	Network traffic
4-6	1984.9	102	0	Transmission error detection
5	2006.7	1514	8	Fuzzy
5-1	2010.0	688	3	Fuzzy control
5-2	2002.8	600	5	Fuzzy set
5-3	2007.4	100	0	Rough set
6	1996.1	784	3	Fluid analysis, fluid simulation
7	1994.0	112	0	Diffusion equation
8	2006.8	104	3	Recommender system
9	2003.4	103	0	Elementary particle physics

The clusters C1–C5 are too big, so they have been divided into sub-clusters C1-1–C5-3

References

Abate, A. F., Nappi, M., Riccio, D., & Sabatino, G. (2007). 2D and 3D face recognition: A survey. *Pattern Recognition Letters, 28*(14), 1885–1906. https://doi.org/10.1016/j.patrec.2006.12.018.

Ayala, G., Leon, T., & Zapater, V. (2005). Different averages of a fuzzy set with an application to vessel segmentation. *IEEE Transactions on Fuzzy Systems, 13*(3), 384–393. https://doi.org/10.1109/TFUZZ.2004.839667.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*. https://doi.org/10.1002/aris.1440370106.

Boston Dynamics. (2018). *BigDog*. https://www.bostondynamics.com/bigdog. Retrieved April 7, 2018.

Broven, J. (2009). *Record makers and breakers: Voices of the independent rock "n" roll pioneers*. University of Illinois Press. https://books.google.com/books?id=uFrYsNvmrvQC&pg=PA187. Retrieved August 19, 2018.

Center for Research and Development Strategy (CRDS), Japan Science and Technology Agency (JST). (2016). *Innovations of basic technology for robots through integration of nanotechnology, information technology, and mechanics—Toward smart and human-friendly robots (in Japanese)*. http://www.jst.go.jp/crds/pdf/2015/SP/CRDS-FY2015-SP-03.pdf. Retrieved April 26, 2018.

Charny, B. (2001). *World's first 3G phone network goes live*. ZDNet. https://www.zdnet.com/article/worlds-first-3g-phone-network-goes-live/. Retrieved September 4, 2018.

Chiou, C. B., Chiou, C. H., Chu, C. M., & Lin, S. L. (2009). The application of fuzzy control on energy saving for multi-unit room air-conditioners. *Applied Thermal Engineering, 29*(2–3), 310–316. https://doi.org/10.1016/j.applthermaleng.2008.02.031.

European Commission (EC). (2002). *Communication from the Commission on the collection and use of expertise by the Commission: Principles and guidelines*. https://publications.europa.eu/en/publication-detail/-/publication/969f5240-ec81-4998-911e-6831d9318919. Retrieved April 30, 2018.

European Commission (EC). (2016). *Foresight and horizon scanning—JRC science hub*. https://ec.europa.eu/jrc/en/research/crosscutting-activities/foresight. Retrieved November 11, 2017.

European Commission (EC). (2017). *Foresight—Research and innovation*. https://ec.europa.eu/research/foresight/index.cfm. Retrieved November 11, 2017.

Fujita, K., Kajikawa, Y., Mori, J., & Sakata, I. (2012). Detecting research fronts using different types of combinational citation. In *17th International conference on science and technology indicators (STI 2012)*. Montreal. http://2012.sticonference.org/Proceedings/vol1/Fujita_Detecting_273.pdf.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science, 122*(3159), 108–111. https://doi.org/10.1126/science.122.3159.108.

Google. (2014). *Google genomics*. https://cloud.google.com/genomics/. Retrieved November 22, 2017.

Grisan, E., Pesce, A., Giani, A., Foracchia, M., & Ruggeri, A. (2004). A new tracking system for the robust extraction of retinal vessel structure. In *The 26th annual international conference of the IEEE engineering in medicine and biology society* (Vol. 3, pp. 1620–1623). IEEE. https://doi.org/10.1109/iembs.2004.1403491.

Haire, M. (2009). The walkman. *TIME*. http://content.time.com/time/nation/article/0,8599,1907884,00.html. Retrieved August 16, 2018.

Huang, S., Sun, Z., & Lester, W. A. (1990). Optimized trial functions for quantum Monte Carlo. *The Journal of Chemical Physics, 92*(1), 597–602. https://doi.org/10.1063/1.458409.

IFPI. (2017). *Global music report 2017: Annual state of the industry*. http://www.ifpi.org/downloads/GMR2017.pdf. Retrieved June 28, 2018.

Innovation Policy Research Center (IPRC). (2013). *Academic landscape system*. http://academic-landscape.com/. Retrieved November 11, 2017.

Ittipanuvat, V., Fujita, K., Sakata, I., & Kajikawa, Y. (2014). Finding linkage between technology and social issue: A literature based discovery approach. *Journal of Engineering and Technology Management, 32,* 160–184. https://doi.org/10.1016/j.jengtecman.2013.05.006.

Iwami, S. (2017). Study on the destination of research via knowledge flows. *Scientometrics, 112*(1), 273–288. https://doi.org/10.1007/s11192-017-2395-x.

Iwami, S., Harada, D., & Kajikawa, Y. (2015a). A study on the extraction of existing science and technology candidates to apply for the cutting-edge academic fields (in Japanese). In *The Japan society for science policy and research management, the 30th annual meeting*. Tokyo.

Iwami, S., Kogo, K., Tacoa, F., Mori, J., Kajikawa, Y., & Sakata, I. (2015b). Cross-disciplinary methodology for detection of collaborative and competitive candidates. In *The 24th international conference for management of technology (IAMOT 2015)*. Cape Town. http://iamot2015.com/2015proceedings/documents/P013.pdf.

Iwami, S., Mori, J., Sakata, I., & Kajikawa, Y. (2014). Detection method of emerging leading papers using time transition. *Scientometrics, 101*(2), 1515–1533. https://doi.org/10.1007/s11192-014-1380-x.

Kajikawa, Y., Ohno, J., Takeda, Y., Matsushima, K., & Komiyama, H. (2007). Creating an academic landscape of sustainability science: An analysis of the citation network. *Sustainability Science, 2*(2), 221–231. https://doi.org/10.1007/s11625-007-0027-8.

Kajikawa, Y., Tacoa, F., & Yamaguchi, K. (2014). Sustainability science: The changing landscape of sustainability research. *Sustainability Science, 9*(4), 431–438. https://doi.org/10.1007/s11625-014-0244-x.

Kimmerle, J., Moskaliuk, J., Harrer, A., & Cress, U. (2010). Visualizing co-evolution of individual and collective knowledge. *Information, Communication & Society, 13*(8), 1099–1121. https://doi.org/10.1080/13691180903521547.

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. https://doi.org/10.1002/asi.20967.

McCarthy, A. (2012). The business of genome analysis and interpretation. *Chemistry & Biology, 19*(8), 927–928. https://doi.org/10.1016/j.chembiol.2012.08.006.

Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering. *Research Policy, 31*(8–9), 1389–1403. https://doi.org/10.1016/S0048-7333(02)00070-7.

Naveed, K., Watanabe, C., & Neittaanmäki, P. (2017). Co-evolution between streaming and live music leads a way to the sustainable growth of music industry—lessons from the US experiences. *Technology in Society, 50,* 1–19. https://doi.org/10.1016/j.techsoc.2017.03.005.

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E, 69,* 66133.

NTT DoCoMo. (2014). *PHS (personal handyphone system)* (In Japanese). http://history-s.nttdocomo.co.jp/list_phs.html. Retrieved August 18, 2018.

Ochiai, A. (1957). Zoogeographical studies on the Soleoid fishes found in Japan and its neighhouring regions-II. *Nippon Suisan Gakkaishi, 22*(9), 526–530. https://doi.org/10.2331/suisan.22.526.

Ojala, A. (2016). Business models and opportunity creation: How IT entrepreneurs create and develop business models under uncertainty. *Information Systems Journal, 26*(5), 451–476. https://doi.org/10.1111/isj.12078.

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology, 61*(9), 1871–1887. https://doi.org/10.1002/asi.21368.

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. https://doi.org/10.1016/j.technovation.2008.03.009.

Shibata, N., Kajikawa, Y., Takeda, Y., Sakata, I., & Matsushima, K. (2011). Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications. *Technological Forecasting and Social Change, 78*(2), 274–282. https://doi.org/10.1016/j.techfore.2010.07.006.

Tadano, K., & Kawashima, K. (2006). Development of 4-DOFs forceps with force sensing using pneumatic servo system. In *Proceedings 2006 IEEE international conference on robotics and automation, 2006. ICRA 2006* (pp. 2250–2255). IEEE. https://doi.org/10.1109/robot.2006.1642038.

Tuli, T. S., Orr, N. G., Robert, D., & Zee, E. (2006). Low cost ground station design for nanosatellite missions. In *North American space symposium*. North American Space Symposium. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.564.3644.

U.S. Department of Health and Human Services (HHS). (2016). *STAR METRICS*. https://www.starmetrics.nih.gov/. Retrieved November 11, 2017.

van der Pol, J., & Rameshkoumar, J.-P. (2018). The co-evolution of knowledge and collaboration networks: The role of the technology life-cycle. *Scientometrics, 114*(1), 307–323. https://doi.org/10.1007/s11192-017-2579-4s.

Vlasic, D., Adelsberger, R., Vannucci, G., Barnwell, J., Gross, M., Matusik, W., et al. (2007). Practical motion capture in everyday surroundings. *ACM Transactions on Graphics, 26*(99), 35. https://doi.org/10.1145/1239451.1239486.

Wikipedia. (2018). History of smart antennas. In *Wikipedia*. https://en.wikipedia.org/wiki/History_of_smart_antennas. Retrieved August 18, 2018.

Yagi, H., & Uda, Shintaro. (1926). Projector of the sharpest beam of electric waves. *Proceedings of the Imperial Academy*. https://doi.org/10.2183/pjab1912.2.49.