An Integrated Approach for Main Path Analysis: Development of the Hirsch Index as an Example

John S. Liu
Graduate Institute of Technology Management, National Taiwan University of Science and Technology 43, Section 4, Keelung Road, Taipei 10607, Taiwan. E-mail: johnliu@mail.ntust.edu.tw

Louis Y.Y. Lu
College of Management, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li 32003, Taiwan. E-mail: louislu@saturn.yzu.edu.tw

This study enhances main path analysis by proposing several variants to the original approach. Main path analysis is a bibliometric method capable of tracing the most significant paths in a citation network and is commonly used to trace the development trajectory of a research field. We highlight several limitations of the original main path analysis and suggest new, complementary approaches to overcome these limitations. In contrast to the original local main path, the new approaches generate the global main path, the backward local main path, multiple main paths, and key-route main paths. Each of them is obtained via a perspective different from the original approach. By simultaneously conducting the new, complementary approaches, one uncovers the key development of the target discipline from a broader view. To demonstrate the value of these new approaches, we simultaneously apply them to a set of academic articles related to the Hirsch index. The results show that the integrated approach discovers several paths that are not captured by the original approach. Among these new approaches, the key-route approach is especially useful and hints at a divergence–convergence–divergence structure in the development of the Hirsch index.

Introduction

Garfield, Sher, and Torpie (1964) suggested that it is possible to "write the history of science" through analyzing citation relationships among science publications. This is based on the assumption that "the history of science is regarded as a chronological sequence of events in which each new discovery is dependent upon earlier discoveries" (p. iii). They demonstrated the idea by drawing and analyzing the "topological network diagrams" (p. iii) of a set of 40 DNA publications. Their idea of tracing discoveries has carried on and has become more convincing and more practicable through the efforts of many subsequent researchers.

In modern-day terminology, the "topological network diagrams" is a citation network where the nodes represent publications and the links represent the citation relationships among these publications. Garfield et al. (1964) established a citation network of DNA publications and then traced the significant publications based on their citation counts. The set of DNA publications that they investigated was small (40 publications) such that the task was manageable. Presently, the accumulated publication set for most of science is relatively large, and without some sort of quantitative method, it would not be possible to "write the history of science" from such a maze of a large citation network.

Hummon and Doreian (1989) addressed the issue by introducing main path analysis. The method helps to escape from the maze by offering the most significant trajectories—the main paths—of a citation network. This is done by first assigning a significance index to each citation link. The index is defined as the accumulated traversal count of a link, assuming that knowledge embedded in earlier works is diffused to works published at a later date. After the significance index is determined, one then conducts a "priority first search" algorithm to trace the significant subsequent followers of an earlier publication. The paths thus obtained are the most significant path of the target scientific field. Hummon and Doreian demonstrated the method on the same DNA citation network discussed in Garfield et al. (1964), and their results are quite convincing.

To conduct a more serious examination of the method, Hummon, Doreian, and Freeman (1990) applied main path analysis to a larger sample—the citation network of the centrality–productivity literature—and later to the social network analysis field (Hummon & Carley, 1993) as well as to the conflict-resolution field (Carley, Hummon, & Harty,

Received July 16, 2011; revised September 14, 2011; accepted October 3, 2011

 • Published online 18 November 2011 in Wiley Online Library (wileyonlinelibrary. com). DOI: 10.1002/asi.21692

1993). Methodologically, Batagelj (2003) proposed a major advancement of the method, enhancing main path analysis by offering efficient algorithms for determining various versions of the significance index. The efficient algorithms were implemented in Pajek, a public-domain social network analysis software (Batagelj & Mrvar, 1998).

Many researchers have applied the main path analysis method since then to investigate the developmental trajectories of various science and technology domains using bibliographical citation data, patent citation data, or both. Moore, Haines, Hawe, and Shiell (2006) examined the genealogy of the concept of social capital in public health. Verspagen (2007) traced the development trajectories of fuel cell technology using patent citation data. Mina, Ramlogan, Tampubolon, and Metcalfe (2007) discovered the growth and transformation of coronary artery disease treatment using both bibliographic and patent citation data. Carlero-Medina and Noyons (2008) studied the development path of the absorptive capacity research field. Lucio-Arias and Leydesdorff (2008) combined main path analysis with HistCite and path-dependent transitions methods in the study of fullerenes and nanotube technology. Harris, Luke, Zuckerman, and Shelton (2009) examined the discovery and delivery literature in secondhand smoke by applying the main path analysis. More recent studies using the method include those by Liu, Lu, Lu, and Lin (in press) on the development path of data envelopment analysis and by Lu, Lin, Liu, and Yu (2011) on the research history of the ethics of nanotechnology.

There are several benefits to applying main path analysis to a citation network. First, it simplifies a complicated citation network to a small number of nodes and links. The analysis provides a satellite view to a given citation network. Under such a view, the paths are like roads on the ground, and only the most significant paths remain whereas paths of lesser significance disappear. Second, it highlights a sequence of major historical-development events, which is very useful to scientists who are considering entering into a particular science and technology domain. Third, it identifies works standing at an important juncture of a field's historical development. These works identified through the main path analysis can be different from those works that have a high citation count. The citation count reckons direct influences while the main path analysis takes indirect influences into account as well.

The method, however, has some limitations, as it either offers a complex "network of main paths," as suggested in Hummon et al. (1990), or one and only the most significant development path, as implemented in Pajek software. For a large citation network, the "network of main paths" does not achieve the goal of simplifying the citation network. On the other hand, the one and only path, which limits our view to the target science or technology domain, is not satisfactory for explorers who are looking for more than the most significant development path. Furthermore, the "priority first search" algorithm traces the most significant route at each juncture when several new ideas are competing. The result obtained may not be the path with the largest overall impact. Indeed, many close-second significant routes also can be neglected in the process.

To our knowledge, there are no studies in the literature which address the aforementioned limitations. This study fills the gap and contributes to the main path analysis literature by proposing several variants of the original method, and suggests that simultaneously applying all these complementary approaches will greatly expand our perspective in exploring a target science and research domain. The variants we propose include: global main path, backward local main path, multiple main paths, and key-route main paths. Each of them provides one or multiple development trajectories of particular significance.

The next section reviews the original main path method and elaborates on its limitations. For each limitation, a solution to improve the original method in that aspect is proposed. We then present a procedure to integrate these new approaches and suggest that integrating these complementary variants can vastly increase our view to the target science or technology domain. Demonstrating the value of the new approach, we apply it to a set of Hirsch index (Hirsch, 2005) papers to explore the history of the Hirsch index research.

Main Paths

We first revisit the main path method as proposed by Hummon and Doreian (1989) and the contributions from Batagelj (2003) in this section. The discussion then proceeds to potential limitations of the method and solutions to overcome these limitations.

Revisiting the Main Path Analysis

In a citation network, knowledge is suggested to flow from cited nodes to citing nodes. The links among these nodes are the conduits of knowledge flow. Given any arbitrary node, the idea proposed disseminates through the conduits until it hits an end node. A sequence of conduits that link the given node to an end node is called a "search path." There can be multiple search paths for any given node, and the significance of each search path can vary. One defines the main path as the most significant search path among all search paths. In other words, the main path is the most important sequence of conduits that spreads the knowledge out from this arbitrary node.

Hummon and Doreian (1989) provided a method to identify the main path. The contribution of their work is twofold. First, they proposed three types of "traversal counts" as the significance index for each link in a citation network. Second, based on these traversal counts, a "priority first search" algorithm is recommended to construct the main path. In other words, they proposed to identify a main path in two steps. First, translate the binary citation network into a weighted network, with the weight of each link indicating the significance of the link. Second, apply some sort of search algorithm to construct the main path.

DOI: 10.1002/asi

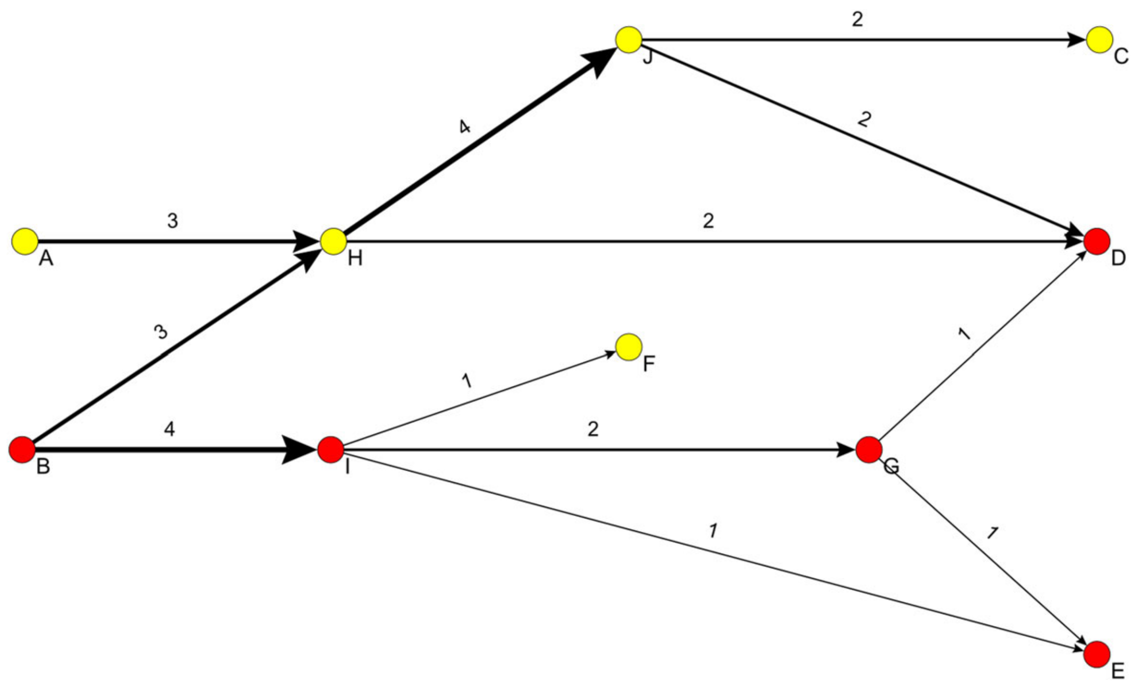

FIG. 1. A simple citation network. [Color figure can be viewed in the online version, which is available at wileyonlinelibrary.com.]

Traversal counts measure the times a citation link has been traversed if one exhausts the search from a set of start nodes to another set of end nodes. The three proposed types of traversal counts are search path link count (SPLC), search path node pair (SPNP), and node pair projection count (NPPC). The SPLC and SPNP are calculated based on all possible search paths emanating from a start node whereas the NPPC is calculated for a subnetwork defined by a start node and an end node. The logic behind using these traversal counts as the significance index is that if a citation link occupies a route through which much knowledge flows, it has to have a certain importance in the knowledge-dissemination process. Furthermore, the nodes on the significant routes also can be inferred to possess important knowledge.

Narin, Pinski, and Gee (1976) proposed to assign an "influence weight" to a publishing entity. The concept first obtains a matrix, where each matrix element contains the number of mutual references among entities under consideration. It then takes the eigenvalue of the normalized matrix to obtain the "influence weight" of each publishing entity. The main path method does not directly assign values to the publishing entities but rather assigns values to citation links between pairs of publishing entities. In other words, the main path method identifies significant "links" rather than important "nodes." The nodes on the significant links are interpreted to nonetheless have certain importance.

More than a decade later, Batagelj (2003) made a significant contribution by offering efficient algorithms for determining the SPLC and SPNP. Batagelj also proposed a new traversal count—the search path count (SPC)—and recommended the SPC over the SPLC and SPNP because of its "nice" properties. In this study, we do not elaborate on the pros and cons of applying each of the traversal counts but follow the recommendation and apply the SPC throughout.

We use a simple citation network in Figure 1 to demonstrate how the SPC for each individual link in a citation network is calculated. One defines a "source" as a node that is cited, but cites no other nodes, and a "sink" as a node that cites other nodes, but is not cited. In other words, sources are the origins of knowledge while sinks are the end points of knowledge dissemination. The network in Figure 1 has two sources, A and B, and four sinks, C, D, E, and F. There are many alternative paths to go from the sources to the sinks. Assuming that one exhausts all efforts in searching out all paths from all the sources to all the sinks, the SPC for each link is defined as the total number of times the link is traversed. For example, Link J-C has an SPC value of 2 because paths A–H–J–C and B–H–J–C pass through it. Link B-I's SPC value is 4 because it is traversed by four paths: B–I–F, B–I–G–D, B–I–G–E, and B–I–E. In the example network, B-I and H-J have the largest SPC value. The larger the SPC value, the more significant the link's role is in transmitting the knowledge.

Regarding search algorithms, the "priority first search" algorithm suggested by Hummon and Doreian (1989) presents that at any node, one always chooses the next link in the path with the highest traversal count as the outgoing link. By applying the choice rule repeatedly until hitting a terminal node, a main path is constructed. Hummon and Doreian tested the proposed method with a DNA citation network prepared by Garfield et al. (1964). The resulting main path for DNA development was checked against two other pieces of evidence: (a) the important DNA works suggested by Asimov (1963) and Garfield et al. (1964), and (b) the central core identified through Q-analysis presented in Hummon and Doreian (1989). Both studies have shown that the nodes on the main path do indicate significant DNA works.

There are many main paths in reality, one for each non-terminal node. For a small citation network such as the DNA network mentioned earlier, the number of main paths is small and relatively easy to manage. For a large citation network, Hummon et al. (1990) indicated that one way to examine the structure of these main paths is to construct a "network of main paths," a network obtained by merging all the main paths together. However, this network of main paths still consists of a large number of nodes. To further simplify this network, Hummon and Carley (1993) proposed the "main path tie frequency" and the "main path endpoint frequency" measures. Tie frequency is the number of time a route (tie) occurs across all main paths. Endpoint frequency is the number of main paths that terminate in an end node. These two measures can be used to identify the most important main paths.

The aforementioned approach seems to be complicated, but a widely available and simpler approach has been implemented in Pajek software (Batagelj & Mrvar, 1998). It finds a one and only path by beginning the search from all nodes that cite no others and picks the link with the largest traversal count emanating from these nodes as the start link. It then chooses the next link in the path with the highest traversal count as the outward link. This method is the one we have adopted and is the departing point of our integrated approach.

Next, we will summarize the "classical" main path method, as implemented in Pajek software and following the terminologies used in Verspagen (2007). One begins with the identification of all sources and sinks in the target citation network and then calculates traversal counts for each link by applying the method described previously. Based on the traversal counts for each citation link, the main path can be defined through the following procedures.

1. Find the link with the largest traversal count from all possible links emanating from all the sources. Assign the beginning node of this link as the start point of the main path. Take the end node of the link as the start point for the next step. If there are ties, take all the tied links into consideration.
2. Find the link with the largest traversal count emanating from the current start point(s). Take the end node(s) of the link(s) as the start point(s) for the next step. If the end node is a sink, stop. If there are ties at each start point, take all the tied links into consideration.
3. Continue Step 2 until all the paths hit a sink.

The aforementioned procedure begins the main path search from all the sources. The intention is to find the single most significant path for the whole network. One also could begin the search from a specific node to identify the most significant path emanating from that particular node. In that case, Step 1 should begin with "Find the link with the largest traversal count from all possible links emanating from a designated node." This would find one of the main paths in the "network of main paths."

The single main path method described earlier suffers from several limitations. First, the path resulting from a priority first search is local; there is no guarantee that this path is the most significant path among all paths in the whole network. Second, searching forward finds the nodes that attract many followers, but what if one wants to find the significant nodes that bring together ideas from many earlier publications? Third, only one single path is acquired. This single path does not serve our need for exploring the secondary paths. Fourth, the link with the largest SPC may not be in the final main path; this is a more serious problem. The remaining portion of this section addresses these limitations and proposes corresponding solutions.

Global Versus Local

The priority first search algorithm as proposed in Hummon and Doreian (1989) is a "local" approach. It repeatedly chooses the link with the largest traversal count emanating from the current start point. The overall sum of the traversal counts along the path identified via this approach may not be the largest among all the paths in the network. We propose to also examine the "global" main path. A global main path is the path that has the largest overall traversal counts. In contrast to the local main path that highlights the progressing significance, the global main path emphasizes the overall importance in knowledge flow.

The problem of finding the path with the overall largest traversal count is similar to the shortest path problem in graph theory. Several algorithms, the Floyd–Warshall (Floyd 1962) algorithm for one, are readily available to solve the problem. However, note that the aim of the global main problem is a reverse of the shortest path problem because its goal is to find the path with the largest traversal counts rather than the smallest traversal counts.

The global main path adds to the analysis a new viewing angle for the significance of the main path. In practice, the local and global main paths may be identical or deviate only slightly. The consistency of the two indicates convergence of the target field. Liu et al. (in press) traced both the global and local main paths for the development of data envelopment analysis. In that study, the two paths are quite similar and deviate only slightly in the early and late stages of the development.

Backward Versus Forward

The original main path analysis searches forward from sources to sinks. One can certainly do the reverse; that is, search backward from the latest papers to the earliest papers. Searching forward is like finding offspring of important contributions whereas searching backward is similar to tracing the roots of current works. When searching forward, the main path algorithm has those papers that attract the most followers getting a place in the main path. Searching backward, in contrast, grants a higher weight to those papers that have taken ideas from the widest number of sources. Technically, in contrast to tracing based on the outward link, searching

DOI: 10.1002/asi

backward can be easily done by tracing the inward link in a citation network. Lucio-Arias and Leydesdorff (2008) had a similar idea and used the term *codification* to describe the process of searching backward.

Searching backward provides additional views to the analysis, but is meaningful only for a local main path search. For the global main path, the notion of search direction is not applicable. The local main path obtained by searching forward and backward can be very different, but again, the consistency of the two indicates convergence of the target field.

Multiple Versus Single

The single main path approach provides only the most significant path. For a discipline that has many subfields, one may want to also discover the important paths at the next level. Assuming the main path is a satellite-view picture of a network on the ground, discovering a secondary path is like flying lower to get more features of the network.

To find multiple local main paths, one basically relaxes the search constraint. For example, one can choose not only the link with the largest traversal count but also the link with the second-largest count or links with a traversal count falling within certain criteria. The looser the constraint, the more detail the network features surface. To obtain multiple global main paths, one chooses those paths with the top overall traversal counts. If more paths are selected, more detail is revealed. The degree of local relaxation and the number of global paths to explore are context-dependent.

Key-Route Search Versus Source–Sink Search

A serious potential problem that the main path approach suffers is that the link with the highest traversal count may not always be included in the main path. To overcome this problem, the suggested solution is to view the main path as an extension of the most significant link and begin a search from both ends of the key-route rather than from the sources. We call this the *key-route search*. It guarantees that this key route is included in the main path. The key-route search procedure is as follows.

1. Select the key-route; it is the link that has the highest traversal count.
2. Search forward from the end node of the key-route until a sink is hit.
3. Search backward from the start node of the key-route until a source is hit.

The search in Steps 2 and 3 can be either local or global. One also can select multiple key-routes and execute the procedure multiple times, each time selecting the link with the next-highest traversal count, to obtain multiple key-route main paths.

An Integrated Approach

As discussed earlier, we have proposed four ways to broaden the perspective to the main path analysis. The global method provides a path that has the most significant overall traversal count. The backward method traces the roots of the current activities. The multiple main paths method allows one to investigate more paths as needed. The key-route search method guarantees that significant top links in the citation network are included in the main paths.

The integrated approach we propose here combines these methods in one analysis. Rather than examine only a single path, the integrated approach opens up the kaleidoscope and provides opportunities to uncover more critical development paths in the target discipline. A typical main path analysis applying the integrated approach would begin by obtaining the standard single forward local main path, followed by examining the global paths and then the single backward local main path. For a large citation network, multiple paths can be added in the process to examine important works at the next level. Finally, one obtains the key-route search paths to make sure that the top significant links are included. Figure 2 displays the procedure of conducting the integrated approach to the main path analysis.

These paths may have certain degrees of similarity. A high degree of similarity indicates concentration of the target discipline and that the development of the target discipline focuses on the overlapping works. A low degree of similarity, on the other hand, is a sign of scattered development and no dominant research direction.

Development Trajectory of the Hirsch Index as an Example

We apply the integrated approach of main path analysis to explore the development trajectory of the Hirsch index. Hirsch (2005, p. 16569) proposed an index to quantify an individual's scientific research output using citation information. Hirsch index h is defined as "the number of papers with citation number $\geq h$." In other words, a researcher has index h if h of his or her papers published over a certain period of years in a certain scientific field have at least h citations each. The index is conceptually simple and has attracted researchers' attention. Many researchers have extended, tested, and debated Hirsch's original concept (Egghe, 2006; van Raan, 2006; Bornmann et al., 2008). Other studies applied the Hirsch index and its variants to evaluate an individual's research output (Cronin and Meho, 2006) and a journal's overall performance (Braun et al., 2006). To a certain extent, these subsequent works form a scientific field around the subject of the Hirsch index.

We collected the research data from two sources: the ISI *Web of Science* (*WOS*) database and a collection of references of review articles on the Hirsch index. A set of academic papers related to Hirsch index research and its application was gathered from the Science Citation Index Expanded, the Social Sciences Citation Index, and the Arts & Humanities Citation Index database of the *WOS*. The data time span ranges from 2005, when the Hirsch paper was published, to June 29, 2011. We begin by obtaining a set of documents by searching for the co-occurrence of the keywords "Hirsch

DOI: 10.1002/asi

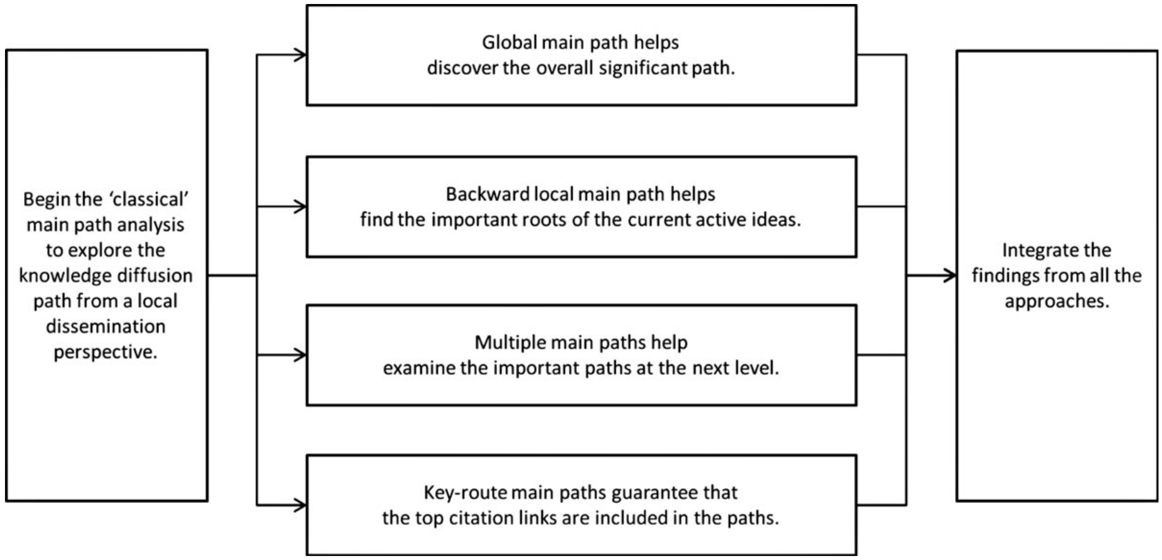

FIG. 2. The integrated approach for the main path analysis.

index" and "citation*" in the *WOS* Topic field. The wildcard notation asterisk at the end of the keyword "citation" dictates the search system to take both "citation" and "citations" into consideration. In addition, we obtain five more sets of documents by replacing the search term "Hirsch index" with "Hirsch indices," "h-index," "h-indices," "h type index," and "h type indices," respectively. These six sets of documents are then combined together by applying logical OR, which consolidates the duplicate documents. It is not surprising that this final set of documents does not include Hirsch's original work because it does not use the terms mentioned earlier; it is added manually to the final document set. In the end, we arrive at a total of 357 papers from this process. Although we believe that high percentages of relevant papers are incorporated into this dataset through the process, a second data source will make the dataset more complete.

The second data source is a collection of references from the following eight review articles: Alonso, Cabrerizo, Herrera-Viedma, and Herrera (2009), Bornmann and Daniel (2007, 2009), Egghe (2010), Norris and Oppenheim (2010); Panaretos and Malesios (2009), Thompson, Callen, and Nahata (2009), and Zhang, Thijs, and Glänzel (2011). This data source consists of a total of 786 articles. We begin the handling of this dataset by manually removing the articles already listed in the first dataset and consolidating the remaining articles on duplicates afterwards. For the 316 articles left, we manually screen out the articles that are not Hirsch-related, or not listed in the *WOS* database. This process provides a total of 64 articles. Finally, we merge the articles obtained from the two sources, which amount to a total of 421 articles.

The citation information for these papers was then collected to establish the links among these papers. In the process, citations to papers other than the target papers were disregarded. Thus, the number of citations taken into consideration for a paper is usually less than its total citation count shown in the *WOS* database. The difference is displayed in columns 2 and 3 in Table 1, which list the relevant data of the top-12 cited papers in our dataset. In the end, the resulting citation network consists of 2,941 links.

Figure 3 shows a partial citation network that is formed by the top-12 cited papers. In the figure, all citation links are marked with the calculated traversal counts (SPCs). The arrow direction goes from the cited article to the citing article. This figure and all the main paths shown thereafter are drawn with Pajek software (Batagelj & Mrvar, 1998). As can be seen from the figure, citation relationships among these highly cited papers are rather complicated, not to mention that the whole citation network is in a much larger scale. The main path analysis is a very efficient method to escape from the maze.

We apply the integrated approach as discussed earlier to obtain local main paths, global main paths, backward local main paths, multiple local main paths, and key-route main paths for the Hirsch index research. Each main path is presented with a figure. In those figures, arrows indicate the direction of knowledge flow, and line thickness reflects the traversal counts of the link. The thicker the line, the more significant is the link. We assign each paper on the main path with a code, which begins with the last name of the first author, followed by the initials of the subsequent authors and then the publication year. If there are duplicate codes, then lower case alphabets are appended at the end. The Appendix lists the codes of all the papers that are shown on the main paths. These papers also are flagged with an asterisk in the References.

Local Main Paths

Figure 4 presents the classical main path; that is, the single forward local main path. Leading by Hirsch's 2005 seminal work, there are 10 studies in total on this main path. In fact, there are more than 10 nodes if the end nodes (the sink nodes)

DOI: 10.1002/asi

TABLE 1. Top-12 cited papers for Hirsch index research.

Total citations[a]	Citations included[b]	Authors	Title	Year Published	Journal
1. 821	364	Hirsch, J.E.	An Index to Quantify an Individual's Scientific Research Output	2005	*PNAS*
2. 199	128	Egghe, L.	Theory and Practise of the g-Index	2006	*Scientometrics*
3. 153	105	van Raan, A.F.J.	Comparison of the Hirsch-Index With Standard Bibliometric Indicators and With Peer Judgment for 147 Chemistry Research Groups	2006	*Scientometrics*
4. 121	70	Hirsch, J.E.	Does the H Index Have Predictive Power?	2007	*PNAS*
5. 105	84	Jin, B.H., Liang, L.M., Rousseau, R., & Egghe, L.	The R-and AR-Indices: Complementing the H-Index	2007	*Chinese Science Bulletin*
6. 101	71	Cronin, B., & Meho, L.	Using the H-Index to Rank Influential Information Scientists	2006	*JASIST*
7. 96	78	Bornmann, L., & Daniel, H.D.	Does the H-Index for Ranking of Scientists Really Work?	2005	*Scientometrics*
8. 98	67	Batista, P.D., Campiteli, M.G., Kinouchi, O., & Martinez, A.S.	Is It Possible to Compare Researchers With Different Scientific Interests?	2006	*Scientometrics*
9. 98	68	Braun, T., Glänzel, W., & Schubert, A.	A Hirsch-Type Index for Journals	2006	*Scientometrics*
10. 92	70	Bornmann, L., & Daniel, H.D.	What Do We Know About the h Index?	2007	*JASIST*
11. 91	69	Egghe, L., & Rousseau, R.	An Informetric Model for the Hirsch-Index	2006	*Scientometrics*
12. 84	67	Glänzel, W.	On the h-Index—A Mathematical Approach to a New Measure of Publication Activity and Citation Impact	2006	*Scientometrics*

PNAS = Proceedings of the National Academy of Science, USA; *JASIST* = *Journal of the American Society for Information Science and Technology*.
[a]The number of citations as indicated in the Web of Science database as of September 6, 2011. [b]The number of citations taken into consideration for the main path calculation.

are taken into consideration. Here, we choose not to discuss the end nodes because they have not yet received any citation.

BornmannD2005 is a short communication in immediate response to Hirsch's original work. It has suggested that the h-index is a promising measure of the quality of a young scientist's work. Glänzel2006 analyzed the basic h-index properties on the basis of a probability distribution model. EggheR2006 broadened the theoretical base of the h-index by extending its definition to an information production process framework. BornmannD2007b is an early review article that discussed the advantages and disadvantages of the h-index and summarized the studies on the convergent validity of the h-index. Schreiber2007a is an empirical work that studied the h-index of 26 nonprominent physicists.

BornmannMD2008 categorized nine variants of the h-index into two types: the "quantity of the productive core" type and the "impact of the productive core" type; the study showed that peer assessments have a higher correlation to the latter than to the former. Schreiber2008b compared the h-index and its variants g-index, A-index, and R-index on the same 26 physicists that were studied in Schreiber2007a. AlonsoCHH2009 reviewed comprehensively the h-index related literature. At the end of the path, CabrerizoAHH2010 introduced the q^2-index, which is another h-index variant.

Among these 10 works, two are review papers and seven are theoretical-oriented works. These theoretical works established the theoretical base, proposed new variants, or categorized and compared the variants of the h-index. The path shows that the works that laid out the theoretical foundation, especially Glänzel2006 and EggheR2006, were the most recognized efforts in the early development of the h-index.

Global Main Path

The global main path complements the local main path in an overall-maximum perspective. This path begins with the same four leading papers as that in the local main path, but deviates afterwards. Rousseau2007 discussed the effect of missing publications on the result of h-index calculation. JinLRE2007 introduced the R-index and the AR-index; the R-index achieves the same goal as does the g-index, but from a different perspective, and the AR-index is the article-age-dependent R-index. Hirsch2007 turned from the output evaluation aspect of the h-index to the issue of its predictive power; the study showed that the h-index can better predict future scientific achievement than can other bibliometric indicators such as total citation count, citation per paper, and total paper count.

Close to the end of the global main path, four works by Scheriber succeed BornmannMD2008 and Schreiber2008b, who also appear in the local main path. Schreiber2009b and Schreiber2009a proposed the h_m and g_m indices, respectively, which take multiple coauthorships into account. Schreiber2010c suggested that the g-index combines the features of the h-index and the A-index in one number. Schreiber2010e compared 20 Hirsch-type indices. Figure 5 exhibits the global main path.

DOI: 10.1002/asi

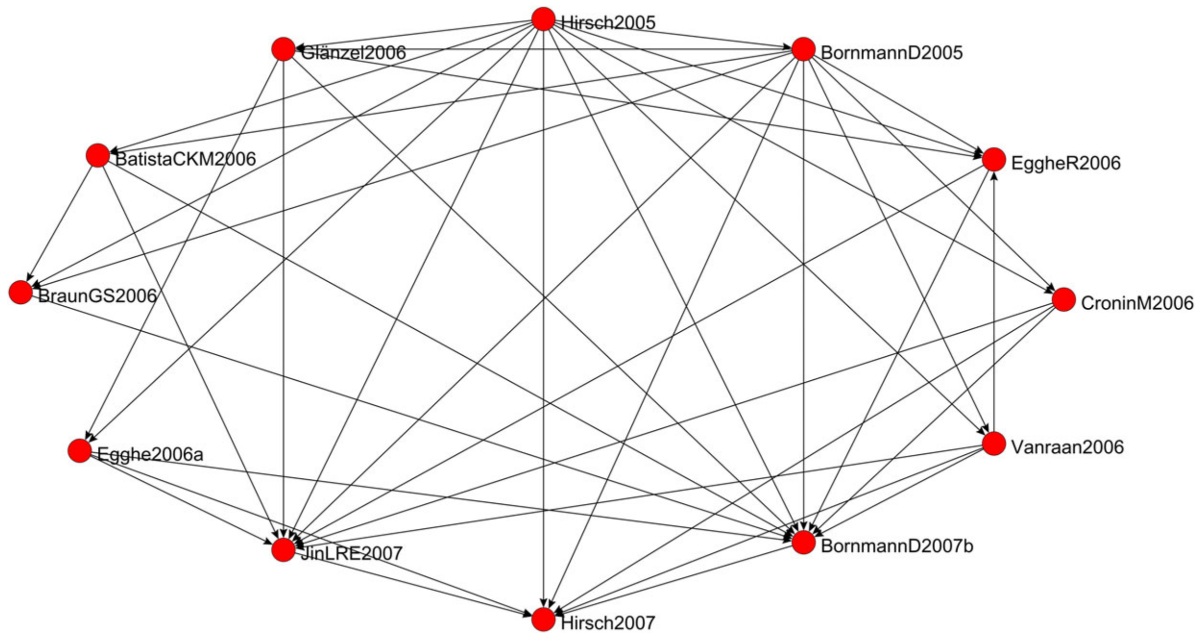

FIG. 3. Partial citation network formed for the top-12 cited papers. [Color figure can be viewed in the online version, which is available at wileyonlinelibrary.com.]

The global main path draws our attention to studies that proposed the variants of the h-index: Jin et al. (2007) and Schreiber (2009b, 2009a). This path, together with the local main path, hints that Hirsch's original idea was quickly echoed with empirical and theoretical works that were all in the spirit of testing and verifying the concept. It also is complemented with various similar approaches. After the concept secured its position in the academic community a few years later, Hirsch himself proposed a new use (predictive power) for the index. At the same time, extensions of the original idea and empirical works that applied the index continued to flourish.

Backward Local Main Path

Figure 6 presents the backward local main path. The major skeleton of the path is a mix of the forward local and the global main paths. Several new routes in the early days of Hirsch index development are uncovered. These routes, Hirsch2005 to Glänzel2006, Hirsch2005 to Vanraan2006, BornmannD2005 to Vanraan2006, and Vanraan2006 to EggheR2006, suggest that there are diverse intellectual roots for the subject other than Hirsch's original work.

This path proposes two new papers: Vanraan2006 and NorrisO2010b. Vanraan2006 is an early empirical study that compared the h-index with the standard bibliometric indicators and with a peer judgment for 147 chemistry groups. Two characteristics of the study are (a) dealing with research groups rather than individuals and (b) using a 3-year window rather than life span. NorrisO2010b is a very recent and extensive review paper on the development of the h-index.

Multiple Main Paths

If we want to examine a little more detail on the development of the Hirsch index research, we can conduct either the multiple local main paths or the multiple global main paths analysis, or both. Here, we choose to visualize more details by conducting the multiple local main paths. The condition "the largest" is changed to "within 20% of the largest" in the procedure described earlier. The number 20% is an arbitrary number that controls the level of detail that we want to visualize. Figure 7 exhibits the multiple local main paths at 20% tolerance.

Egghe2006a finally enters into the scene. Egghe2006a proposed the g-index, which is a much used h-index variant that better takes into account the citation scores of the top articles of an author. Egghe2006a's introduction in this main path demonstrates the power of the integrated approach. In fact, the traversal count of the route Glänzel2006 to EggheR2006 is 33,286, or only a little larger than that of the route Glänzel2006 to Egghe2006a, which is 32,780. The original main path algorithm forced an unfair negligence on Egghe2006a when extending from Glänzel2006. The integrated approach brings this important work back to our attention.

Key-Route Main Paths

Now that we have examined the forward local, global, backward local, and multiple local main paths, do they cover enough the significant routes and works in the development of the Hirsch index? Table 2 provides a hint to answering the question. It lists the top-20 routes with the highest traversal counts. None of the main paths discussed so far contains all

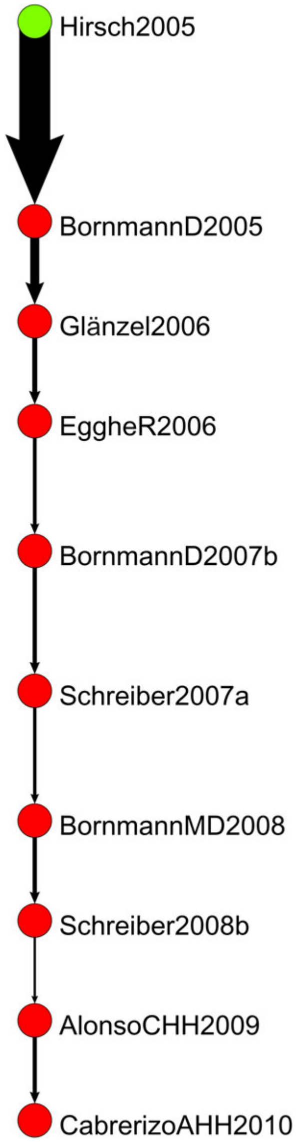

FIG. 4. Forward local main path. Thicker line indicates higher traversal count. The network was drawn with Pajek software. [Color figure can be viewed in the online version, which is available at wileyonlinelibrary.com.]

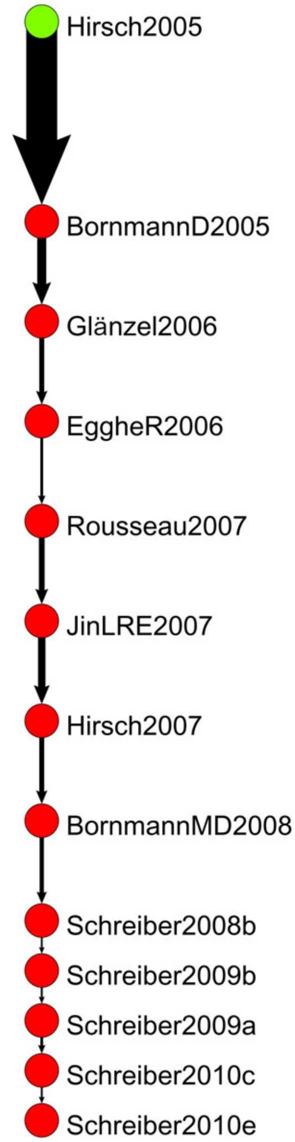

FIG. 5. Global main path. Thicker line indicates higher traversal count. The network was drawn with Pajek software. [Color figure can be viewed in the online version, which is available at wileyonlinelibrary.com.]

of the top-20 routes. To obtain paths that leave out none of the top-20 routes, we follow the key-route search procedure discussed in the previous section. Figure 8 presents the key-route main paths at 20 key-routes that apply the global search approach. Here, one visualizes virtually a combination of all the paths that we have discussed.

Several new routes surface in these main paths; along these routes, four studies stand out: CroninM2006, BatistaCKM2006, BraunGS2006, and Burrell2007b. CroninM2006 applied the h-index to rank 31 influential information scientists. The study found that there is a strong positive correlation between the h-index and citation counts and confirmed that the h-index is a tool with merits. BatistaCKM2006 proposed the h_I-index, which takes the coauthorship effect into account. This index could be used to compare scientific research output performance in different research fields. BraunGS2006 compared the h-index and the impact factor of journals and suggested that an h-type index would be a useful supplement to journal impact factors. Burrell2007b proposed a stochastic model for the h-index.

This key-route main path exhibits a divergence–convergence–divergence structure that properly summarizes the development of the Hirsch index research. The concept of Hirsch's (2005) original work was initially disseminated into five papers and then consolidated in BornmannMD2008 after diverse contributions from several researchers. The five most significant early hubs for emitting Hirsch's idea were BornmannD2005, Glänzel2006, Vanraan2006, CroninM2006, and BatistiaCKM2006. BornmannD2005 and CroninM2006 were among the first to empirically approve the idea. Glänzel2006 echoed Hirsch's idea using a probability distribution model. Vanraan2006 applied the Hirsch index to

DOI: 10.1002/asi

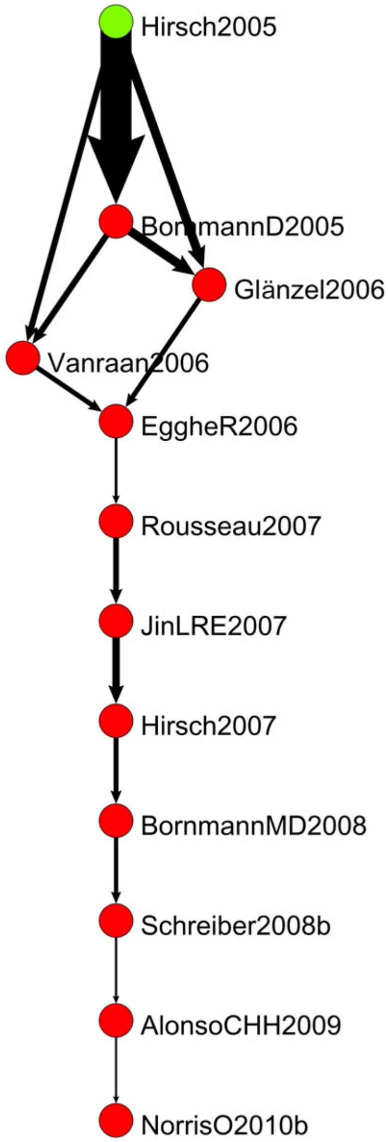

FIG. 6. Backward local main path. Thicker line indicates higher traversal count. The network was drawn with Pajek software. [Color figure can be viewed in the online version, which is available at wileyonlinelibrary.com.]

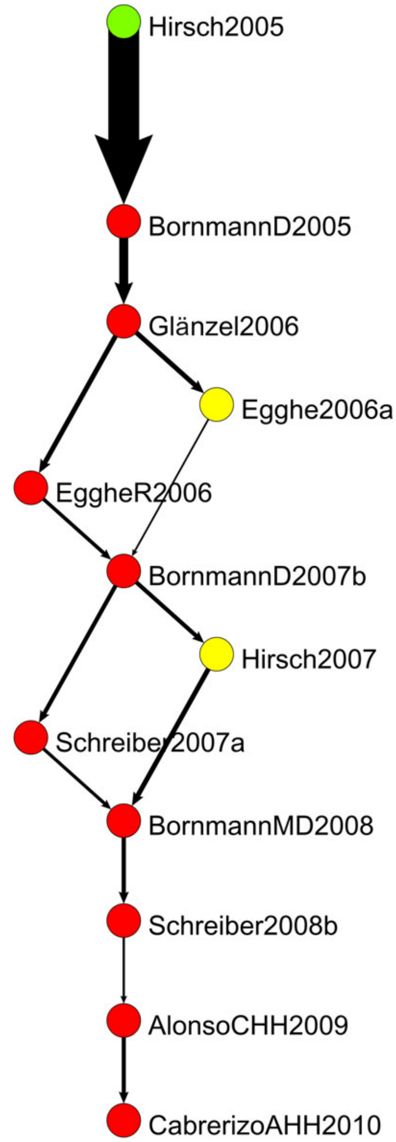

FIG. 7. Mulitple forward local main paths at 20% tolerance. Thicker line indicates higher traversal count. The network was drawn with Pajek software. [Color figure can be viewed in the online version, which is available at wileyonlinelibrary.com.]

a set of empirical data. BatistaCKM2006 proposed a variant of the h-index. Many more contributions followed, including Egghe2006, EggheR2006, Rousseau2007, JinLRE2007, Hirsch2007, and so on. All of these converged to the summarizing works of BornmannMD2008 and Schreiber2008b. The research then diverged into two major paths. The two paths, nevertheless, show no significant difference in their research direction because they basically focus on inventing new h-index variants and comparing the existing variants.

The key-route main path as shown is a very good tool to visualize the development structure of a scientific research field. It typically contains the basic elements of the local and the global main paths. One may then ask "Why not explore the key-route path in the first place?" It certainly can be done that way, but one loses the opportunity to observe other main paths and to clarify the priority of significance.

Zhang et al. (2011) studied a collection of h-index-related articles and found that the literature could be separated into four clusters: applications of the h-index outside the field of informetrics/scientometrics (Cluster 1), h-index-related problems and applications in the field of informetrics (Cluster 2), method-related applications of the h-index (Cluster 3), and methodology/theoretical studies of the h-index (Cluster 4). Most articles on the main paths belong to Cluster 4 of Zhang et al. This is not a surprise, as the methodologies and theories are the foundation of sustainable scientific research.

Discussion

The main path analysis provides a satellite view of a citation network. The view makes the dominant development

TABLE 2. Top-20 routes for Hirsch index research.

Traversal Counts (SPC)	Routes	In forward local	In global	In backward local
1. 208,839	Hirsch2005 ⇒ BornmannD2005	✘	✘	✘
2. 60,596	Hirsch2005 ⇒ Glänzel2006			✘
3. 60,596	BornmannD2005 ⇒ Glänzel2006	✘	✘	✘
4. 49,632	JinLRE2007 ⇒ Hirsch2007		✘	✘
5. 40,993	Hirsch2005 ⇒ Vanraan2006			✘
6. 40,993	BornmannD2005 ⇒ Vanraan2006			✘
7. 38,654	Rousseau2007 ⇒ JinLRE2007		✘	✘
8. 33,286	Glänzel2006 ⇒ EggheR2006	✘	✘	✘
9. 33,286	Vanraan2006 ⇒ EggheR2006			✘
10. 33,024	Hirsch2007 ⇒ BornmannMD2008		✘	✘
11. 32,780	Glänzel2006 ⇒ Egghe2006a			
12. 28,800	BornmannMD2008 ⇒ Schreiber2008b	✘	✘	✘
13. 28,750	BornmannD2007b ⇒ Schreiber2007a	✘		
14. 28,200	BornmannD2007b ⇒ Hirsch2007			
15. 27,856	AlonsoCHH2009 ⇒ CabrerizoAHH2010	✘		
16. 25,081	BornmannD2005 ⇒ CroninM2006			
17. 25,081	Hirsch2005 ⇒ CroninM2006			
18. 24,762	EggheR2006 ⇒ BornmannD2007b	✘		
19. 22,354	Hirsch2005 ⇒ BatistaCKM2006			
20. 22,354	BornmannD2005 ⇒ BatistaCKM2006			

SPC = search path counts.

paths of a field clearly visible; however, this clear view is achieved at the cost of sacrificing many other significant development paths. The integrated approach proposed herein helps uncover many of the "lost," yet significant, development paths from various perspectives. The key-route search is especially useful in this regard. Nevertheless, the method is still embedded with some limitations. Understanding the causes and consequences of these limitations will prevent us from overinterpreting the analysis results.

Yet Another Quantitative Method

The main path analysis is primarily a quantitative method, the same as other citation-based methods. The number of citations may be totally or partly relevant or may in fact be irrelevant to its impact. For example, a large amount of citations in the now virtually rejected "cold fusion" field in physical chemistry does not mean real advances. In addition, nonscientific factors may play a part when an author decides to cite an article (Bornmann & Daniel, 2008; Case & Higgins, 2000). Many other issues on the direct use of quantitative citation information in analysis also have been discussed in Garfield (1979), MacRoberts and MacRoberts (1989), and so on.

Data Source Affects the Results

Selecting Google Scholar, *WOS*, or *Scopus* as a data source will probably result in different main paths, although no such evidence has been reported in the literature. This conjecture is based on the fact that different databases contain different datasets. Some articles may be missing from any of the databases. If the missing article is of certain importance or a potential main path article, then the results will be largely affected. In this example of Hirsch index research, we consider only those articles that are included in the *WOS* database; therefore, one should keep this in mind when interpreting the analysis results.

Relevancy of Citation Is Not Considered

The main path method extracts the significant paths from a given citation network, and takes them as the development trajectory under the assumption that a citation link indicates diffusion of a piece of knowledge from a cited work to the citing document. The degree of relevancy between the citing and cited documents is not considered in the main path analysis. Citing for the reason of complete adoption of an idea and citing for the reason of a mere hint are treated equally; this is why the subsequent articles on the Hirsch index main paths do not always show direct relevancy.

Who Cited You Matters

Occasionally, some highly cited papers are not seen in any of the main paths of the research field. For example, two of the top-12 cited papers in this field, Batista et al. (2006) and Braun et al. (2006), ranked eighth and ninth, respectively, as shown in Table 1, are not seen in any of the main paths except the key-route main paths. They also could have been missed there if the multiple key-route search stopped at the top-18 routes rather than at the top-20 routes.

One common reason for this is that many citations are not from papers in the same research field. For example, in Table 1, Hirsch's (2005) original work has 821 citations, yet only less than half ($n = 364$) of these citations are from the papers in our dataset; that is, more than half of the citations are from works not directly related to Hirsch index research. This does not seem to be the case for Batista et al. (2006) and Braun

DOI: 10.1002/asi

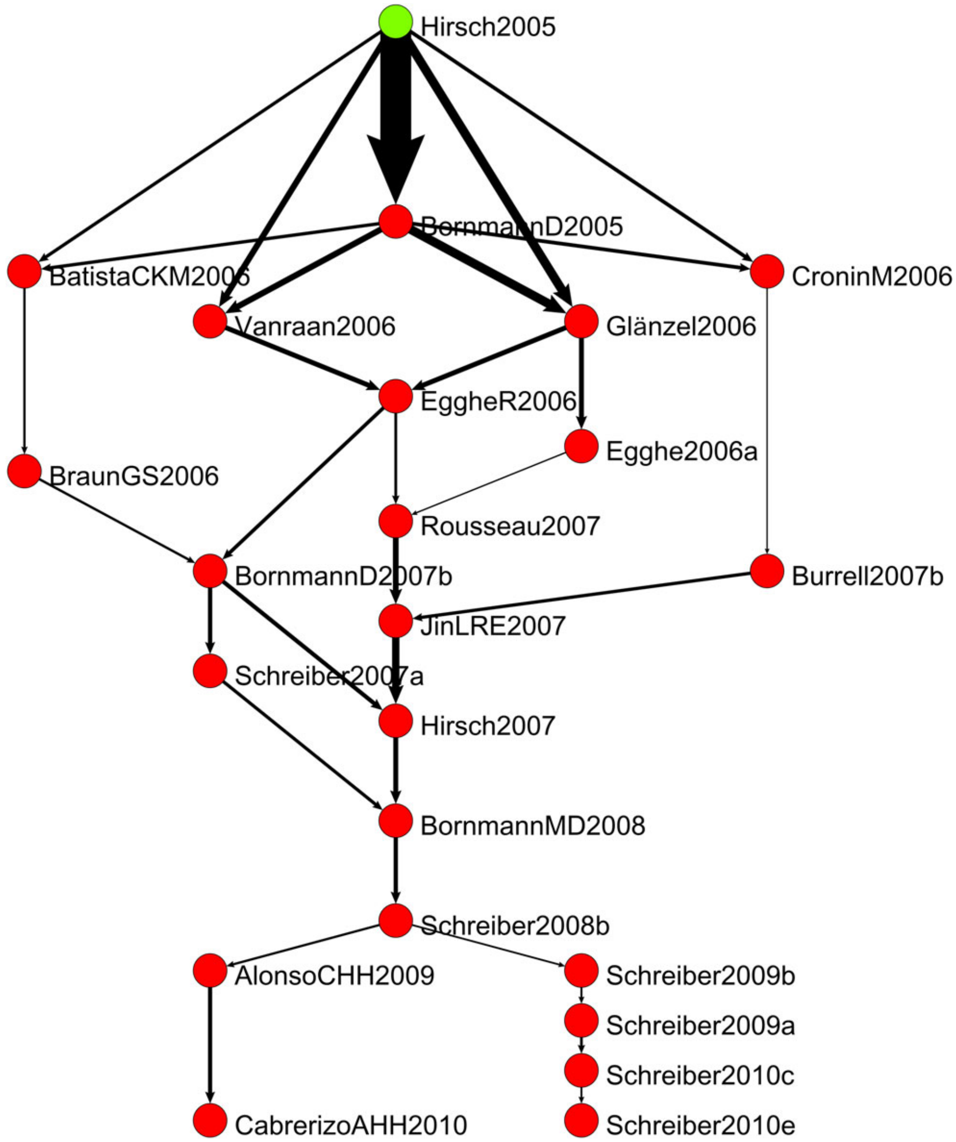

FIG. 8. Key-route main path. Thicker line indicates higher traversal count. The network was drawn with Pajek software. [Color figure can be viewed in the online version, which is available at wileyonlinelibrary.com.]

et al. (2006) because their in-field citations are relatively high. Another possible cause is embedded in the mechanism of the main paths method. The main paths are determined based on traversal counts of citation links, and traversal counts are measured by counting the times a citation link has been traversed if one exhausts the search from a set of start nodes to another set of end nodes. By this definition, traversal counts of a route leading to a paper depend not only on its citation count but also on the citation count of the papers that cite this paper. This implies that citations by an influential paper will boost a paper's historical position in the development trajectory.

Both BraunGS2006 (Braun et al., 2006) and EggheR2006 (Egghe & Rousseau, 2006) have around the same citation counts (68 vs. 69, respectively) in our dataset. The largest traversal count among all routes leading to BraunGS2006 is from BatistaCKM2006 (Batista et al., 2006) at 15,350; yet, the largest traversal count among all routes leading to EggheR2006 is 33,286 from Vanraan2006 (van Raan, 2006). This large difference mostly comes from the reality that a relatively powerful paper, Jin, Liang, Rousseau, and Egghe (2007), cited EggheR2006, but not BraunGS2006. This explains why BraunGS2006 barely secured a position in the multiple key-route main paths. In fact, Jin et al. cited a paper similar to BraunGS2006 by Braun, Glänzel, and Schubert (2005) that was published in *The Scientist*, a high-status magazine not listed in the *WOS* database.

Review Papers Are More Favorable

It is intuitively understandable that the main path method favors those articles that have received high citations. Less immediately obvious is the fact that the method also favors articles that have high citations to other articles. Indeed, an

DOI: 10.1002/asi

article with a larger number of references has more chances to increase the traversal counts of the links between itself and the articles citing it versus those with a smaller number of references. Applying the main path as a river analogy, a river with more branches pouring into it has more chances to become wider downstream than does a river with fewer branches.

This effect boosts the traversal counts of high-citing documents such as review papers. This, in addition to the fact that review papers are usually popular citing targets, makes review papers frequent entities on the main path. Many review papers are highly cited not because of their original ground-breaking results but rather due to their comprehensive summary of results of a field. Although review papers certainly have their contributions to knowledge dissemination, it is still arguable whether review papers deserve recognition on the main path. Additional qualitative judgment is certainly invited for review papers on the main path.

Other Voices Are Not Heard

It has been reported that the "preferential attachment" or the "rich get richer" phenomenon does exist in citation networks (Jeong, Néda, & Barabási, 2003). The Hirsch index has been a well-liked research subject since its inception. Inevitably, articles that supported the idea gathered a good amount of citation in the beginning and rode the phenomenon better than did those that were critical. Accordingly, the articles on main path are mostly the mainstream articles that positively support the h-index. Articles that are relatively conservative to the idea (Costas & Bordons, 2007; Vinkler, 2007, 2009) or question its adverse effect on scientists' behavior (Kotov, 2010; Williamson, 2009) are certainly not favored by the method.

Conclusion

We have proposed several new types of main paths in addition to the main path introduced by Hummon and Doreian (1989): global, backward, multiple, and key-route main paths. Together, these complementary main paths have greatly enhanced our capability to capture significant paths from the complicated citation relationships among a set of documents in a science or a technology domain. The key-route main path is the most relevant among the variants and especially is a good tool to visualize the development structure.

Research on the Hirsch index is used to demonstrate the merits of this integrated approach. The forward local and global main paths together, as expected, show that the field's development evolved around empirical works that confirmed its usefulness (BornmannD2005, Schreiber2007a), theoretical works that expanded its mathematical foundation (Glänzel2006, EggheR2006), and new variants that broadened its perspectives (JinLRE2007). The backward local main path further highlights the important role of several early empirical works, including Vanraan2006. The multiple local main paths remind us of the equally important knowledge diffusion role of Egghe2006a and EggheR2006 to latter works. The key-route main path hints at a divergence–convergence–divergence structure in the development of the Hirsch index. It should be noted nevertheless that all these results cannot escape from the method's limitations. We suggest that it is always good practice to verify the results of the main path analysis with domain experts.

This study adds two contributions to the bibliometric literature. First, it opens up new possibilities in searching for relationships among scientific information and greatly increases our capability in exploring the development of a target science and research domain. Second, it explores the development of the Hirsch index research in a way that has not been done before. The resulting main paths will make it much easier for researchers who are considering entering into the Hirsch index research to grasp the evolution structure of the field.

In the future, the main path analysis could be improved in at least one area. The current methodology works on binary citation networks; that is, all citations are treated equal. The relevancies between any citing–cited pair are always assumed to be the same, although obviously this is not true. Text mining provides an improvement opportunity for this oversimplification problem. Recent advances in text mining have spread over a wide spectrum of scientific disciplines and have had a significant impact in the scientific community. One could apply text-mining techniques to scale the relevancies between any citing–cited pair of documents. This would turn the traditional binary citation networks into weighted networks. Constructing algorithms to allow main path analysis to work in such a context would be a significant improvement to the methodology.

Acknowledgments

We thank two anonymous reviewers for their constructive comments, which have greatly improved the accuracy and readability of this article. This work is partially supported by Taiwan's National Science Council Grants NSC 98-2410-H-155-037-MY2 and 99-2410-H-011-003.

References

*Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., & Herrera, F. (2009). h-index: A review focused in its variants, computation and standardization for different scientific fields. Journal of Infometrics, 3(4), 273–289.

Asimov, I. (1963). The genetic code. New York: New American Library.

Batagelj, V. (2003). Efficient algorithms for citation network analysis. University of Ljubljana, Institute of Mathematics, Physics and Mechanics Department of Theoretical Computer Science, Preprint Series, 41, 897.

Batagelj, V., & Mrvar, A. (1998). Pajek—Program for large network analysis. Connections, 21(2), 47–57.

*Batista, P.D., Campiteli, M.G., Kinouchi, O., & Martinez, A.S. (2006). Is it possible to compare researchers with different scientific interests? Scientometrics, 68(1), 179–189.

*Bornmann, L., & Daniel, H.D. (2005). Does the h-index for ranking of scientists really work? Scientometrics, 65(3), 391–392.

*Bornmann, L., & Daniel, H.D. (2007). What do we know about the h index? Journal of the American Society for Information Science and Technology, 58(9), 1381–1385.

DOI: 10.1002/asi

Bornmann, L., & Daniel, H.D. (2008). What do citation counts measure? A review of studies on citing behavior. Journal of Documentation, 64(1), 45–80.

Bornmann, L., & Daniel, H.D. (2009). The state of h index research. Is the h index the ideal way to measure research performance? EMBO Reports, 10(1), 2–6.

*Bornmann, L., Mutz, R., & Daniel, H.D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. Journal of the American Society for Information Science and Technology, 59(5), 830–837.

Braun, T., Glänzel, W., & Schubert, A. (2005). A Hirsch-type index for journals. The Scientist, 19(22), 8.

*Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. Scientometrics, 69(1), 169–173.

*Burrell, Q.L. (2007). Hirsch's h-index: A stochastic model. Journal of Informetrics, 1(1), 16–25.

*Cabrerizo, F.J., Alonso, S., Herrera-Viedma, E., & Herrera, F. (2010). q2-index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core. Journal of Informetrics, 4(1), 23–28.

Carlero-Medina, C., & Noyons, E.C.M. (2008). Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field. Journal of Informetrics, 2(4), 272–279.

Carley, K.M., Hummon, N.P., & Harty, M. (1993). Scientific influence: An analysis of the main path structure in the Journal of Conflict Resolution. Knowledge: Creation, Diffusion, Utilization, 14(4), 417–447.

Case, D.O., & Higgins, G.M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. Journal of the American Society for Information Science, 51(7), 635–645.

Costas R., & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. Journal of Informetrics, 1(3), 193–203.

*Cronin, B., & Meho, L. (2006). Using the h-index to rank influential information scientists. Journal of the American Society for Information Science and Technology, 57(9), 1275–1278.

*Egghe, L. (2006). Theory and practise of the g-index. Scientometrics, 69(1), 131–152.

Egghe, L. (2010). The Hirsch index and related impact measures. Annual Review of Information Science and Technology, 44, 65–114.

*Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. Scientometrics, 69(1), 121–129.

Floyd, R.W. (1962). Algorithm 97: Shortest path. Communications of the ACM, 5(6), 345.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? Scientometrics, 1(4), 359–375.

Garfield, E., Sher, I.H., & Torpie, R.J. (1964). The use of citation data in writing the history of science. Philadelphia, PA: Institute for Scientific Information.

*Glänzel, W. (2006). On the h-index—A mathematical approach to a new measure of publication activity and citation impact. Scientometrics, 67(2), 315–321.

Harris, J.K., Luke, D.A., Zuckerman, R.B., & Shelton, S.C. (2009). Forty years of secondhand smoke research: The gap between discovery and delivery. American Journal of Preventive Medicine, 36(6), 538–548.

*Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences, USA, 102(46), 16569–16572.

*Hirsch, J.E. (2007). Does the h index have predictive power? Proceedings of the National Academy of Sciences, USA, 104(49), 19193–19198.

Hummon, N.P., & Carley, K. (1993). Social networks as normal science. Social Networks, 15(1), 71–106.

Hummon, N.P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. Social Networks, 11(1), 39–63.

Hummon, N.P., Doreian, P., & Freeman, L.C. (1990). Analyzing the structure of the centrality–productivity literature created between 1948 and 1979. Science Communication, 11(4), 459–480.

Jeong, H., Néda, Z., & Barabási, A.L. (2003). Measuring preferential attachment in evolving networks. Europhysics Letters, 61(4), 567–572.

*Jin, B.H., Liang, L.M., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. Chinese Science Bulletin, 52(6), 855–863.

Kotov, N.A. (2010). Fraud, the h-index, and Pasternak. ACS Nano, 4(2), 585–586.

Lucio-Arias, D., & Leydesdorff, L. (2008). Main-path analysis and path-dependent transitions in HistCite-based historiograms. Journal of the American Society for Information Science and Technology, 59(12), 1948–1962.

Liu, J.S., Lu, L.Y.Y., Lu, W.M., & Lin, B.J.Y. (in press). Data envelopment analysis 1978–2010: A citation-based literature survey. Omega–The International Journal of Management Science.

Lu, L.Y.Y., Lin, B.J.Y., Liu, J.S., & Yu, C.Y. (2011). Ethics in nanotechnology: What's being done? What's missing? Working paper.

MacRoberts, M.H., & MacRoberts, B.R. (1989). Problems of citation analysis: A critical review. Journal of the American Society for Information Science and Technology, 40(5), 342–349.

Mina, A., Ramlogan, R., Tampubolon, G., & Metcalfe, J.S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. Research Policy, 36(5), 789–806.

Moore, S., Haines, V., Hawe, P., & Shiell, A. (2006). Lost in translation: A genealogy of the "social capital" concept in public health. Journal of Epidemiology & Community Health, 60, 729–734.

Narin, F., Pinski, G., & Gee, H.H. (1976). Structure of the biomedical literature. Journal of the American Society for Information Science and Technology, 27(1), 25–45.

*Norris, M., & Oppenheim, C. (2010). The h-index: A broad review of a new bibliometric indicator. Journal of Documentation, 66(5), 681–705.

Panaretos, J., & Malesios, C. (2009). Assessing scientific research performance and impact with single indices. Scientometrics, 81(3), 635–670.

*Rousseau, R. (2007). The influence of missing publications on the Hirsch index. Journal of Informetrics, 1(1), 2–7.

*Schreiber, M. (2007). A case study of the Hirsch index for 26 non-prominent physicists. Annalen Der Physik, 16(9), 640–652.

*Schreiber, M. (2008). An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the A-index, and the R-index. Journal of the American Society for Information Science and Technology, 59(9), 1513–1522.

*Schreiber, M. (2009a). Fractionalized counting of publications for the g-index. Journal of the American Society for Information Science and Technology, 60(10), 2145–2150.

*Schreiber, M. (2009b). A case study of the modified Hirsch index h_m—Accounting for multiple coauthors. Journal of the American Society for Information Science and Technology, 60(6), 1274–1282.

*Schreiber, M. (2010a). Revisiting the g-index: The average number of citations in the g-core. Journal of the American Society for Information Science and Technology, 61(1), 169–174.

*Schreiber, M. (2010b). Twenty Hirsch index variants and other indicators giving more or less preference to highly cited papers. Annalen Der Physik, 19(8), 536–554.

Thompson, D.F., Callen, E.C., & Nahata, M.C. (2009). New indices in scholarship assessment. American Journal of Pharmaceutical Education, 73(6), Article 111.

*van Raan, A.F.J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics, 67(3), 491–502.

Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. Advances in Complex Systems, 10(1), 93–115.

Vinkler, P. (2007). Eminence of scientists in the light of the h-index and other scientometric indicators. Journal of Information Science, 33(4), 481–491.

Vinkler, P. (2009). The π-index: A new indicator for assessing scientific impact. Journal of Information Science, 35(5), 602–612.

Williamson, J.R. (2009). My h-index turns 40: My midlife crisis of impact. ACS Chemical Biology, 4(5). 311–313.

Zhang, L., Thijs, B., & Glänzel, W. (2011). The diffusion of H-related literature. Journal of Informetrics, 5(4), 583–593.

DOI: 10.1002/asi

Appendix

Papers on the Hirsch Index Main Paths

Article codes	Articles
AlonsoCHH2009	Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., & Herrera, F. (2009). h-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, 3(4), 273–289.
BatistaCKM2006	Batista, P.D., Campiteli, M.G., Kinouchi, O., & Martinez, A.S. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179–189.
BornmannD2005	Bornmann, L., & Daniel, H.D. (2005). Does the h-index for ranking of scientists really work? *Scientometrics*, 65(3), 391–392.
BornmannD2007b	Bornmann, L., & Daniel, H.D. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58(9), 1381–1385.
BornmannMD2008	Bornmann, L., Mutz, R., & Daniel, H.D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, 59(5), 830–837.
BraunGS2006	Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, 69(1), 169–173.
Burrell2007b	Burrell, Q.L. (2007). Hirsch's h-index: A stochastic model. *Journal of Informetrics*, 1(1), 16–25.
CabrerizoAHH2010	Cabrerizo, F.J., Alonso, S., Herrera-Viedma, E., & Herrera, F. (2010). q^2-index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core. *Journal of Informetrics*, 4(1), 23–28.
CroninM2006	Cronin, B., & Meho, L. (2006). Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57(9), 1275–1278.
Egghe2006a	Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152.
EggheR2006	Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, 69(1), 121–129.
Glänzel2006	Glänzel, W. (2006). On the h-index—A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67(2), 315–321.
Hirsch2005	Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, USA*, 102(46), 16569–16572.
Hirsch2007	Hirsch, J.E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences,USA*, 104(49), 19193–19198.
JinLRE2007	Jin, B.H., Liang, L.M., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6), 855–863.
NorrisO2010b	Norris, M., & Oppenheim, C. (2010). The h-index: A broad review of a new bibliometric indicator. *Journal of Documentation*, 66(5), 681–705.
Rousseau2007	Rousseau, R. (2007). The influence of missing publications on the Hirsch index. *Journal of Informetrics*, 1(1), 2–7.
Schreiber2007a	Schreiber, M. (2007). A case study of the Hirsch index for 26 non-prominent physicists. *Annalen Der Physik*, 16(9), 640–652.
Schreiber2008b	Schreiber, M. (2008). An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the A-index, and the R-index. *Journal of the American Society for Information Science and Technology*, 59(9), 1513–1522.
Schreiber2009a	Schreiber, M. (2009a). Fractionalized counting of publications for the g-index. *Journal of the American Society for Information Science and Technology*, 60(10), 2145–2150.
Schreiber2009b	Schreiber, M. (2009b). A case study of the modified Hirsch index h_m—Accounting for multiple coauthors. *Journal of the American Society for Information Science and Technology*, 60(6), 1274–1282.
Schreiber2010c	Schreiber, M. (2010a). Revisiting the g-index: The average number of citations in the g-core. *Journal of the American Society for Information Science and Technology*, 61(1), 169–174.
Schreiber2010e	Schreiber, M. (2010b). Twenty Hirsch index variants and other indicators giving more or less preference to highly cited papers. *Annalen Der Physik*, 522(8), 536–554.
Vanraan2006	van Raan, A.F.J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491–502.

DOI: 10.1002/asi