Visualization and quantitative study in bibliographic databases: A case in the field of university–industry cooperation

Feng Feng [a], Leiyong Zhang [a], Yuneng Du [b,*], Weiguang Wang [c]

[a] *School of Public Affairs, University of Science and Technology of China, China*
[b] *College of Economics & Management, Anhui Agricultural University, Hefei 230036, China*
[c] *Chinese Academy of Sciences, China*

A R T I C L E   I N F O

A B S T R A C T

CiteSpace is a visual document analysis software, by which performances and trends of certain disciplines can be displayed for a given period. Moreover, the evolution of a frontier research can be explored by such software as well. This research focuses on the visualization and quantitative study in bibliographic databases by taking the university–industry collaboration studies as an example. Using the Web of Science (WOS), 587 publications and over 30,000 references were selected for analysis, which produced the following results: (1) Our method can clearly reveal the key elements of certain disciplines, such as the largest share of publications, the most frequently cited authors and journals in the university–industry cooperation research field; (2) The relationships among the frequently cited authors, references, journals and keywords can be explained visually in the university–industry cooperation research field; (3) Of special note is that the potential problems and evolutionary trends of certain research fields such as university–industry cooperation can also be ascertained via our method; (4) In general, according to the case study, our visualization and quantitative method evolved a new research framework to evaluate the performance of some research areas.

1. Introduction

Quantitative research of literatures appeared in the early 20th century but had not formed an independent discipline until Pritchard (1969) first proposed that the terminology Statistical Bibliography should be replaced by Bibliometrics. Since then the theory and practice of literature studies on the basis of bibliometrics has become widespread in academic research (Diem & Wolter, 2013). The metamorphosis of Bibliometrics generated by the development of networks and computer technology has made graphical study and visualization research of literatures possible. That images contain much more information than digits or words of the same size is a foregone truism (Ma & Xi, 1992). This becomes evident when looking at all kinds of visualization software. Exploration and prediction of frontier science has increasingly become more popular over the years, and analyzing text messages and citation information of literatures by computer provides a new perspective of Bibliometrics. Complicated phenomenon and analysis results can be obtained in the process of visualization research. Using the images

---

generated by computer software, researchers can quickly understand the research status and be able to forecast the possible research directions in the future (Chen, 2006). Thus new study areas will be identified through visualization study, yet it must be noted that the basis of visualization study is co-citation analysis (Ma & Xi, 1992).

Co-citation analysis theory suggests that any new knowledge is derived from existing ones. If two papers are cited by one or more papers, then they have co-citation relationship, and the number of citations is called the co-citation degree. A high co-citation degree means that the two papers are closely related to each other (Small, 2003). White and McCain (1998) tried to describe the subject structures of information science with the method of Author Co-citation Analysis (ACA), which were developed by them since 1981 (White, 1990). Cluster analysis, multi-dimensional scaling and factor analysis were applied in their study with the help of SPSS. After that, lots of studies on the description and analysis of different areas appeared, and some retrieval systems and analysis software were developed based on ACA. Traditional ACA method gained huge success, but some researchers still questioned it. The arguments were focused on the following points: (1) whether Pearson correlation coefficient r was suitable for the measure of correlation between the authors (See Ahlgren, Jarneving, & Rousseau, 2003; Bensman, 2004; Leydesdorff & Zaal, 1988; White, 2003); (2) which matrix should be used to generate the similarity coefficient matrix (See Kruskal & Wish, 1978; Leydesdorff & Vaughan, 2006; Miguel, Moya-Anegón, & Herrero-Solana, 2008; Waltman & Eck, 2007). To solve this dilemma, the optimization of the ACA was developed into two directions. Some researchers tried to choose proper similarity measurement according to the specific data, or better ways to generate a similarity matrix, by which visual maps can be created with the help of SNA (Social Network Analysis) software such as Pajek, Ucinet or VxOrd. Other researchers tried to use the co-citation matrix directly, for example, White (2003) introduced Pathfinder Network Scaling (PFNETs) into ACA, and the original co-citation data in the co-citation matrix was used directly. In the visual maps which were created with PFNETs, the nodes represented the authors, the lines between nodes represented weighted path, and the co-citation number was the weight (White, 2003).

In the two directions above, visual maps are important concepts. In some sense, visualization is the future direction for bibliographic studies. As mentioned above, since images contain much more information than words or data, this area can be called mapping knowledge domains. In studies, co-citation analysis of authors, titles, keywords and institutions can be explored. The performance of certain study areas can be shown to the readers intuitively by maps with authors, literatures, journals and institutes as nodes in it. CiteSpace is a typical tool for co-citation network analysis and visualization. Based on some concepts in information science (research front, intellectual base and time-variant duality), CiteSpace can generate two complementary views, viz., cluster views, and time-zone views (Synnestvedt, Chen, & Holmes, 2005). One of the functions of CiteSpace is drawing visual co-citation maps. Separate co-citation networks are generated first and then a whole map is combined with the separate ones. Important literatures can be recognized in the map due to their prominent features. Thus, the finding of Turning Points can be simplified to the searching of key nodes in the visualization maps, and the evolution of the area can also be detected and monitored with the key nodes (Chen, 2003). Another function of CiteSpace is in detecting the hottest topics and predicting future research trends. Hybrid network of cited papers and their citing papers are mapped by the software and burst terms of the area are detected by an algorithm called burst detection (Chen, 2006), which reveal the research trends.

This paper tried to use CiteSpace to develop visualization and quantitative study in bibliographic databases. University–industry collaboration was selected to be the target area, because this field has become a hot issue in recent years. For a long time, the university–industry collaboration was an important issue related to management studies, science and technology policy studies, innovation studies, industrial organization or network, and science of science studies (Agrawal, 2001; Alice, 2007; Hancock & McCurry, 1993; McMillan & Hamilton, 2003; Thune, 2007). University–industry relationship is not just a link or a connection, it is rich in meanings. Universities not only draw R&D funds from enterprises to develop science and technology, but also cultivate more talents through the talent exchange program with firms. At the same time, industry may increase profits and market share via the technical progress of universities (Cao, Zhang, Feng, & Du, 2013; Feng, Zhang, Du, Ma, & Fu, 2012). Generally, cooperation of universities and industry must go through an interface or media which can be tangible such as diffusion of technology transfer offices, technology transfer center, or intangible such as technology, knowledge, patents or even information (Nelson, 2001; Shane, 2005; Siegel, Waldman, & Link, 2003; Thursby, Jensen, & Thursby, 2001). Heterogeneity and complementarity are the two key reasons in the formation process of university–industry cooperative relation (Kumar, 1994). Upon upgradation of cooperation, university and industry may build a certain economic entity together, which is common in practice (Jonas, Göran, Inger, John, & Mats, 2006; Pickard, 1944). Previous studies have given lots of details of collaborations between universities and industry, but there is little concern for bibliometric analysis of the university–industry collaborations studies. Although several studies have emphasized the importance of literature analysis on university–industry links (Aurora & Luisa, 2012), to our knowledge, a more quantitative and graphical approach based on bibliometric techniques has yet to be undertaken. As a consequence, the university–industry cooperation is a perfect example to construct our visualization and quantitative research framework in bibliographic databases, since the visual images with graphical techniques are really needed in the university–industry cooperation study (Aurora & Luisa, 2012).

In this paper, visualization and quantitative research framework was created through the following regimen: the history of the university–industry collaboration studies was mapped and hot topics in recent years were found; using bibliometrics analysis method with the help of Web of Science and the program of CiteSpace II, the trend of university–industry collaboration studies was depicted quantitatively (Chen, 2004, 2006); Panoramic information was revealed and the key literatures and key authors of the university–industry collaboration areas were identified in our work (Hou & Hu, 2013). The ultimate

aim of this paper is to use the university–industry collaboration issue as a target to form a framework of visualization and quantitative research with CiteSpace.

2. Methodology and data

Bibliometrics is a set of methods to quantitatively analyze academic literature by the knowledge of mathematics, physics and computer science. Generally, the development of science follows the rules of conversion between general science and scientific crisis (Kuhn, 2012). The view of paradigm conversion is well known in almost all subjects, so it is possible to identify and detect the changes in certain areas based on cluster analysis of document co-citation (Small, 2003). The research of Small (2003) proposed this before many other modern visualization researches, but the conversion processes in clusters of different years were ignored in his work. In recent years, the analysis of complex networks has provided a potential pathway for bibliometrics (Albert & Barabási, 2002). These ideas are applied to CiteSpace by computer software (Chen, 2004, 2006). The principles and algorithm of CiteSpace are as follows.

(1) Divide-and-conquer

The computation time of a problem is related to the scale. Sometimes a huge problem may be very hard to solve. The idea of divide-and-conquer claims that a huge problem should be divided into several smaller ones, thus making them easier to conquer.

(2) Success breeds success

A paper will have higher probability of being read if it was more frequently cited, and may be cited even more in future research. The citing probability of a paper may be in proportion to the number of exiting citing papers. Success Breeds Success is one of the principles in the development of CiteSpace.

(3) Pathfinder network scaling

Pathfinder Network Scaling (PFENTs) shares the same math model with Social Network Analysis. In this algorithm, most of the links in the network are deleted and only the most important relations are retained. The aim of PFENTs is to simplify the complex network. Semantic Analysis, Social Network Analysis and Virtualization Analysis are introduced into the ACA methods with the introduction of PFNETs, so PFNETs has great advantages compared to traditional ACA method. In PFNETs, the papers, keywords, titles and authors are treated as the nodes of network, and the nodes are assumed to be interconnected with each other by weighted paths. The co-citation frequency of the subjects is the weight and only the shortest path is shown in the network. The key nodes form informal communication networks and the trends of the field will be controlled by them. Conversely, if a lot of nodes contain high co-citation frequencies, the discipline branches will be formed automatically, and the specialized clustering procedure is not required. The maps of PFNETs are mapped with the original co-citation frequency data, so the results can be more abundant than the traditional ACA method. In addition, PFNETs reduce the complexity of the traditional ACA method, and the results are more reliable.

(4) Minimum spanning trees

Assuming $G = (V, E)$ is an undirected weighted graph and $G\prime$ is a subgraph of G. If $G\prime$ contains all the nodes of G, then we can call it a Spanning Tree of G. The sum of weights of the spanning trees can be called the Cost of the Spanning Tree. The spanning tree with the least cost is called the Minimum Spanning Tree. In the modern graph theory, the Minimum Spanning Trees can be gained by the Prim Algorithm (Prim, 1957) and the Kruskal Algorithm (Kruskal, 1956) with the help of computer programming.

(5) Expectation maximization clustering method

Expectation Maximization Clustering Method (referred to as EM Clustering Method) is one of the most important maximum likelihood estimation algorithms in statistics. Missing variables X can be estimated using the existing data via the EM clustering algorithm. Conditional expectation can be got in E-Stage (Expectation) and the maximum likelihood estimation value can be got in M-Stage (Maximum). In EM clustering algorithm, assuming J is latent variable, U is given data and Θ is the maximum posterior probability, then

$$\Theta^* = \text{argmax}\Theta \sum J P(\Theta, J \mid U).$$

EM clustering algorithm is the main clustering method in CiteSpace, which uses the specific parameters (such as the publication year, author, time cited or half-life period) to cluster. Clustering criterion is identified via the statistical analysis by EM Clustering Method. The same clustering literature is marked with the same color and shown in the maps.

CiteSpace II is a Java program and it is very useful for visualization and analysis of literatures in bibliometrics. It was developed by Chaomei Chen in the year of 2004. The version upgrades and updates continually and the version of this paper used is CiteSpace-v3.0.R2 beta. Generally speaking, the input of CiteSpace is just the output of ISI (i.e. the saved format of Web of Science). CiteSpace has a built-in function of data conversion, so there is no need to convert the original data into correlation matrix. All the documents must be put into the same folder and the name of each document should begin with "download" and finish with ".txt" (such as download-S.T.policy-extinction-2013.txt). Data storage path and project
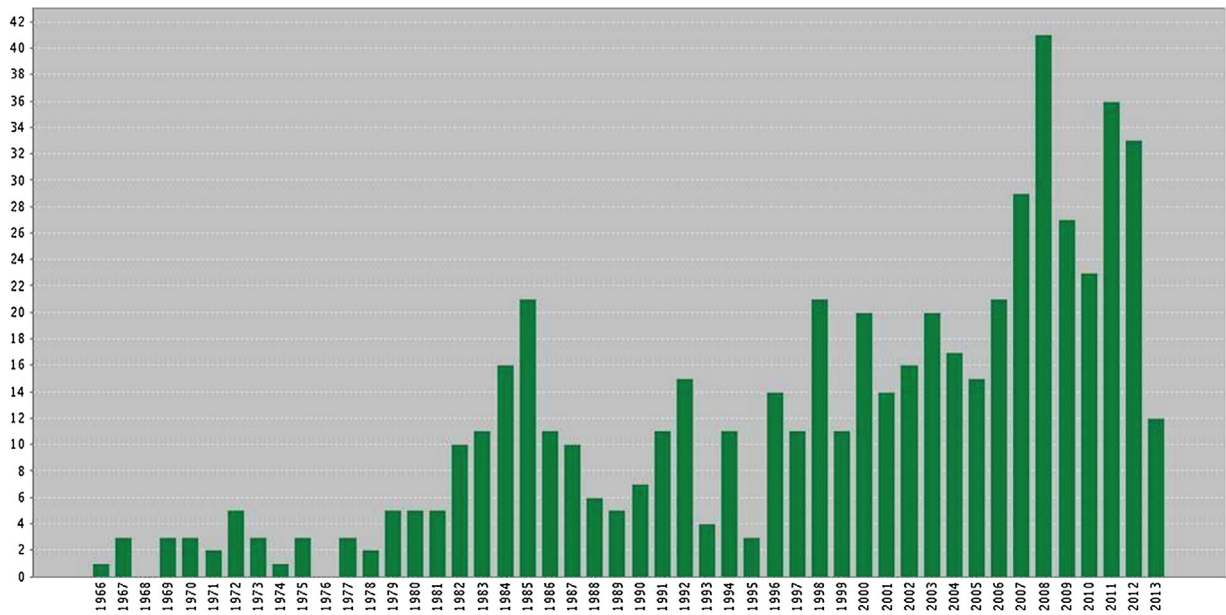
Fig. 1. Number of articles published on university–industry collaborations, 1966–2013, from articles in the literature database built for this paper. *Note*: All papers were published in 281 journals.

storage path should be designated before the running of the program. All the settings can be finished in the main interface of CiteSpace II.

This paper used CiteSpace II to analyze the trend of university–industry collaboration. The data was derived from the database of Web of Science. Web of Science is one of the most important literature databases and citation information sources. To get the original data, terms "academic–industry" and "university–industry" were retrieved from Web of Science by setting database as "SCI-EXPANDED, SSCI, CPCI-S, CCR-EXPANDED, IC". Time span was limited to "all years". To ensure the accuracy of the results, we screened the details of each paper carefully and finally 587 records were selected to build the database of this paper. Although Web of Science only contained parts of all the papers about university–industry collaborations, their references covered almost all the important articles in this area. In this case, we depict a map of the evolution of studies on university–industry collaborations. For bibliometrics analysis, the useful publications are always the top ones, which indicate that the papers restricted in Web of Science can hardly have a negative effect on our analysis. This paper set the time scale to 2 years. The breaking of the time can increase the speed and accuracy of the software. At the same time, key nodes and temporal patterns of the evolving discipline can be recognized more easily.

However, the citations of recent publications have higher potential to be cited in the future, and their importance might not emerge as yet. In our opinion, the citation of one paper is discrete and few methods can give a completely accurate normalized result of citation. In the future this problem will continue to be a concern.

3. Basic depiction of university–industry collaboration studies

Since the year of 1966, studies on university–industry collaboration have continued over the years. In the year of 2008, the largest number of papers (41 total) focused on this topic, accounting for 7.25% of all 587 publications, followed by the year of 2011 (6.36%). During the period of 2005–2012, related researches were the most active in this field (almost 40%) (cf. Fig. 1).

All in all the 587 articles were cited 4520 times during 1966–2013 (i.e. Jun. 15, 2013), and excluding self-citations it was 4112 times according to our results. On the one hand, in contrast to the number of publications, the number of citations has been increasing since the year of 1999, which means the impact of previous studies on university–industry collaborations to new ideas was becoming huger. On the other hand, the topic of university–industry collaborations was becoming hotter (cf. Fig. 2). It should be noted that only an overview of the citations is shown in this section, and more comprehensive and accurate analysis will be conducted in the following sections. For the 587 papers, the average of citations was 7.99 (Fig. 3).

The works of 25 authors listed in Table 1 account for almost 20% of the publications in this field. Among these, Leydesdorffb L. (Worked for Department of Science and Technology Dynamics, Netherlands) is the author who published the highest number in this field (14 publications), followed by Etzkowitz (State University of New York, US) with 9 publications. It should be noted that Leydesdorffb and Etzkowitz are collaborators in many works. They have made great contributions to the development of university–industry collaboration studies and have expanded the "Mode 2" model to a Triple Helix of university–industry–government relationship (Etzkowitza & Leydesdorffb, 2000; Leydesdorffb & Etzkowitza, 1996).
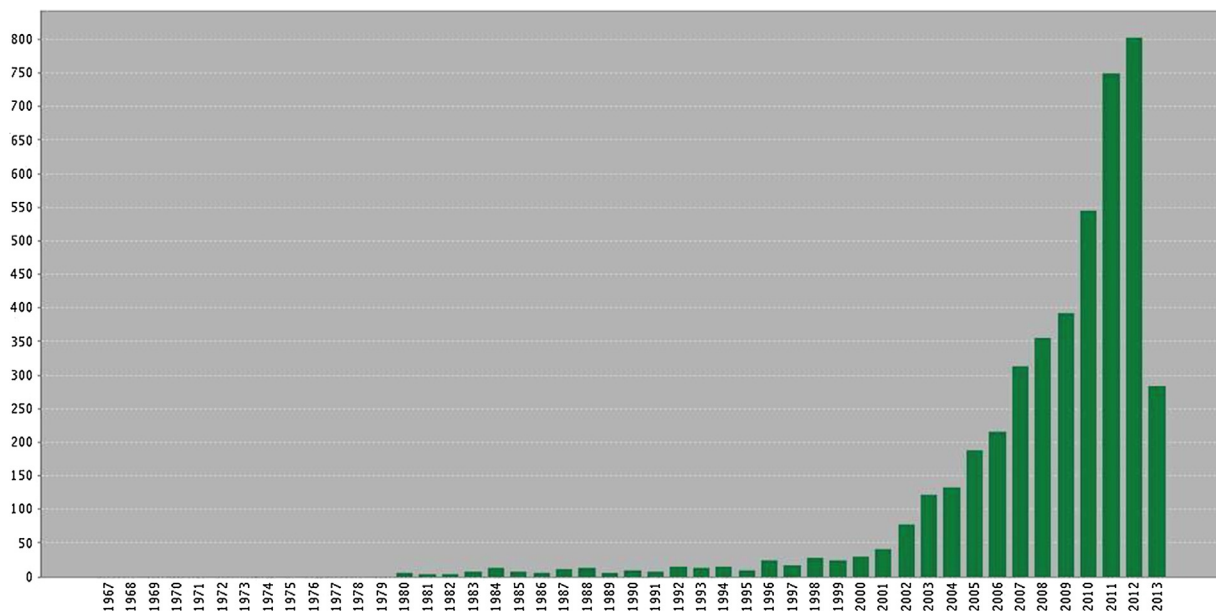
Fig. 2. Number of citations of articles each year, 1967–2013, from articles in literature database built for this paper.

Abstracts of papers of the American Chemical Society published the most papers in the field of university–industry collaborations (A total of 41, i.e. 6.99%). As is becoming increasingly evident, chemistry is one of the areas where academia and industry most closely collaborate. The utilization and industrialization level of technology in chemistry is also the highest (Goddard & Dasgupta, 1992). There are more articles concerned with the field of university–industry collaboration in chemistry. So it is not surprising that most papers are published in a chemical journal. Journals such as *Research Policy* (33 in total), *Science* (16 in total), *Technovation* (17 in total), *High Education* (12 in total), *IEEE Transactions on Education* (12 in

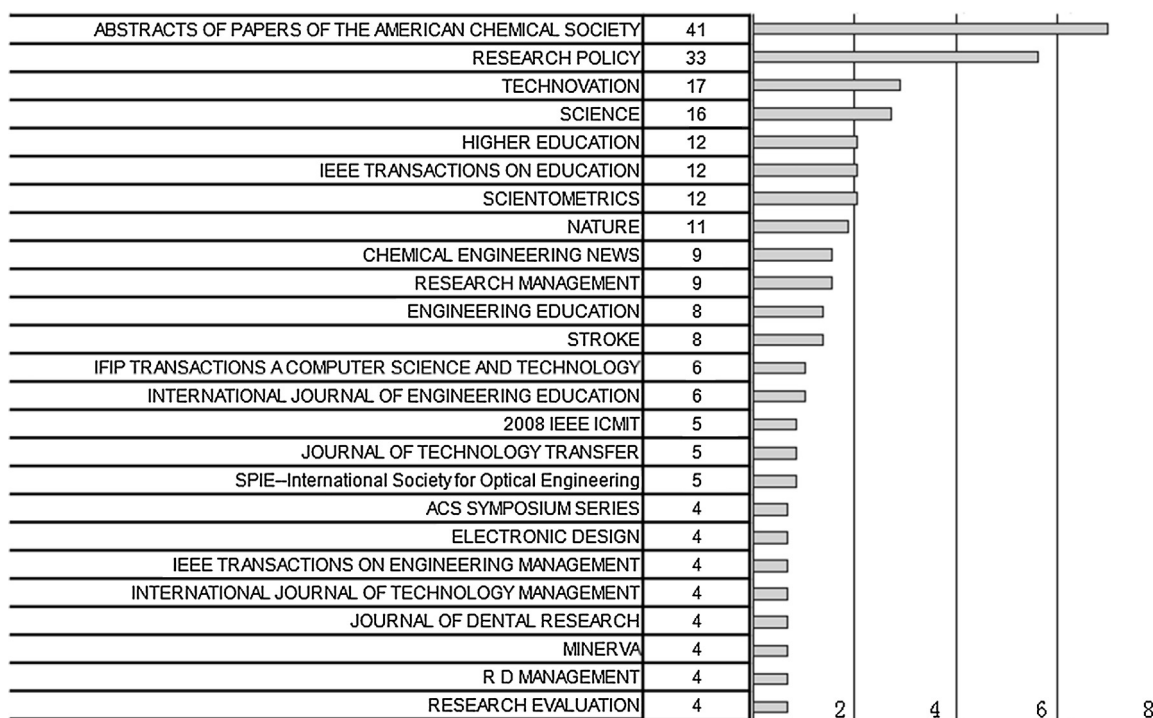**Fig. 3.** Top 25 journals with articles in the field of university–industry collaboration. *Note*: The gray bars denote the proportion of articles published in the journal to the total of 587 publications. It may be noted that just 25 journals are listed here. Some journals that published 4 articles may have not appeared in the result.

Table 1
Top 25 authors in the field of university–industry collaboration studies.

Author	Number	Proportion	Author	Number	Proportion
Anonymous	14	2.385%	Welsh R.	4	0.681%
Leydesdorff L.	11	1.874%	Albers G.W.	3	0.511%
Etzkowitz H.	9	1.533%	Arvanitis S.	3	0.511%
Fisher M.	6	1.022%	Azagra-Caro J.M.	3	0.511%
Blumenthal D.	5	0.852%	Brostrom A.	3	0.511%
Abramo G.	4	0.681%	Burrington J.D.	3	0.511%
Armstrong J.A.	4	0.681%	Campbell E.G.	3	0.511%
Biscotti D.	4	0.681%	Carayannis E.G.	3	0.511%
D'angelo C.A.	4	0.681%	Chen J.	3	0.511%
Geuna A.	4	0.681%	D'este P.	3	0.511%
Kawasaki K.	4	0.681%	Di Costa F.	3	0.511%
Perkmann M.	4	0.681%	Dickson D.	3	0.511%
Phillips D.I	4	0.681%			

total), *Scientometrics* (12 in total), *Nature* (11 in total) are associated closely with academic–industry cooperation, and their publications in this field appear more frequently.

4. Reference analysis results

4.1. Mapping and analysis on references

The references analysis is one of the most important contents of bibliometrics, and the influence of the authors and the papers can be given as results. References of papers in the dataset were analyzed by CiteSpace II, and the top 30 papers of each year in the dataset were selected for analysis. With the help of CiteSpace II, the top 10 cited papers of all years can be found according to their frequency of being cited (cf. Table 2). Then visualization of the data can be developed by CiteSpace II. The operation of CiteSpace II is very simple and the original data do not need to be converted into a matrix. Different maps can be drawn via the same data sample and the evolution features of data can be shown from different perspectives. The nodes and lines are marked with different colors by CiteSpace II, which show the time-varying characteristics of the literatures. Color rings of nodes represent the citation characteristics of the nodes at different times (the nodes may be the authors, keywords, journals and so on), and colors of the lines represent the time that co-citation frequency of lines reached the threshold.

In all the references, the paper written by Cohen (Camegie Mellon University) and published in *Management Science* was cited the most in our database. It should be noted that another of Cohen's articles published in *Administrative science quarterly* figured also in the top 10. Cohen's papers focused on the collaborations of public research and industrial R&D, on learning and innovation (Cohen & Levinthal, 1990; Cohen, Nelson, & Walsh, 2002). He is also one of the earliest researchers who was concerned about the link or relationship between the industry and academia. Other papers in the top 10 were also linked closely to the topic of Cohen's paper.

The results also suggested that references which were cited the highest number of times were also the ones which have high centralities. Centrality is an indicator given by CiteSpace II, showing the ability of one reference to link with other references. Generally, references with high centrality are the ones which combine papers together in this field. The results indicate that university–industry collaboration studies now are inclined to review the previous studies on relative topics. Rosenberg's work has the highest centrality which means that this paper has more relevance with other papers in some sense (Rosenberg & Nelson, 1994). The second one published by Gibbons also has a high centrality (Gibbons et al., 1994).

Looking at the map of university–industry collaboration references (cf. Fig. 4), we were able to find that the most important studies were linked closely. It meant that those researches were combined sufficiently, and studies in this field still focus on the ideas created in the past. In some sense, it is a good thing because most studies were continuous and progressive, but it should be noted that new studies have no effect on the old ideas. On the other hand, some branches sprouted out of the largest core of the map. It means that researchers' works in this field have started some explorations that venture beyond the previous. However, their work was formed unsystematically, which meant that those researches were still in a primary stage. The points or nodes located far from the core were in different fields or even different subjects. So we can just focus on the big ones in the center (cf. Fig. 4). Five publications are chosen by CiteSpace II and shown in the map (cf. Fig. 4), and they are the most important papers in the references. Besides Cohen's, Rosenberg's and Gibbons' works mentioned above, Jaffe (1989) developed an exploratory research with state-level time-series data on corporate patents, corporate R&D and university research. It is obvious that his work concerns a type of cooperation between universities and enterprises. A significant effect on the corporate and an indirect effect on local innovation were found in his study. The areas which Jaffe pointed out (such as Drugs, Medical Technology, Electronics, Optics and Nuclear Technology) were all associated closely with university–industry studies. Geisler's work in 1989 (published on Cooperative) was cited 5 times and did not appear in Table 2, but it was a core paper in some sense. It did not relate to other core papers but formed a sub-area around itself. Three questions (Why do universities and industry cooperate? What are the more effective mechanisms for such cooperation?

Table 2
Top 10 cited references of articles in the literature database of this paper.

Study	Citations by U–I literatures (Centrality)	Citations by [Google scholar]	Type of source
Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). Links and impacts: the influence of public research on industrial R&D. Management science, 48(1), 1–23.	41 (0.12)	1353	Journal article
Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. Research policy, 29(2), 109–123.	34 (0.01)	2845	Journal article
Meyer-Krahmer, F., & Schmoch, U. (1998). Science-based technologies: university–industry interactions in four fields. Research policy, 27(8), 835–851.	30 (0.03)	640	Journal Article
Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: a new perspective on learning and innovation. Administrative science quarterly, 128–152.	29 (0.03)	19,751	Journal article
Rosenberg, N., & Nelson, R. R. (1994). American universities and technical advance in industry. Research policy, 23(3), 323–348.	28 (0.49)	1104	Journal Article
D'Este, P., & Patel, P. (2007). University–industry linkages in the UK: What are the factors underlying the variety of interactions with industry. Research Policy, 36(9), 1295–1313.	27 (0.01)	378	Journal article
Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). The new production of knowledge: The dynamics of science and research in contemporary societies. Sage.	26 (0.14)	9604	Book
Nelson, R. (1993). National innovation systems: a comparative analysis. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.	22 (0.02)	6049	Journal article
Jaffe, A. B. (1989). Real effects of academic research. The American Economic Review, 957–970.	20 (0.10)	2760	Journal article
Lee, Y. S. (1996). 'Technology transfer' and the research university: a search for the boundaries of university–industry collaboration. Research policy, 25(6), 843–863.	20 (0.02)	430	Journal article

Note: Citations by google scholar was also listed in this table for comparison. Nine sources came from journal articles and one source was a book.

What are the main barriers to such cooperation?) were focused on in this paper. Though Geisler developed a review about the main issues in this area, it was far away from the biggest cluster (see Fig. 4). Finally, most core papers appeared in or after 1989.

Map of timeline mode of references showed that the most important researches were concentrated in a certain time such as the year of 1989, 1994 and 2002 (cf. Fig. 5). Their themes or methods may be close. Jaffe found the significant effect of university research on corporate patents. His studies also showed that university research appeared to have an indirect effect on local innovation by inducing industrial R&D spending (Jaffe, 1989). Jaffe's work started to reveal the mechanism and motivation of the cooperation of universities and enterprises. There was no doubt that his article brought a new idea to the university–industry collaboration study. Geisler made a review about the university–industry collaboration, and gave a systematic exposition of studies in this field before 1989 (Geisler & Rubenstein, 1989). However, like the situation shown in Fig. 4, the most important achievements in this area appeared after 1989 according to our results (cf. Fig. 5). Gibbons argued that the technology and the knowledge belong to the same kind of products, so they can be exchanged in all kinds of markets. His view was so novel that it gave lots of new ideas to later scholars (Gibbons et al., 1994). In the same year, Rosenberg claimed that universities helped in improving technology progress to a large degree. As an example, he pointed out that the relations between American universities and firms may be inefficient and unstable (Rosenberg & Nelson, 1994). The contribution of his work is that he started to research the improvement policy of the university–industry collaboration beyond the argument and confirmation of the collaboration.

We also found that Cohen's paper is the biggest core in map of timeline mode in Fig. 5. The article was published in 2002, and it was also the most important literature in our analysis above. His work indicated that university research largely generates new ideas for industrial R&D projects, and the key channels through which university research impacts industrial R&D included published papers and reports, public conferences and meetings, informal information exchange, and consulting (Cohen et al., 2002). Though Cohen contributed greatly to the study on the university–industry collaboration, his work didn't contain the impact of industry on academia.
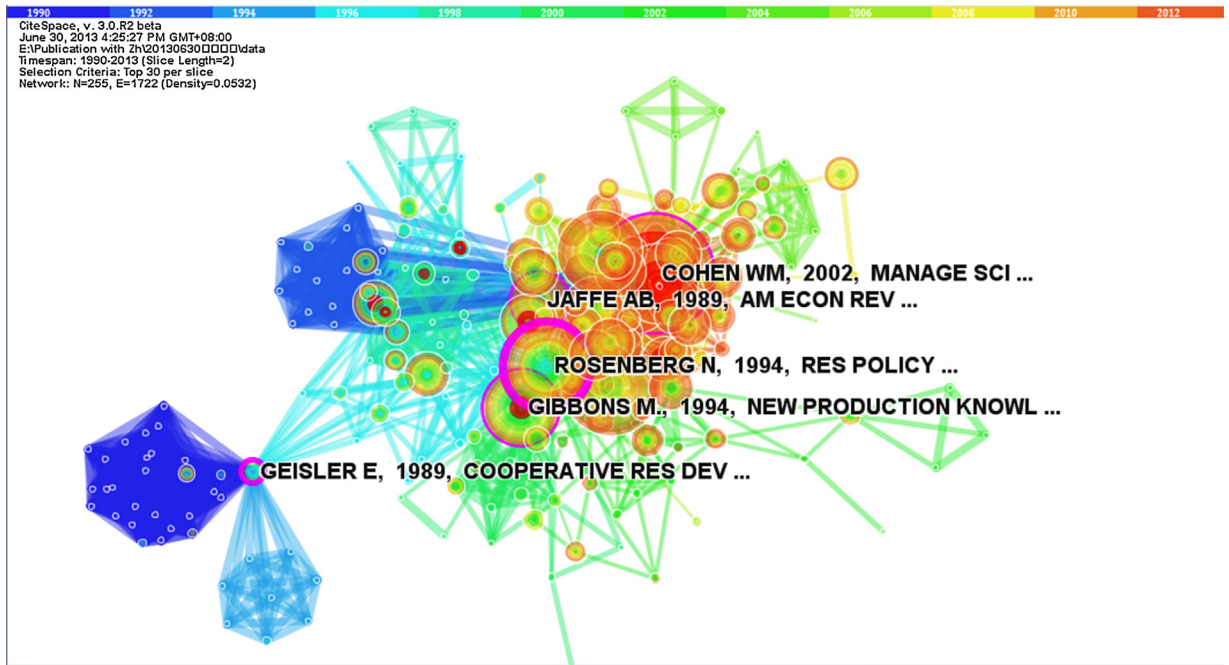
Fig. 4. Map of university–industry collaboration references. *Note*: Considering the large number of references, this paper selected the top 30 papers of every year for analysis. With the help of CiteSpace II, all papers were placed in the map as nodes.

The important ideas and papers in the field of university–industry collaboration were proposed in the 1990s, however, some important ideas were put forward only in 2000, and in 2008. Great achievements were made in this area, but it should be noted that there were no very important references after the year of 2008 (cf. Fig. 6). This means that more innovative work needs to be carried out right now and in the future. Of course, new ideas might not be confirmed so early on. Hence there may be some good ideas or important papers after 2008 (see the small clusters in 2010 and 2012) which were not
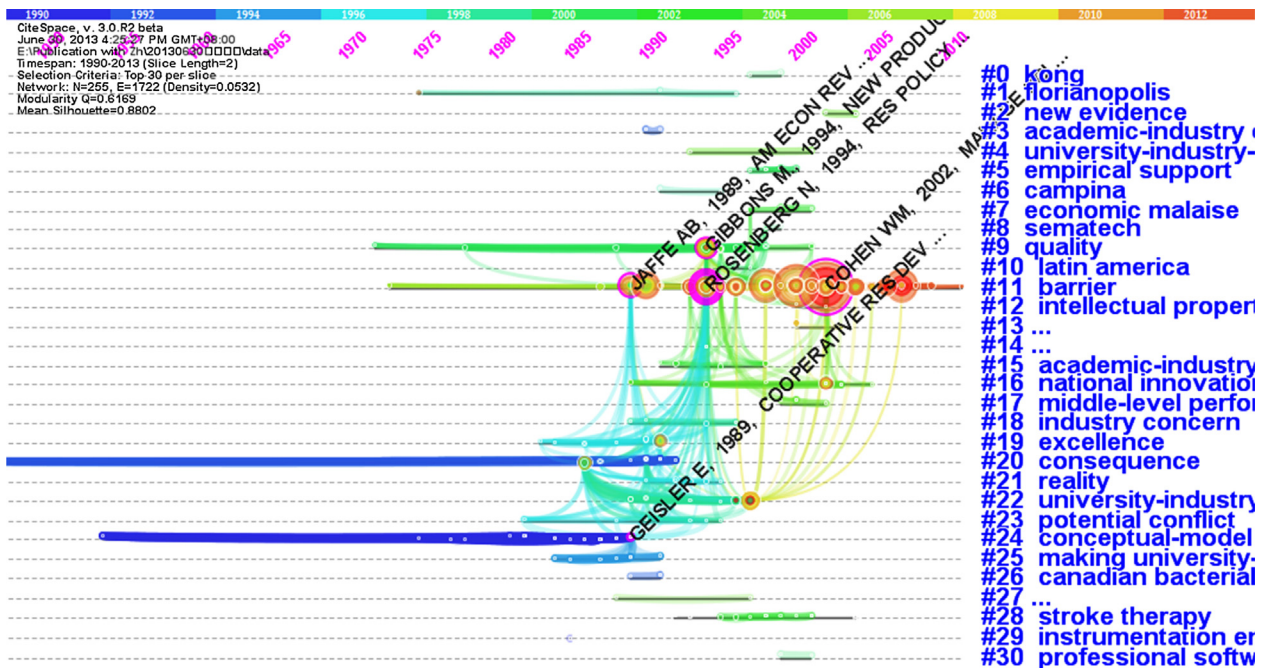

Fig. 5. Map of timeline mode of references (mapping using CiteSpace II). *Note*: Slice length was 2 and top 30 authors were selected per slice. ("slice length" is special usage in CiteSpace which means "time interval").
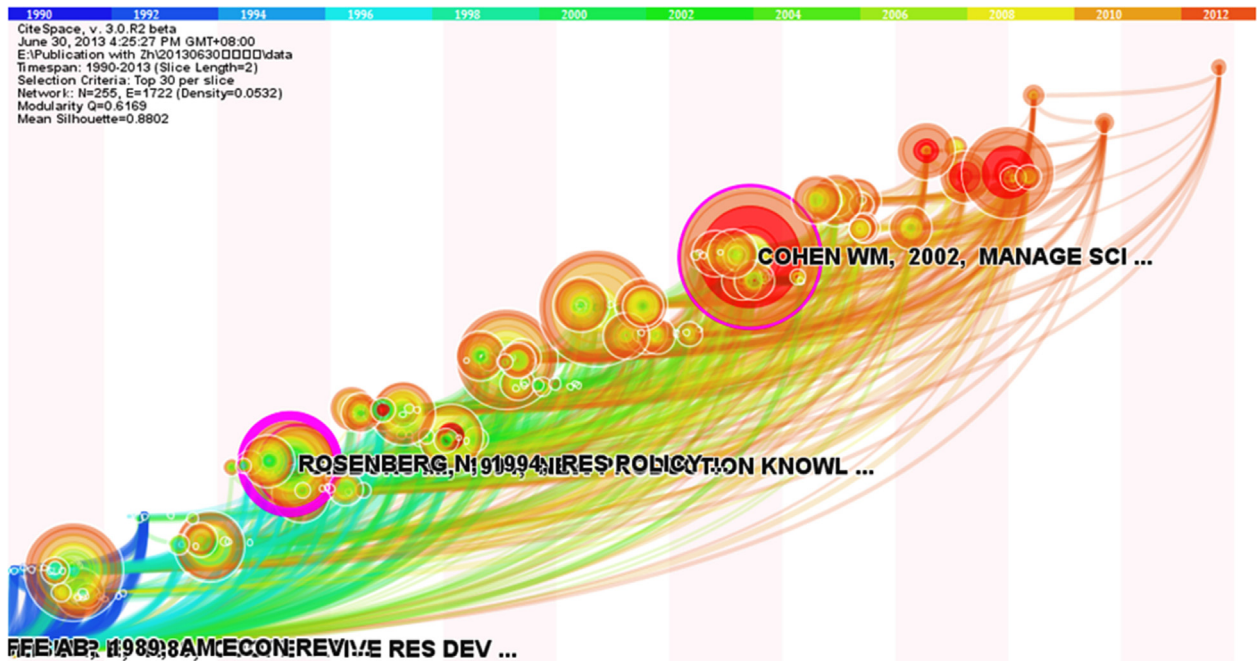
Fig. 6. Map of timeline mode of the references. *Note*: Slice length was 2 and top 30 articles were selected per slice.

referenced fully. In the future, these two nodes may be developed into cores, and we may see their influences increase. But up to now, Cohen's work in 2002 has been the core of all the references.

### 4.2. Mapping and analysis on authors and journals of references

Etzkowitz (State University of New York) was the most important author according to our results in the period 1990–2013 (cf. Table 3). His work was fully cited by the articles in our dataset (86 times) and had a great impact on other authors. The centrality of his work in the map was 0.31, quite a high score among all the authors in this area (cf. Fig. 7). By referring to the map of references we can conclude that some authors' work played an important role in this field in the past, but have not been of concern with the passage of time. The most important authors were linked closely (cf. Fig. 7). When looking at Etzkowitz's work, we found that he paid lots of attention to university–industry–government relationship. In Etzkowitz's opinion, universities play a more important role in the knowledge economy than ever before. Entrepreneurial universities can provide Knowledge Space, Consensus Space, and Innovation Space to area development. At the same time, Etzkowitz tried to find out a mechanism for universities to expand their core function from just teaching and research to include development of economy and society. Universities were an important part of the university–industry–government relationship (Etzkowitz, 2004, 2008).

Looking at Table 3, it can be concluded that authors in America were the most active researchers in the field of university–industry collaboration. 80% of the authors in the top 10 were from the US. The 10th author in the list was from Germany and his articles were cited 30 times by the papers in our dataset. Mansfield (1991) tried to estimate the extent to which technological innovations in various industries had been based on recent academic research and the time

Table 3
The top 10 authors of the references by CiteSpace II.

Author	Freq.	Centrality	Institutes
Etzkowitz H.	86	0.31	Social Science Division, State University of New York, US
Cohen W.M.	62	0.06	Camegie Mellon University, US
Mansfield E.	45	0.14	Center for Eco. and Tech., University of Pennsylvania, US
Blumenthal D.	37	0.13	Brigham and Women's Hospital, Boston, US
Mowery D.C.	35	0.01	Haas School of Business, University of California, US
Rosenberg N.	35	0.12	Department of Economics, Stanford University, Stanford, US
Etzkowitz H.	35	0.03	Social Science Division, State University of New York, US
OECD	30	0.03	Organization for Economic Co-operation and Development
Meyer-Krahmer F.	30	0.01	Fraunhofer Institute for Systems and Innovation Research, Germany
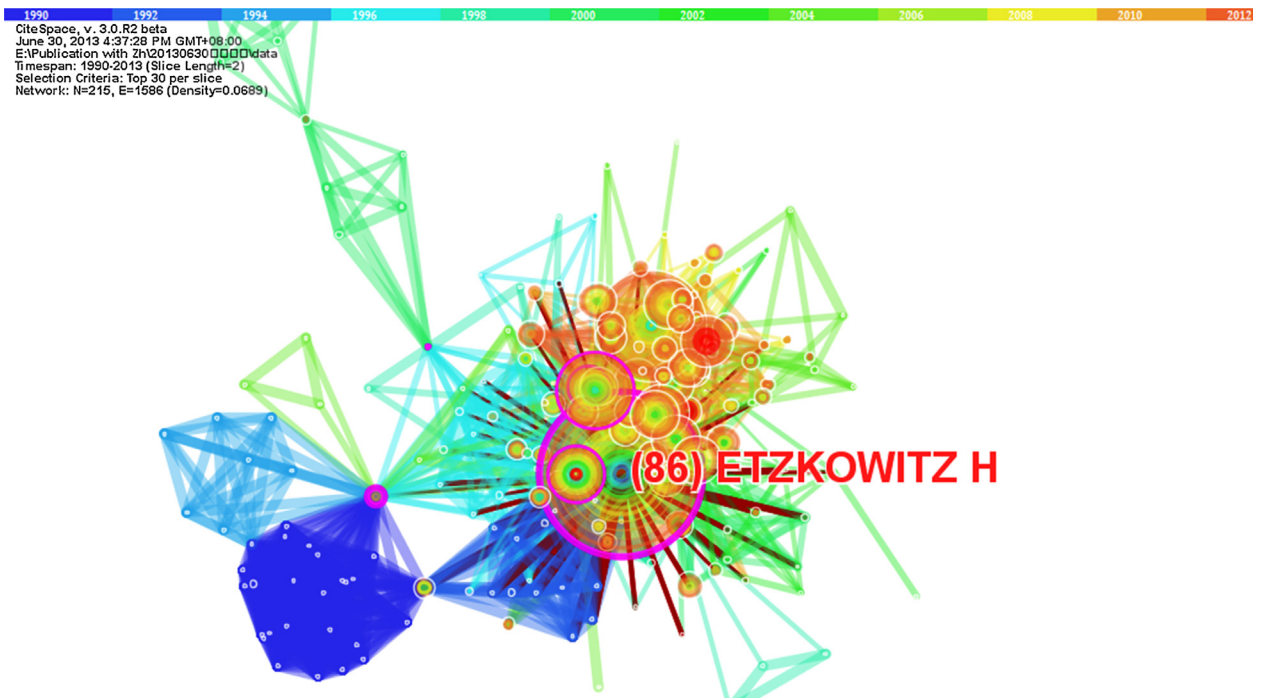
Fig. 7. Main authors studied in university–industry collaboration. (mapping using CiteSpace II). *Note*: Slice length was 2 and top 30 authors were selected per slice.

lags between the investment in research projects and the industrial utilization of their findings. Mowery claimed that there were close relations between technology and economy growth. He proposed that universities play important roles in the "knowledge-based" economies of modern industrial and industrializing states as sources of trained "knowledge workers" and ideas flowing from both basic and more applied research activities (Mowery & Sampat, 2005). Unlike American scholars, the only German scholar (Meyer-Krahmer) in the top 10 was interested in the interactions between universities and industries (Meyer-Krahmer and Schmoch, 1998). In comparing Figs. 4 and 7, we find that Cohen's paper has the highest influence during the sample period but Etzkowitz was the most influential author. The reason may be that Etzkowitz had more high-influence papers but none has excelled beyond Cohen's most influential paper.

Similar to the previous situations in Fig. 6, the most important authors in this field who appeared were from earlier years. Although the number of people who focused on the area of university–industry has been increasing year by year, the most number of core authors were concentrated in the 1990s (cf. Fig. 8). This shows that no author has carried out groundbreaking work over the last decades. Some authors did some influential work after 1998, but no author has really done anything significant beyond the erstwhile authors. Another reason may be because their research production have only been published a very short while ago, and so they have not yet been read and cited fully. Look at the left part of Fig. 8, Etzkowitz, Cohen and Blumenthal appear related closely. It means that their researches may be similar to each other in a certain period of time. The same situation applies to Rosenberg, Mansfield and Gibbons (cf. Fig. 8).

The maps of the journals on university–industry collaboration studies gave similar information as the results above. The most important journals contact each other tightly, and only a few journals were far from the main cluster core, meaning that some marginal areas had focused on the fields of university–industry collaborations studies (cf. Fig. 9). Generally speaking, the propagation of one field to other fields suggests the maturation of the total or part of the field (Such as the journal *The American journal of medicine* or *The New England Journal of Medicine* in Fig. 9). So we can conclude that researches in this area have been well applied into practice.

By careful analysis of the list of journals of the references, we found that except the specialized areas of innovation management, other fields like chemistry, industrial management, medicine and engineering were the main application areas of the study on university–industry collaboration.

According to our results of analysis on the journals of references, Science, The Journal of Technology Transfer, Research Policy, Science and Public Policy, Administrative Science Quarterly and The New England Journal of Medicine were the core journals focusing on the university–industry collaboration studies. Similar to the situations of the references and the authors, the dark parts of the map suggest that some of the journals were less concerned in recent years (cf. Figs. 9 and 10), and which also indicate the change and evolution of this area. For example, Science was the most important journal in 1990, but The Journal of Technology Transfer, Research Policy and Science and Public Policy subsequently became more and
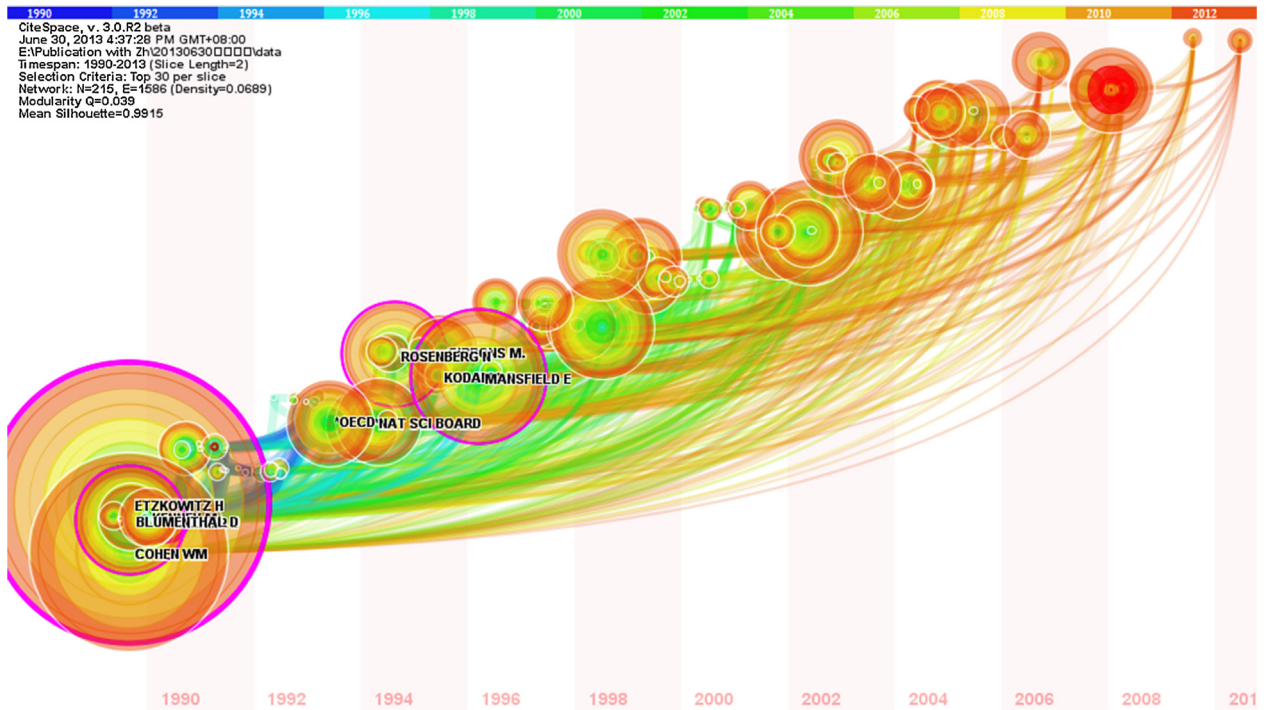
Fig. 8. Map of timeline mode of authors who studied university–industry collaboration, mapping by CiteSpace II. *Note*: Slice length was 2 and top 30 authors were selected per slice.
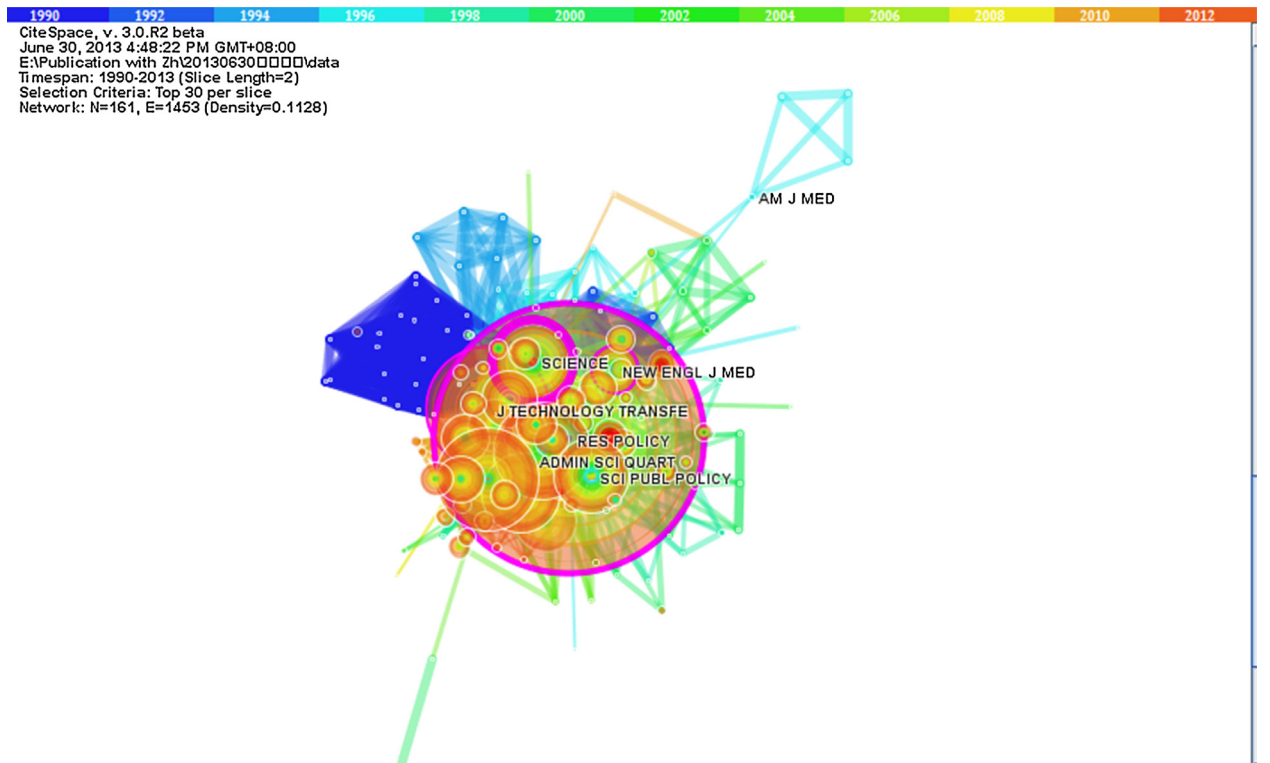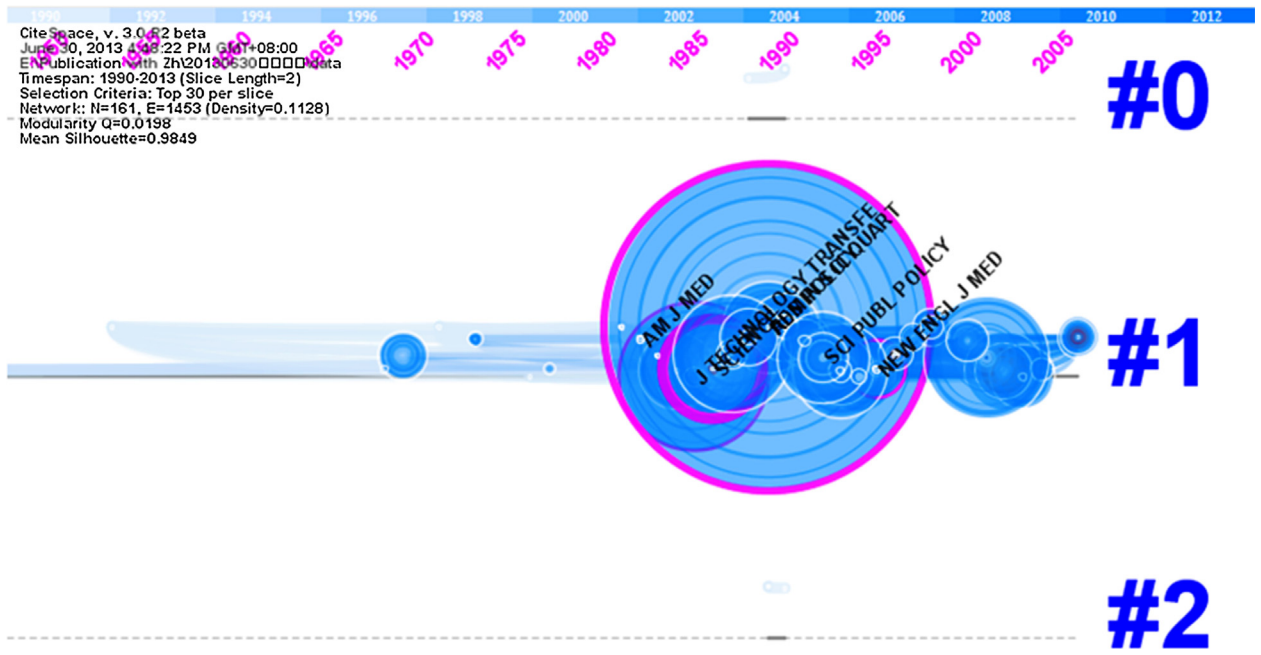

Fig. 9. Map of journals of references studied in university–industry collaboration, mapping by CiteSpace II. *Note*: Slice length was 2 and top 30 authors were selected per slice.

Fig. 10. Map of timeline mode on journals of references studied in university–industry collaboration, mapping by CiteSpace II. *Note*: Slice length was 2 and top 30 authors were selected per slice.

more important in this area. Generally speaking, papers in Science focused more on the basic theory, and journals such as The Journal of Technology Transfer, Research Policy and Science and Public Policy were concerned more with the practice. So we may infer that the theory of university–industry collaboration gradually matured and could guide the practice. This inference confirms the above conclusion (researches in the area of U–I collaboration have been well applied into practice, see the analysis of Fig. 9).

4.3. Mapping and analysis on keywords and topics of references

The analysis of keywords may yield information about the topics in the field. With the help of CiteSpace II, the top 20 keywords were chosen and listed in Table 4. As we see in the table, "innovation" was the word which appeared most frequently of all the keywords in references we selected, followed by "science" and "R&D". Similarly, words such as "technology", "transfer", "knowledge", "collaboration" and "networks" also took a lot of space. In the light of the these results, we can state that regardless of whether they were formal or informal, topics of university–industry collaboration have focused a lot on the basis of the collaborations such as the characteristics of the firms and universities or ways to cooperate. Some specialized disciplines appeared in the list, such as biotechnology. In some sense, they may be important applied areas of theory in university–industry collaboration.

Table 4
Top 20 keywords of the references of papers in our dataset.

Freq.	Centrality	Keyword	Freq.	Centrality	Keyword
67	0.23	Innovation	17	0.12	Collaboration
41	0.08	Science	14	0.03	Performance
39	0.06	Research-and-development	14	0.01	Knowledge transfer
30	0.15	Knowledge	12	0.01	Networks
36	–	Technology-transfer	12	0.03	Absorptive-capacity
30	–	Firms	11	0.01	Japan
23	0.32	Biotechnology	10	0.00	Spillovers
20	0.03	University–industry Collaboration	10	0.00	Patents
20	0.22	Technology	10	0.01	Systems
18	0.10	Academic research	10	0.01	Intellectual property

Note: Some keywords referred to the same thing (such as firms and firm, technology transfer and technology-transfer) but the program cannot recognize this anomaly, so we merged them manually. Their data of centrality were lost as a result of the merger.
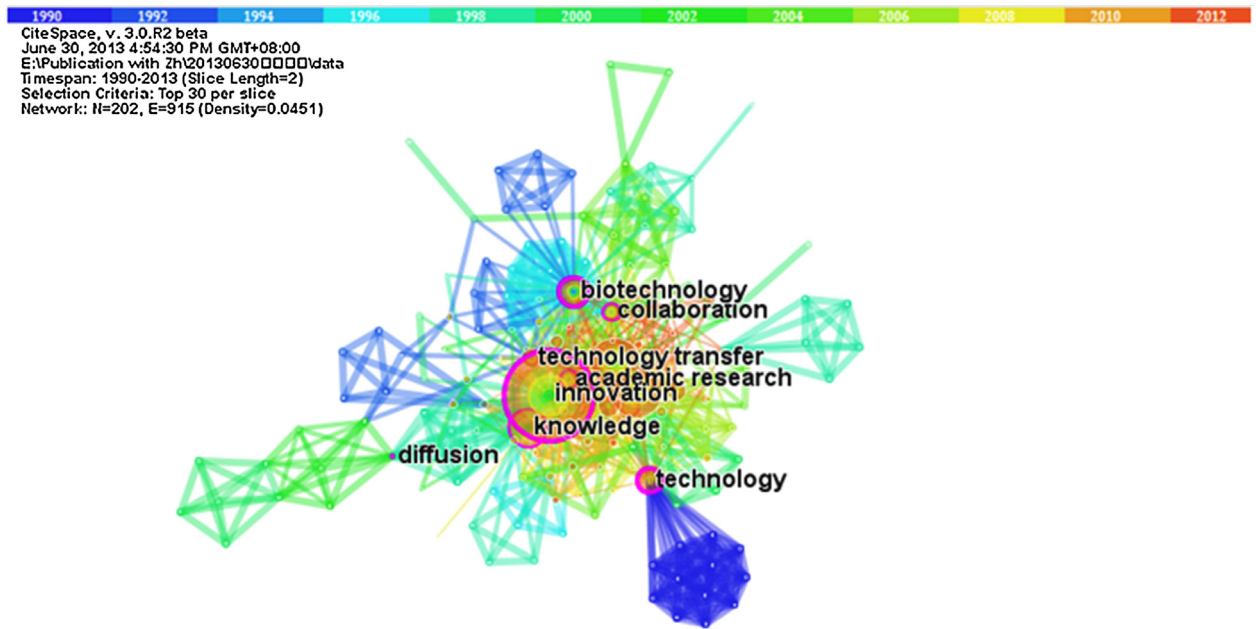
Fig. 11. Map of keywords of references studied in university–industry collaboration, mapping by CiteSpace II. *Note*: Slice length was 2 and top 30 authors were selected per slice.

When all the keywords were mapped in Fig. 11, we were able to assess the collaborations of the keywords of the references. It is very clear that the most important keywords were connected closely, which meant that main researches in this field were interlinked. All the keywords appeared in Fig. 11 connected tightly to the knowledge of university–industry studies. The branches of the core in the figure represented other specialized areas which have collaborations with university–industry studies, such as "biotechnology", "chemistry" or "medicine". These areas were important but not the core studies of the theory of university–industry collaboration. The core keywords were "technology transfer" (technology transfer or patent transfer from universities to enterprises, and a common university–industry collaboration form between universities and industries), "academic research" (the important function of universities), "innovation" (a comprehensive cooperation process between universities and industries). "Knowledge" and "technology" were the content in the technology transfer. Some topics, such as "diffusion" and "technology", had formed branches. It meant that some new theory had been proposed and developed to a certain extent, but they did not affect the previous theory. Of course, it can also mean that correlation of some researches were not strong.

To find out the trend of the studies on university–industry cooperation, we divided the reference dataset by time for analysis. Articles in the period 1990–2009 and 2010–2013 were studied respectively. Within the changes of frequency rank and centrality of keywords, the trends of topics can be sought out. Since we need to analyze change of centrality, the same meaning words are not merged in this part.

Changes of the keywords gave lots of interesting information on the trends in this field. The top No. 1 keyword was "innovation", whether for the period 1990–2009 or 2010–2013. It meant that innovation was the most close and most active topic in field of university–industry collaboration. Some keywords were important in the period 1990–2009, but they disappeared from the top 20, even top 100 keywords in 2010–2103, such words as "intellectual property", "policy", "Japan", "commercialization", "absorptive-capacity", "collaboration". It means that these topics had been popular in the past, but have lost import in recent years.

On the other hand, some keywords did not appear in the top 20 topics in the period 1990–2009 but they appeared in the top 20 topics in 2010–2013, which means these topics may turn out to be the new hot topics currently and in the near future. Keywords such as "firms", "technology", "R&D", "networks" belonged to these topics (cf. Fig. 12). Many researchers claimed that firms or enterprises were the main part of the innovation system (Perkmann & Walsh, 2007; Powell & Grodal, 2005). So the topic of firms assumed more importance in recent years and we believe that in the very near future, the topic about firms in university–industry collaboration studies will become hotter. Other topics such as "networks", "R&D" may also heat up with the development of complex collaborations between large numbers of universities and firms.

If the centralities of the keywords in the periods 1990–2009 and 2010–2013 are considered, some interesting changes of this field are worth mentioning (cf. Fig. 13). Firstly, the changes shown in the figure are very similar to the ones in Fig. 12. This indicates that our results are stable in some sense. Secondly, centralities of some keywords such as "technology" or "collaboration" became 0.00 after the year 2009. This indicates that basic theories and concepts of university–industry
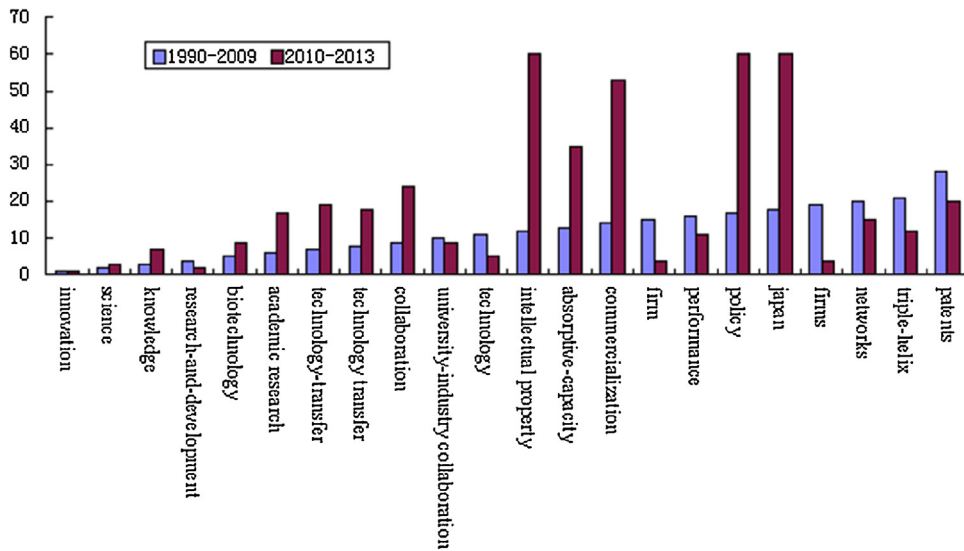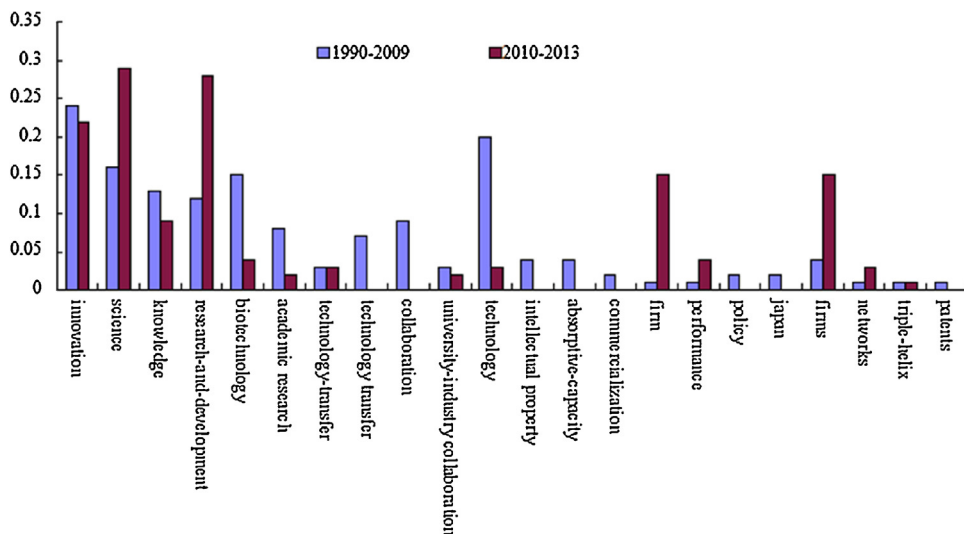
Fig. 12. Change of the rank of the keywords. *Note*: Since the merger of same words may lose the centrality of the keywords, we did not merge words such as firms and firm. Keywords of the period 1990–2009 were listed in the figure one by one. So by just comparing the height of the graph, we may find out the trend of topics in a field. If the rank falls rapidly, it may illustrate that this topic is not of much concern now as it was in the past (the histogram on the right is longer than the one on the left). Similarly, if the rank rises rapidly, then it may be concluded that this topic is a hot topic now (the histogram on the left is longer than the one on the right).

collaboration have developed well. Lastly, centrality of some topics such as "firms", "R&D", "networks" went up rapidly after 2009, which means that these topics may be the new core of studies in this field.

5. Summary of case study

Firstly, ideas for university–industry collaboration had been proposed for a long time, and can even be traced to the 1960s. Leydesdorff had published the largest number of papers focused on the university–industry collaboration during the sample period, and *Abstracts of papers of the American Chemical Society* had published most publications in this field during this time. Some new theories had appeared but they had not become the cores yet.

Fig. 13. Change of the centrality of the keywords. *Note*: Since the merger of same words may lose the centrality of the keywords, we did not merge words such as firms and firm. By comparing the high degree of the trend of core topics in a field. If the centrality falls rapidly, it may illustrate that this topic is no longer a core topic as it was. (The histogram on the left is longer than the one on the right in the figure). Similarly, if the centrality rises rapidly, it can be concluded that this topic is more a central topic now than it was in the past (the histogram on the right is longer than the one on the left in the figure).

Cohen's publication published in *Management Science* was the most important paper by far in all the references of the field of university–industry collaboration studies. The most important publications in terms of references on our dataset connected to each other closely, and they almost all appeared before 2008. After 2008, no very important papers appeared (or some important papers may have been published in this area, but they are not cited fully yet because they cannot be recognized by CiteSpace).

All important cited authors were linked tightly. Etzkowitz was the most important cited author and he enjoyed the highest centrality in all the authors of references. American authors have contributed the most to this area up to now and they are the most active researchers in the world. The most important cited authors appeared in the 1990s after which no breakthrough has been made in this field for over two decades (During the sample period, some new findings could also be important but they need more time to be cited and recognized).

Science, The Journal of Technology Transfer, Research Policy, Science and Public Policy, Administrative Science Quarterly and The New England Journal of Medicine were the core journals that focused on the university–industry collaboration studies in the sample period. The achievements and theories of university–industry collaborations have been applied in some other areas such as chemistry, industrial management, medicine and engineering. This is an indication that the U–I field has come of age in some sense.

"Innovation" was the hottest topic in the field of university–industry collaborations. Topics on "firms", "technology", "R&D", "networks" may be the prospective hottest research topics in the near future. Conversely, topics about "technology" or "collaboration" may be of less concern than they were in the past.

6. Conclusions

Virtualization study in bibliographic databases can provide us the performance and trends in certain fields. Maps contain much more information than words or data. Maps of literatures allow deeper analysis. CiteSpace II is a very useful program for virtualization study in bibliographic databases. This paper has attempted to develop a framework of visualization and quantitative research via CiteSpace II by taking the field of university–industry collaboration as a case. Interesting results were obtained with the help of CiteSpace II. The results show that our framework can deal with the literature data well.

Unlike words and data, more information can be given by visualization maps. Principles and algorithms are discussed in this paper. Publications and trends in the field of university–industry collaboration studies were analyzed via CiteSpace II. The most important articles, references, authors and keywords in the university–industry studies were picked out for analysis during the sample period, and the development process as well as the current status of this field were discovered. Connections between important authors, journals, references and keywords were evaluated in our work. We found that some theories of university–industry collaborations were mature in some sense, and they have been well applied in other areas.

By comparing the top 20 key words in the period from 1990 to 2009 with the top 20 in 2010–2013, we identified the change of the researchers' attentions easily. With the use of CiteSpace II, figures were worked out, moreover structures, links of references, authors and keywords were extracted. From all the information shown in the figures we claim that the study on university–industry collaboration has formed some cores but no breakthrough has been made after the year of 2008 (or some breakthroughs have been made but they have not been properly appreciated yet. A lot of work can be done in the near future in this field.)

According to our study, creative work will be the most important task in the foreseeable future, and studies on the applications of the university–industry collaboration theory will also be valuable. Having demonstrated the powerful function of CiteSpace II in this paper, more work can be carried out in other fields using the program. Visualization and quantitative study or evaluation of performance in bibliographic databases of some more areas will prove very useful in the authors' opinion.

It may be noted that there are some shortcomings in our research. Firstly, the dataset was not comprehensive and many important journals were excluded. Though cited references of papers in our dataset can describe situations quite well, we lost some opportunities to process those papers directly. Besides, figures made by CiteSpace II cannot contain all data because of its computational capability, so some minor details could be ignored.

Acknowledgements

References

Agrawal, A. (2001). University-to-industry knowledge transfer: Literature review and unanswered questions. *International Journal of Management Reviews, 3*, 285–302.

Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, *54*(6), 550–560.

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*(1), 47.

Alice Lam. (2007). Knowledge networks and careers: Academic scientists in industry-university links? *Journal of Management Studies*, *44*(6), 993–1016.

Aurora, A. C. T., & Luisa, M. (2012). A bibliometric portrait of the evolution, scientific roots and influence of the literature on university–industry links. *Scientometrics*, *93*, 719–743.

Bensman, S. J. (2004). Pearson's r and author cocitation analysis: A commentary on the controversy. *Journal of the American Society for Information Science and Technology*, *55*(10), 935.

Cao, Z., Zhang, L., Feng, F., & Du, Y. (2013). Study on the characteristics and cultivating path of the industry–university symbiotic networks: Based on the small-world network model and the theory of symbiosis? *Asian Social Science*, *9*(1), 15–21.

Chen, C. (2003). *Mapping scientific frontiers*. London: Springer.

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences*, *101*(Suppl.), 5303–5310.

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.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). Links and impacts: The influence of public research on industrial R&D. *Management Science*, *48*(1), 1–23.

D'Este, P., & Patel, P. (2007). University–industry linkages in the UK: What are the factors underlying the variety of interactions with industry. *Research Policy*, *36*(9), 1295–1313.

Diem, A., & Wolter, S. C. (2013). The use of bibliometrics to measure research performance in education sciences. *Research in Higher Education*, *54*(1), 86–114.

Etzkowitz, H. (2004). The evolution of the entrepreneurial university. *International Journal of Technology and Globalisation*, *1*(1), 64–77.

Etzkowitz, H. (2008). *The triple helix: University–industry–government innovation in action*. London: Routledge.

Etzkowitza, H., & Leydesdorffb, L. (2000). The dynamics of innovation: From National Systems and Mode 2 to a Triple Helix of university–industry–government relations. *Research Policy*, *29*(2), 109–123.

Feng, F., Zhang, L., Du, Y., Ma, L., & Fu, M. (2012). Study on the classification and stability of industry–university–research symbiosis phenomenon: Based on the logistic model? *Journal of Emerging Trends in Economics and Management Sciences*, *3*(1), 116–120.

Goddard, W., & Dasgupta, S. (1992). The materials and molecular simulation center of the Beckman Institute at Caltech – A model for university–industry–government research collaborations. *Abstracts of Papers of the American Chemical Society*, *204*(16), 79–80.

Geisler, E., & Rubenstein, A. H. (1989). University–industry relations: A review of major issues. In *Cooperative research and development: The industry–university–government relationship*. The Netherlands: Springer.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: Sage.

Hou, J., & Hu, Z. (2013). Review on the application of CiteSpace at home and abroad. *Journal of Modern Information*, *33*(4), 99–103.

Hancock, R. E., & McCurry, T. E. (1993). Canadian Bacterial Diseases Network: A new approach to university–industry relationships. *Clinical and Investigative Medicine*, *16*(4), 306–313.

Jaffe, A. B. (1989). Real effects of academic research. *American Economic Review*, *79*(5), 957–970.

Jonas, L., Göran, T., Inger, L., John, S., & Mats, B. (2006). Collaboration uncovered: Exploring the adequacy of measuring university–industry collaboration through co-authorship and funding. *Scientometrics*, *69*(3), 575–589.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, *7*(1), 48–50.

Kruskal, J. B., & Wish, M. (1978). . *Multidimensional scaling* (Vol. 11) London: Sage.

Kuhn, T. S. (2012). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kumar, N. (1994). Industry–university symbiosis. *International Journal of Engineering Education*, *10*, 127–132.

Lee, Y. S. (1996). 'Technology transfer' and the research university: A search for the boundaries of university–industry collaboration. *Research Policy*, *25*(6), 843–863.

Leydesdorffb, L., & Etzkowitza, H. (1996). Emergence of a Triple Helix of university–industry–government relations. *Public and Science Policy*, *23*(5), 279–286.

Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology*, *57*(12), 1616–1628.

Leydesdorff, L., & Zaal, R. (1988). *Co-words and citations relations between document sets and environments*. Elsevier.

Ma, F., & Xi, M. (1992). Status and trends of bibliometric? *Journal of Information Science*, *13*(5), 7–17.

Mansfield, E. (1991). Academic research and industrial innovation. *Research Policy*, *20*(1), 1–12.

Meyer-Krahmer, F., & Schmoch, U. (1998). Science-based technologies: University–industry interactions in four fields. *Research Policy*, *27*(8), 835–851.

McMillan, H. E., & Hamilton, R. D. (2003). The impact of publicly funded basic research: An integrative extension of Martin and Salter. *IEEE Transactions on Engineering Management*, *50*(2), 184–191.

Miguel, S., Moya-Anegón, F., & Herrero-Solana, V. (2008). A new approach to institutional domain analysis: Multilevel research fronts structure. *Scientometrics*, *74*(3), 331–344.

Mowery, D. C., & Sampat, B. N. (2005). *Universities in national innovation systems*. The Oxford Handbook of Innovation.

Nelson, R. (1993). *National innovation systems: A comparative analysis*. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.

Nelson, R. R. (2001). Observations on the post-Bayh-Dole rise of patenting at American universities. *Journal of Technology Transfer*, *26*(1–2), 13–19.

Perkmann, M., & Walsh, K. (2007). University–industry relationships and open innovation: Towards a research agenda. *International Journal of Management Reviews*, *9*(4), 259–280.

Pickard, R. H. (1944). Professional Service in universities technical colleges and industry. *Nature*, *154*(July–December), 832.

Powell, W. W., & Grodal, S. (2005). *Networks of innovations*. London: The Oxford Handbook of Innovation.

Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Technical Journal*, *36*(6), 1389–1401.

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, *25*(4), 348–349.

Rosenberg, N., & Nelson, R. R. (1994). American universities and technical advance in industry. *Research Policy*, *23*(3), 323–348.

Shane, S. A. (2005). *Economic development through entrepreneurship: Government, university and business linkages*. Cheltenham: Edward Elgar.

Siegel, D. S., Waldman, D., & Link, A. (2003). Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: An exploratory study. *Research Policy*, *32*, 27–48.

Small, H. (2003). Paradigms, citations, and maps of science: A personal history. *Journal of the American Society for Information Science and Technology*, *54*(5), 394–399.

Synnestvedt, M. B., Chen, C., & Holmes, J. H. (2005). *CiteSpace II: Visualization and knowledge discovery in bibliographic databases*. AMIA annual symposium proceedings (Vol. 2005) American Medical Informatics Association.

Thune, T. (2007). University–industry collaboration: The network embeddedness approach? *Science and Public Policy*, *34*(3), 158–168.

Thursby, J. G., Jensen, R. A., & Thursby, M. C. A. (2001). Objectives, characteristics and outcomes of university licensing: A survey of major us universities. *Journal of Technology Transfer, 26*(1), 59–72.

Waltman, L., & Eck, N. J. V. (2007). Some comments on the question whether co-occurrence data should be normalized. *Journal of the American Society for Information Science and Technology, 58*(11), 1701–1703.

White, H. D. (1990). Author co-citation analysis: Overview and defense. *Scholarly Communication and Bibliometrics, 84*, 106.

White, H. D. (2003). Author cocitation analysis and Pearson's r. *Journal of the American Society for Information Science and Technology, 54*(13), 1250–1259.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*(4), 327–355.