Chapter 3
Science Mapping Tools and Applications

Abstract We introduce the design and applications of a few influential science mapping tools, namely CiteSpace, VOSviewer, and CitNetExplorer, such that one can utilize these freely available tools to study a scientific domain of interest. CiteSpace provides a variety of metrics and indicators concerning trends and patterns in scientific literature. Many of these metrics are explained further with illustrative examples from applications of CiteSpace. CiteSpace also includes extensions that are particularly made to generate examples in the book. Three examples of systematic scientometric reviews using CiteSpace are included to illustrate relevant concepts and analytic functions.

Keeping Abreast of Scientific Frontiers

Keeping abreast of the development of a scientific domain is challenging for many reasons. It is time-consuming to search and gather relevant information adequately. There are numerous ways to describe the same topic, so it is challenging to come up with a comprehensive list of keywords for a scientific domain. Furthermore, we are most likely unfamiliar with the domain we plan to search for in the first place. How do we maximize the coverage of our search with our limited knowledge of the target?

One of the scenarios that we need to deal with has become less common because we are more likely to find at least some relevant publications now than we were just a few years ago. Imagine that we carefully formulated a search query, but our query didn't lead to any usual articles. In other words, it seems we have to revise our initial query so that we can at least find something. Once we found relevant articles, there are ways to expand the search and find more relevant publications.

An effective way to minimize the risk of missing anything important in a scientific domain is to see the basis that other researchers have built on. Most likely we can learn valuable information from others that we would probably never think of.

Scholarly Publication

Researchers today find themselves with increasingly more options to publish their work, ranging from the traditional peer-reviewed archival publications to self-directed newsflash-like tweets. Although researchers now have many more options than ever before, the essential process remains the same. Briefly speaking, researchers come across a research question that they can do something about. Of course, finding the right research question is currently also more of an art than science. As we have seen in Hilmeier's series of questions, a critical step in research is to understand the status of the research question in a broad context chronologically and domain ontologically.

Scientific discoveries are made rarely in the order that makes the most logical sense; otherwise, science would be reduced to a simple and straightforward logical reasoning process. Scientists and researchers need to publish their work in order to establish or maintain their intellectual impact in the scientific community. Novelty, originality, interesting, and creativity are among the few criteria that are held strongly in scholarly publication, especially through those venues guarded by various forms of peer reviews. Along the line of novelty, reviewers commonly criticize the lack of originally, the inadequacy of a claimed novelty, or an inadequately established connection to prior work by others in the field. The strongest argument for a novelty is almost certain not the one that simply claims no one has ever done it before or the equivalent cliché that we are the first who did so and so.

Sociologists have noticed that the easiest way to attract people's attention is to challenge the beliefs of your audience to the extent that they would be curious enough to listen to what you have to say. On the other hand, going too far in this direction may put off your audience altogether if it starts to sound ridiculous to their current mindset. In fact, sociologists suggest a few strategic moves that may boost the novelty of your next research question. For example, numerous theories were proposed to explain what happened at the KT boundary that led to the extinction of dinosaurs. The widely known theory was the one that focused on an asteroid impact on the earth and its atmosphere. A competing theory suggested that it was the lava from the insider of the earth rather than what from the sky. Similarly, after the September 11 terrorist attacks, researchers realized that people may still develop post-traumatic stress disorder (PTSD) symptoms even they were never near to a site of trauma, which was previously believed to be impossible. Prior to the September 11 terrorist attacks, PTDS research suggests that a first-hand physical experience of a trauma is essential for developing PTDS. However, researchers found that many people who did not have direct experiences through the trauma because they are thousands of miles away from New York.

The subjectivity of evidence means that the role of a piece of information as evidence is subject to the mindset or the mental model of individuals. The same piece of information can be used by different individuals to support different arguments. As the rest of the universe seems to be redshifted from us, does it mean we are at the center of the universe? As everyone can see the sun rises and sets, they

may still come up with different interpretations concerning whether we are at the center of the universe or the earth is orbiting around the sun.

In addition to the subjectivity of information, the uncertainty of the collective knowledge of the scientific community is another fundamental concept that we should bear in mind. We know from sociological perspectives of scientific change that scientists are driven by their desire to establish and consolidate their recognition and reputation in the scientific community or beyond. They seek to attract attention from their peers with novel ideas and astounding findings. The most active areas would be where we know little about the subject. Once we know more and more about a subject area, the level of uncertainty is likely to reduce. Thus, the level of uncertainty is an integral part of our knowledge of an area of research. Indeed, the knowledge of the uncertainty of knowledge is a type of meta-knowledge, which tells us the epistemological status of our knowledge. As we have seen in Shneider's evolutionary model of a scientific discipline, the meta-knowledge of a discipline may tell us which stage of the evolution the discipline is going through (Shneider 2009). Is it still at the first stage when researchers in the specialty are trying to conceptualize a new line of research? Is it at the stage when researchers are concentrated on building the right tools to augment their studies as Galileo was building his telescope?

Citation-Based Analysis

There are many types of scientific publications. We will primarily focus on two of them that are most likely to reveal relevant scientific knowledge of a domain: articles that report original research and reviews of a research topic.

Each formally published scientific article typically consists of the following components:

- A title and sometimes a subtitle.
- A list of authors and their affiliations.
- An abstract, structured or unstructured.
- A list of keywords assigned to the article by the authors.
- A list of keywords assigned to the article by indexing services such as the Web of Science.
- The main body of the article, including text, figures, tables, equations, and other materials.
- An acknowledgement to reviewers, researchers, or research funding or sponsorship.
- A list of references cited in the article.

Terms such as noun phrases appeared in the title of an article can be used to compute how often two terms appear within an article or even at the sentence level. Such terms are called co-occurring terms. For example, the four terms highlighted

in yellow in the title are co-occurring terms. Similarly, connections between terms in the abstract can be established in terms of their co-occurrences too.

A citation is an instance in which an article explicitly refers to a previously published article. Eugene Garfield conceived the idea of citation indexing, which taps into association of ideas found in scientific publications (Garfield 1955). The referred article is called a reference or a cited reference, for example, the citations to (Price 1965; van Raan 2000; Abt 1998) in the example shown in Fig. 3.1. In a scientific publication, especially in an original research article, references are cited for specific reasons in connection to an argument of the article. The fact that these references are cited by the same article means that they are co-cited references. In other words, they are cited together. A co-citation relationship between two references implies that, from the point of view of the author of the citing article, the two references are related to one another through the content of the citing article. For instance, one can infer from the text that the co-citation relation between (Price 1965) and (van Raan 2000) is probably because both of them are relevant to properties of transient articles. There may be more instances in which these two references are cited together in the same article later on. Multiple co-citation instances of the same pair of references may strengthen their co-citation relation in terms of the quantity. On the other hand, co-citations in different contexts may increase the diversity of the nature of the co-citation relation.

Traditionally, co-citation relations are established based on the references listed at the end of an article rather than based on an inspection of co-citation instances in the body of the article. In other words, we know that (Price 1965; van Raan 2000) are co-cited by the article because they are both included in the reference list of the
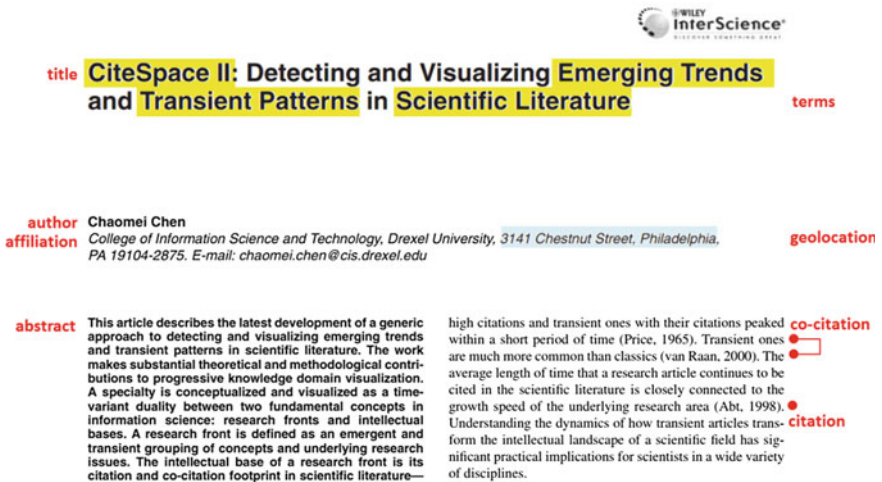

Fig. 3.1 The meta-data of a research article—a 2006 JASIST article on CiteSpace II (Chen 2006). The article is the 2nd of the 10 Google Scholar classic papers in Library and Information Science published in 2006

article rather than we found the sentences that mentioned them on the same page. The simplistic traditional approach is largely due to the accessibility of full text of scientific publications in the mid 20th century when the now famous Science Citation Index (SCI) was conceived by Eugene Garfield. Initial volumes of SCI were themselves created on punch cards and printed on papers. It is not until recent the access to full text articles is gradually taken for granted. As it becomes easier to access the full text of scientific articles, one would wonder what we would miss by deriving co-citation relations from the end-of-article reference list as opposed to pinpoint co-citation instances directly in full text.

The 2006 JASIST paper on CiteSpace is 19-page long, including 18.5 pages of text. The first citation is on the first page to (Price 1965) and the last citation is on page 18 to (Smalheiser and Swanson 1998). Intuitively, the co-citation connection between (Price 1965) and (van Raan 2000), both on the first page, is much more meaningful than the co-citation connection between (Price 1965) and (Smalheiser and Swanson 1998), which spans over 18 pages of text. It seems one should take into account this type of distance between the locations of two references whenever possible. In contrast, due to the limitation of data, co-citation relations in a traditional co-citation analysis cannot further differentiate their strengths within an article. We investigated the effect of the proximity of co-citation locations in the full text articles published in six bioinformatics journals and found that co-citations at the sentence level provide a good approximation to the overall co-citation patterns identified at the article level. It is therefore our recommendation that whenever possible the proximity of co-citation locations should be taken into account, preferably at the citation context level, which is commonly defined as a sentence that contains a citation instance and one or two neighboring sentences before and after the citation sentence.

The Metaphor of a Knowledge Space

An intuitive metaphor of the scientific knowledge is a knowledge space or the universe of entities and relations and various aggregations at higher levels of knowledge representation such as facts, rules, claims, hypotheses, speculations, and other types of elements represented. Stars and quasars in the universe of knowledge would represent concepts and their connections. Each published article would introduce some changes into the existing universe of knowledge. For example, an article may introduce new connections between existing concepts. A more innovative article may introduce a set of new concepts and their relationships all at once. The brightness of a star can indicate how active a concept is. A concept is more active if the concept is being involved in more and more recently published articles. In contrast, if a concept has not been mentioned for a long time, its brightness would become dimmer. Interconnections between concepts can be introduced by an article and subsequently reinforced by additional articles later on. Connections that have not been actively discussed over a long period time may weaken their

strengths. Newly published articles may alter the structure of the underlying knowledge space by adding new concepts and new interrelations.

The universe metaphor is not the only one that is intuitive. For example, an alternative metaphor is a neural network that models how human brains function when we learn from a typically large amount of data. The astounding performance of AlphaGo is one of the many impressive applications of artificial intelligence techniques, especially the so-called deep learning techniques. Through deep learning, multiple layers of interconnected neurons are adaptable to recognize patterns at various levels of granularity. For example, tasks that are used to be very challenging for computers but effortless for human beings such as recognizing a human face or a handwriting now can be reliably done through deep learning techniques. We will return to this exciting topic later in our book.

The universe metaphor has the advantage of visualization-congruence, which means it comes natural to derive a visualization design that would fit nicely with what users may expect from their understanding of the universe of astronomical objects such as stars, galaxies, the Milky Way, and the Great Wall. We would expect that many types of changes in the universe of knowledge are as visible and observable as we have seen in the universe of astronomical objects.

We use an interactive visual analytic tool CiteSpace to demonstrate how to generate a systematic review of a scientific field. A traditional systematic review of a field is typically written by someone who has developed a substantial under-standing of the field. A typical review article contains over 100 cited references. A systematic review is valuable in the course of the development of a field. Derek Price, a pioneer of the scientometric field, once estimated that a fast-growing field probably needs to have a systematic review paper after every 50 original research articles. Systematic reviews thus serve the role of summarizing what has been achieved by the original research articles since the last review article. However, a field may be too young to have a readily available systematic review. Existing reviews may not give enough attention to the topics that we are particularly interested in. In other words, it is quite possible that our best bet would be simply to review the literature all by ourselves.
Doing the review by ourselves has several distinct advantages:

- We choose the depth and breadth of the topics to cover.
- We choose when it is the time to do it as we need it.
- We develop a deeper understanding of the topics and their connections along the way.

The major challenge is the lack of the knowledge of the domain as a whole or that of a few specific areas of the domain. On the other hand, this would be true to any potential researchers who are planning to review the literature of a scientific domain. Given the scale and the volume of today's scientific literature, it is unlikely for an individual to master the depth and breadth of a subject domain. Many new research students face the challenge when they search for potential dissertation topics. By any standard, it is a time-consuming task to sift through hundreds of

hand-picked articles to identify some potential research questions. More importantly, identifying a research question cannot be done meaningfully without a good understanding of what characterizes it in a broader context. In other words, the amount of the effort required to articulate a research question properly is probably about the same as, if not more than, the amount of the effort required to develop a good understanding of a field. As we have seen in Heilmeier's Catechism, researchers need to figure out not only the status of their research problems in a potentially boundless context but also how to articulate the status most effectively to anyone who might concern.

Let's assume that the analyst does not have any special training in the target subject domain, which is most likely the case for many who need to find out more about the domain in the first place. The first thing our analyst needs to find out about the universe of knowledge is its structure or how the various galaxies are organized in the space, how they are related to one another, how long they have been there, what changes are taking place, and what one may expect to see in the future. Within a specific galaxy, our analyst would be interested in stars that stand out in one way or another. What is the brightest star? Which one has the greatest mass? Which one is the most unstable one? Which one is on its way to collapse? Which one is about to collide and merge with another one?

A visualization tool to our analyst is like a telescope to an astronomer or a GPS to a driver. The resolution of a visualization tool is determined by the resolution of the underlying data. A high-resolution GPS would be more useful for us to navigate through a dense and complex road layout than a low-resolution GPS. A high-resolution telescope would be more powerful for us to see finer structures of astronomical objects than a less powerful telescope. The resolution of a visualization of co-cited references in scientific publications can be measured in terms of the number of pages or the distance between the locations of co-cited references. Take the references cited in the 2006 JASIST article on CiteSpace (Chen 2006). In a traditional co-citation analysis, the resolution is the same 18-pages to all the co-cited references. The 18-page resolution is the maximum possible distance between two references cited in the article. If the full text of the article is accessible, then the resolution can be further improved by replacing the 18-page distance with the actual distance between the locations of two citations. If two references are cited multiple times in the same article, then the co-citation proximity can be defined through several options. For example, we can use the minimum distance between the locations of two citations. Alternatively, we can use the median distance to represent the strength of the co-citation link.

CiteSpace: Visualizing and Analyzing a Knowledge Domain

CiteSpace is an interactive visual analytic tool written in Java (Chen 2004, 2006; Chen et al. 2010). It is freely available. The motivation behind the development of CiteSpace is to enable researchers to conduct a systematic review of a scientific

field with little relevant domain knowledge or no prior knowledge of the domain at all. It is suitable for a new research student to search for potential dissertation topics, for an experienced researcher to keep abreast of the development of an established field of study (Chen 2017), or for a scientist to explore emergent trends in one or more research areas (Chen et al. 2012, 2014a, b).

CiteSpace is not intended to replace the role of conventional systematic reviews. Rather, CiteSpace aims to provide a computational approach that can be easily applied by the vast majority of researchers to meet their own needs. The procedure is repeatable at practically no cost to the analyst so that the analyst can generate a new review whenever necessary.

It is not our intention to use tools such as CiteSpace to eliminate the role of domain expertise in interpreting the analytic results of a CiteSpace application. On the contrary, we precious the value of domain expertise and we believe any domain expertise is hard to come by. We want to provide a tool for areas where domain expertise is not readily accessible or not available in a timely manner.

CiteSpace is probably the first computer application that is specifically designed to support the visual analysis of scientific literature. The development of CiteSpace has been particularly inspired by a number of pioneering software systems that have been made freely available, notably Pajek for analyzing large networks (Batagelj and Mrvar 1998), information visualization toolkits such as prefuse (Heer 2007), software programs generously shared by Loet Leydesdorff. Many wonderful and relatively new systems are made freely available, including VOSViewer (Van Eck and Waltman 2010), CitNetExplorer (Van Eck and Waltman 2014), and Gephi.

CiteSpace is unique in several ways in comparison with other systems that also take science citation data as the input. First of all, CiteSpace is designed to support the analyst to obtain a good understanding of the development of a scientific domain, or a knowledge domain. The unit of analysis is a subject domain, which means all the landmark publications and articles that have played a critical role in the holistic view of the knowledge domain as a complex adaptive system. With the support of CiteSpace, our analyst should be able to develop a good sense of the fundamental issues and major methods associated with the research domain. Second, the focus on a domain of knowledge is reinforced by various visual encoding that characterizes patterns and features with reference to underlying theories of the development of a scientific domain. For example, a cluster of co-cited references provides a representation of the intellectual base of a research specialty. The nature of inter-cluster relationships is underlined by cited references with strong betweenness centrality scores. The boundary-spanning or brokerage implications of such references are supported by theories such as the Structural Hole Theory (Burt 1992) and as a focal point in a paradigm shift from a Kuhnian point of view. CiteSpace is designed in such a way that the search for critical information of the development of a scientific field is turned to the visual search of patterns and features that standout in an overview of the domain.

Figure 3.2 shows an overview of terrorism research (1996–2003). We will explain the details shortly, but for now let us check what features would draw our attention, assuming we know nothing about this research domain. What we can see
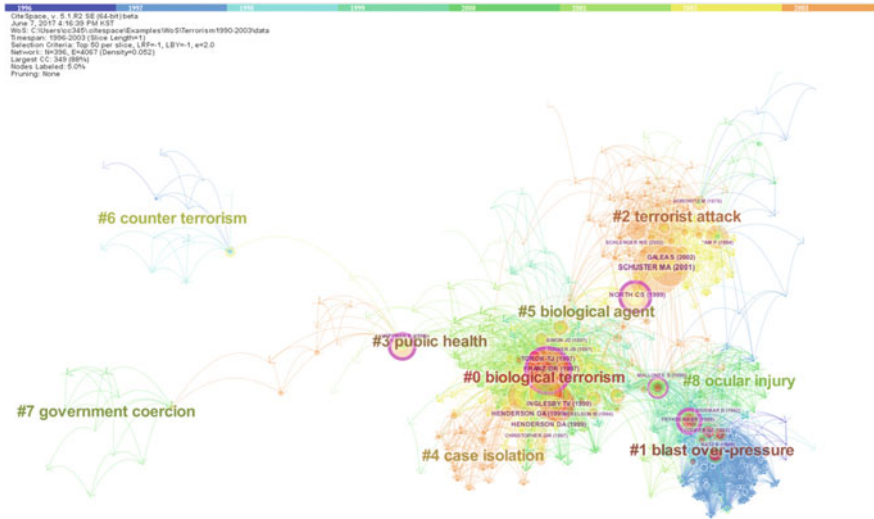
Fig. 3.2 An overview of terrorism research (1996–2003)

effortlessly is a property called preattentativeness, which means they will get our attention within the first 200 ms. It is the time required to redirect our attention. This property is also called standout or popout. In the visualized network of cited references, our attention is likely to be directed towards the few big circles in purple. Then we could also notice some discs in red. Next we would probably explore the surrounding areas of these focal points and perhaps read the text labels in different colors and font sizes. At the highest level of granularity, we could see 3–5 concentrations (clusters) in different colors. The legend above the visualization indicates the navy blue color is associated with 1996 on the left and the orange color is associated with 2003 on the right. The area in the orange color, with a label "#2 terrorist attack," must essentially correspond to the year 2003. In contrast, the area in navy blue, labeled as "#1 blast over-pressure," must be connected to the year 1996. Further inspections would reveal that purple circles seem to be positioned between different areas such as NORTH CS (1999) between #2 terrorist attack and #0 biological terrorism on the upper right region of the display, HOFFMAN B (1998) between the mainland and the peninsula stretching into the west (#7 government coercion), and MALLONEE S (1995) linking the #0 biological terrorism and #8 ocular injury in the lower right region.

As you can see, we are able to identify a small number of elements in the research domain without referencing to any specific domain knowledge. Evidently these elements must play some special roles in further understanding the research domain. What makes these elements standout in the overview of the research landscape is due to the co-citation patterns found in scientific articles written by researchers in the scientific community. In other words, these patterns reflect something profound shared by individual researchers because the emergence of a

pattern requires the consensus, endorsements, and reinforcements of many researchers.

Now we can identify some of the major characteristics of a knowledge domain. It may consist of multiple inter-connected topic areas. The development of each of these topic areas is likely to last for a period of time, which may have a variable duration. The key to the inter-relationship between two topic areas is largely hold by the brokerage node or nodes that connect the two topic areas. Our analyst can reach this level of understanding of a knowledge domain through a visual inspection that won't take much longer than a few minutes, although our experience shows that most of the users would eagerly dive into the juggles of specific references before pondering the overall structure of the forest and the implications of the structure on subsequent exploration of the domain's landscape.

Ben Shneiderman, a pioneer in visual information seeking and human-computer interaction in general, proposed a simple mantra that summarizes the strategy of visual information retrieval for designers as well as for end users (Shneiderman 1996). Shneiderman's mantra states "overview first, zoom and filter, then details on demand." The first step—overview first—is to form a hierarchical organization of a domain and its topic areas. The entire research community behind the domain as a whole can be considered as a single specialty. Each topic area corresponds to a subset of the overarching specialty. Such subsets can be considered as distinct specialties in their own rights. The best way to understand the nature of a specialty is not only to see what topic it focuses on but also how it distinguishes itself from specialties associated with its neighboring topic areas. In the terrorism research (1996–2003) example, understanding that bioterrorism is a major concern in the research domain is one thing, but a deeper understanding of practical implications of bioterrorism on healthcare and the preparedness of emergency responders is a significant step towards understanding what a research domain is really about. Achieving an understanding at this strategic level brings numerous advantages to our analyst in subsequent exploration of the knowledge domain. Once we have established an organizing framework based on the writings of many active researchers in the domain, one can easily categorize newly published research and recognize in what sense the new research is novel. Answering Heilmeier's questions is no longer as challenging as they seemed to be before our inspection of the overview of the domain.

Visual Exploration of Scientific Literature

The general procedure of visually exploring the scientific literature of a knowledge domain consists of several basic steps:

- collecting data
- configuring representation models
- generating interactive visualizations of the domain.

Data Collection

The goal of the data collection step is simple: to collect data that can adequately and accurately represent the domain in question. In practice, this is easier said than done. First, we are probably not familiar with the domain of our choice. We may be very interested in the domain, but we are probably not aware of various terminologies or jargons that have been used in scientific writings to describe topics relevant to the domain. Furthermore, concepts, theories, and practices may have evolved over time. The best case scenario is when we are familiar with the vocabulary of the domain, have an easy access to a domain ontology or thesaurus, and have a domain expert on the team. In the toughest scenario we would have none of them. In general, it is likely that we are somewhere in between. A common strategy is to snowball the query-and-refinement process so that as we learn more and more about the domain, we are better able to characterize what the representative data would look like.

More sophisticated search strategies are possible to improve the quality and efficiency of data collection further. For example, if we are familiar with a theory of the development of scientific knowledge, then we may derive a complex set of queries such that we can cover various aspects of the target domain systematically. In a recent example, we found it effective to organize our queries with reference to a theory of the evolution of a scientific discipline. The theory was proposed by Alexander Shneider. It is simple and intuitive. According to the theory, the evolution of a scientific discipline goes through four distinct stages in sequence: conceptualization, tool construction, tool application, and knowledge codification. The conceptualization is the first stage of the evolution. New ideas are conceived, although a lot of details remain unknown. The tool construction stage focuses on developing instruments that would be necessary to investigate the research questions conceived at the conceptualization stage. The tool application stage is when the application of enabling and augmentative techniques to the research questions result in new discoveries and new knowledge. When we formulate a complex query for relevant articles, we can include sub-queries that would cover specific aspects of an evolving scientific domain. For example, we can use one query to specify the basic concepts of the domain, use another query to specify the types of tools that are particularly relevant to the domain, and yet another query to specify applications of the research method.

Configuration of Representation Models

A key concept in CiteSpace is the time slicing technique. The idea is similar to the concept of a sliding window. A long period of time can be time sliced into a series of adjacent time slices. A snapshot of the domain knowledge can be represented by scientific articles published within the corresponding time slice. A time slice can be

a one-year window or a multi-year window. The primary effect of time slicing is to enhance the impact of research in a particular year. Adjacent time slices can overlap with each other. One of the effects of overlapping time slices is to smooth the variations over time. Besides, it makes more sense to consider that articles published in December and in January next year should belong to the same group as well as articles published in June and July. The effect of allowing an overlapping sliding window is a smoother transition of various patterns. The traditional network analysis without time slicing is a special case when the width of the window becomes the entire time interval. If we allow the duration of overlapping years to vary from 0 to the entire time interval, then the traditional method is a special case when the overlapping years become the entire duration of the time interval. In the following examples, we use non-overlapping time slices for its simplicity.

The clarity of a network is typically affected by several factors. An excessive number of links in a network would make it harder to differentiate salient patterns from common linkages. There are many strategies for reducing the number of links. Some of them make clear-cut decisions, whereas others follow sophisticated criteria that take into account local structural properties or even global structural properties.

Link Selection

Removing weak ties from a network is a commonly adopted strategy. Weak ties are often associated with a higher level of uncertainty, including underrepresented connections. There are many ways to select the weak ties to remove and they tend to impact the remaining network differently. The simplest way is to rank all the links in a network by their strengths and remove links from the bottom of the list, for example, by removing links with the strength below a cut-off threshold or by removing the 20% of the links with the lowest strengths. The downside of this approach is the risk of removing nodes that do not have strong links to survive. Although one may argue that we do not lose much anyway considering those nodes do not have strong ties with the rest of the network, weak ties may bring us valuable and unanticipated information. According to a famous study entitled the strengths of weak ties in social networks, the value of weak ties lies in their potential role in informing us something that may be unexpected. From an information scientist's point of view, any information that surprises us is a learning opportunity because it shows that our current belief, or our mental model, is inadequate, inconsistent, or even totally invalid. Weak ties in a social network imply a connection between people from different social circles. Information from different social circles is more likely to bring us something new as opposed to information from the same circle of friends.

According to sociologist Burt (1992), the potential value of the information flow is not because the ties are weak; rather, it is because weak ties are more likely in the position to connect different groups of individuals. The more broadly we are exposed to different ideas, diverse perspectives, and alternative interpretations, the

more likely we are able to come up with creative solutions and better handle a complex situation. According to Burt, our positions in a social network are not equal because the chance of seeing a diverse range of information flowing by is different. The difference, according to Burt can have profound consequences because one can translate such potential to a competitive edge.

We have learned at least two things from the above discussion: (1) we should avoid removing weak ties simply because they are weak, and (2) some nodes are worth our attention more than others because they may indicate where the competitive edges are or where the creativity is. Let us see if we can meet the two criteria simultaneously. Instead of dealing with all the links in a single list, take all the nodes in the network and consider links that connect each node to the rest of the network. This arrangement makes it possible to retain all the nodes while removing relative weak ties from each node. Along this line of reasoning, we can come up with additional methods to reduce the number of links but preserve global properties of the network. In CiteSpace, the user can tie the number of links proportional to the number of nodes in the network. We know that the least number of links to connect N nodes is $N - 1$ and the maximum number of links in a fully connected undirected network is $N*(N - 1)/2$. Turning a network to a minimum spanning tree will give us a network of $N - 1$ links. However, we have shown in our previous research that using a minimum spanning tree to approximate the original network has several drawbacks despite its advantages such as computationally simple and efficient. Given a network, there may be multiple minimum spanning trees. Arbitrarily picking one of them does not justify the validity of the resultant representation. A more convincing solution is to retain all the minimum spanning trees if we cannot justify selecting one of them only. This is indeed what Pathfinder network scaling can offer.

Pathfinder network scaling is a link reduction technique that can impose a triangle inequality condition across the entire network (Schvaneveldt 1990). Pathfinder network scaling is able to retain the most salient paths in an associative network. A network that satisfies the triangle inequality condition throughout the network is called a Pathfinder network. Comparing with link reduction techniques such as threshold-based methods, Pathfinder network scaling is theoretically sound. Although initial implementations of Pathfinder network scaling are computationally expensive, fast-algorithms have been developed, especially by a group of scientometrics at the University of Granada. The Pathfinder network is the set union of all the minimum spanning trees of the original network. Alternative paths connecting the same pair of source and target nodes are allowed simply because we do not have reasons to discriminate them, just like our travelers can choose a cheaper multi-city flight as well as a faster but more expensive non-stop flight between two cities.

The length of a citation link from the source article published in year Ys to a target article published in Yt provides information that could be useful for understanding the long-term impact of the target article. If we can afford to ignore citations to target articles published long time ago, then we can remove such citations from the network modeling steps. This parameter in CiteSpace is called

Look Back Years (LBY). It is common to cut off by LBY at 5–8 years. Consistent with the universe metaphor, we can choose to focus on our connections within a radius of our choice.

Node Selection

The construction of a network may also impose restrictions on what kinds of scientific publications would qualify to participate in the modeling process. In other words, we need to decide what kinds of publications should contribute a representation of the knowledge of a scientific domain. Why do we assume that we should select a subset of publications to portray the knowledge structure of a scientific domain rather than including all of them in the process?

No matter what data sources we use it is very difficult, if not impossible, to obtain a collection that we can truly claim to have a coverage of 100%. Both Google Scholar and Elsevier have access to tens of millions of publications. The Web of Science and Scopus are representative but not comprehensive. Pragmatically, increasing the current coverage by 10% may cost extra 90% of efforts and resources. More importantly, what can we learn from the extra 10% coverage that we cannot possibly learn from what is currently covered? Besides, if we are going to apply the same methodology to an extended coverage, what would make the extra information standout and avoid becoming sidelined by the existing high-profile features? Thus, it is more important to have the quality data than aiming to collect the data that may cover everything. We need to be selective in data collection as well as in analytic methodologies. As long as we bear in mind the scope of our data sources, we do not have to perfect the dataset before starting analyzing it. In fact, an iterative strategy is likely to work more effectively than perfectionism that focuses on one step of the process alone because each step is a learning process and an opportunity to refine our process.

The node selection process determines not only what is relevant in the sense of information matched by information retrieval models but also evaluative indicators such as citations and altmetrics. Evaluative indicators provide information regarding the perceived value of an entity in the universe of knowledge. The value and the relevant of a piece of information may not necessarily correlate. In other words, a highly relevant piece of information may have little value to our analyst who is constructing a systematic review of a domain. In terms of the value of a citation to a publication, it is probably not so much how many times it has been cited so far; rather, it probably matters a whole lot more if thought leaders in the research domain or potentially relevant domains cited it. The widely known PageRank algorithm follows the same principle—the significance of a webpage should be recursively determined by the significance of pages that refer to the webpage. Along this line of reasoning, we should pay more attention to what an article has to say if it is written by a Nobel Prize Laureate, if it has been cited by Turing Award recipients

or by others who have an established reputation in science and technology, or it has been widely cited for reasons that remain unknown.

Citations and altmetric scores are valuable information on how fellow researchers' react upon a scientific publication. Broadly speaking, the more citations an article has received, the more likely that the article has generated an impact on the research community. The more an article has been viewed and downloaded, the more likely the article is interesting. Using citations as an indicator of research impact is controversial. Some argued that since each citation instance may be motivated differently, it may not make sense to add them up as if each of them is equally accountable. Some argued even further that since some citations are supportive, some are neutral, and some are even challenging the original work, lumping these instances of different nature does not make any sense. Others have questioned the assumption that each citation reflects something about the knowledge of a domain because many mistakes or errors in citing a reference evidently show that one cannot assume everyone reads what they cite. Furthermore, researchers have found some citations even distorted the intended meaning of the original source.

There are at least two ways out of these controversies. One is to further classify the types of citations. The other is to clarify the significance of being cited. A good example of the former is the Shepard's Citation Signal in LexisNexis. For each legal case, for example, Miranda v. Arizona, 384 U.S. 436, the Shepard's Citation Signal identifies the types of citations, i.e. signals, that the case has been cited, including warning, questioned, caution, positive treatment, negative treatment, and criticized by. As shown in Fig. 3.3, instances of citations, or citing decisions, are classified into several types. For example, the Miranda v. Arizona (384 U.S. 436) case has been cited in dissenting opinions of the U.S. Supreme Court in cases listed as the items 21–23, i.e. Florida v. Powell, Montejo v. Louisiana, and Dickerson v. United States. Classifications such as the Shepard's Citation Signal are currently rare in scientific literature. The Web of Science and Scopus do not currently provide any citation information below the article level. In other words, we have no other options except assuming the an article cites all its references uniformly even if we know that this is not a good assumption to make. CiteSeer and Google Scholar are probably the most widely known resources of scientific literature where one can find contextual information of a citation. However, we are not aware of any large-scale resources of scientific publications that enable users to search citations by specific classifications of citations.

The latter way to reconcile much of the controversies or the uneasiness surrounding the use of citation counts as an indicator of scholarly impact is to clarify what we mean by impact. Many assume that the impact implies a positive outcome and that one should rule out any negative impact. It is our view that the term scholarly impact should include both positive and negative influences produced by a scholarly contribution. A failure at one level of consideration can be valuable at another level of thinking. Einstein once said the value of his research in his later years is to stop another fool to make the same mistake. If we can learn from a scientific publication either something to follow or something to avoid, it has a
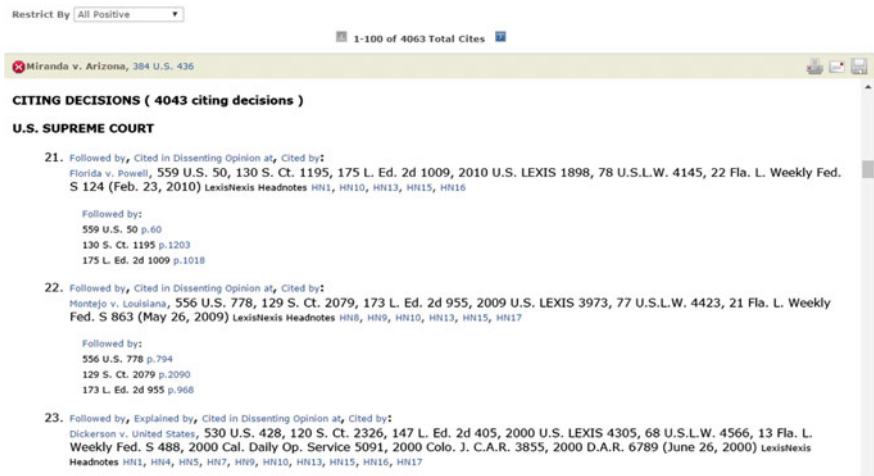
LexisNexis® *Academic*

Restrict By All Positive   ▼

☐ 1-100 of 4063 Total Cites 🖼

❌ Miranda v. Arizona, 384 U.S. 436   🔽 ▢ 💾

CITING DECISIONS (4043 citing decisions)

U.S. SUPREME COURT

21. Followed by, Cited in Dissenting Opinion at, Cited by:
     Florida v. Powell, 559 U.S. 50, 130 S. Ct. 1195, 175 L. Ed. 2d 1009, 2010 U.S. LEXIS 1898, 78 U.S.L.W. 4145, 22 Fla. L. Weekly Fed.
     S 124 (Feb. 23, 2010) LexisNexis Headnotes HN1, HN10, HN13, HN15, HN16

     Followed by:
     559 U.S. 50 p.60
     130 S. Ct. 1195 p.1203
     175 L. Ed. 2d 1009 p.1018

22. Followed by, Cited in Dissenting Opinion at, Cited by:
     Montejo v. Louisiana, 556 U.S. 778, 129 S. Ct. 2079, 173 L. Ed. 2d 955, 2009 U.S. LEXIS 3973, 77 U.S.L.W. 4423, 21 Fla. L. Weekly
     Fed. S 863 (May 26, 2009) LexisNexis Headnotes HN8, HN9, HN10, HN13, HN15, HN17

     Followed by:
     556 U.S. 778 p.794
     129 S. Ct. 2079 p.2090
     173 L. Ed. 2d 955 p.968

23. Followed by, Explained by, Cited in Dissenting Opinion at, Cited by:
     Dickerson v. United States, 530 U.S. 428, 120 S. Ct. 2326, 147 L. Ed. 2d 405, 2000 U.S. LEXIS 4305, 68 U.S.L.W. 4566, 13 Fla. L.
     Weekly Fed. S 488, 2000 Cal. Daily Op. Service 5091, 2000 Colo. J. C.A.R. 3855, 2000 D.A.R. 6789 (June 26, 2000) LexisNexis
     Headnotes HN1, HN4, HN5, HN7, HN9, HN10, HN13, HN15, HN16, HN17

Fig. 3.3 Shepard's analysis definitions

direct impact on our thinking. It has an impact! We should not narrowly limit the meaning of impact to positive ones only.

As we can see, the process of selecting qualified sources is iterative in nature. Initially, we may select publications with many citations already or publications that have no citations yet but have been tweeted and retweeted a lot. At the next level, after we analyze the selected publications, we may be able to apply increasingly sophisticated selection criteria. For example, we may select publications that are known to have strong betweenness or eigenvector centrality scores in the network we have analyzed. We may select articles that are known to have sharp increases in their citation counts. We may also want to focus on articles that have been cited by researchers from at least five specialties. Of course, we may want to pay attention to articles that specifically criticized particular publications.

Interactive Visualizations

Vannevar Bush was the head of the U.S. Office of Scientific Research and Development (OSRD) during the World War II. He envisaged how the mankind's knowledge can be collectively organized by association, the same way as how the human memory works (Bush 1945). Highly connected information resources such as the Internet and the Wikipedia are commonly considered as being inspired by Vannevar Bush's visionary MEMEX. Navigating in such a universe of knowledge

is called trailblazing. The navigator forges trails that represent new connections. As the metaphor of a universe of knowledge may imply, we need to make inter-galactic travels and study information at different levels of granularity. We may be interested in specific causal relations between a virus and a disease. We may be interested in how similar methods are used in different disciplines of science. Interactive visualization is an integral part of visual analytics. It enables us to explore or forage information at various levels of granularity and trace connections across areas where different perspectives may apply.

Ben Shneiderman's mantra for visual information seeking should be very helpful here. It is intuitive and simple to follow. In addition to the useful mantra, it is a good idea for an analyst to get familiar with a few other theories concerning the process of search and what we may expect to find. As people often say, you will only find what you look for. The better we are theoretically prepared, the better position we will be able to place ourselves to recognize potentially relevant patterns. Otherwise, we may miss important clues even if they are right in front of us. We have discussed a few theories of scientific change at the beginning of the book. We will frame our interpretations with these theories and characterize what we would expect to see if the theory is true.

Structural Variation Analysis

The structural variation theory considers the body of a scientific domain's knowledge as a complex adaptive system (Chen 2012, 2014). Its global structure may be altered significantly by newly published articles, or by semantic predications conveyed by these articles. According to the theory, articles that have the potential to trigger global changes are transformative in nature and they are the ones that are most likely to influence the course of the further development of a scientific field. How do we measure such potentials?

If we represent the domain knowledge as a network, then the modularity measure of the network can be very useful for us to assess the global structure of the network. The modularity of a network is defined with reference to a partition of the network. If we can divide the network into smaller components and minimize inter-component connections, then the modularity quantifies the degree to which the resultant components can be separated from one another. The modularity's value ranges from 0 to 1. The highest value of 1 means that the network is completely modularized by the chosen partition. In contrast, the lowest modularity value of 0 means that these components are tightly coupled and one cannot separate them in any meaningful way.

Figures 3.4 and 3.5 illustrate how the system adapts to the publication of the groundbreaking paper by Watts and Strogatz (1998). The network was derived from 5135 articles published on small-world networks between 1990 and 2010. The network of 205 references and 1164 co-citation links is divided into 12 clusters with a modularity of 0.6537 and the mean silhouette of 0.811. The red lines are made by
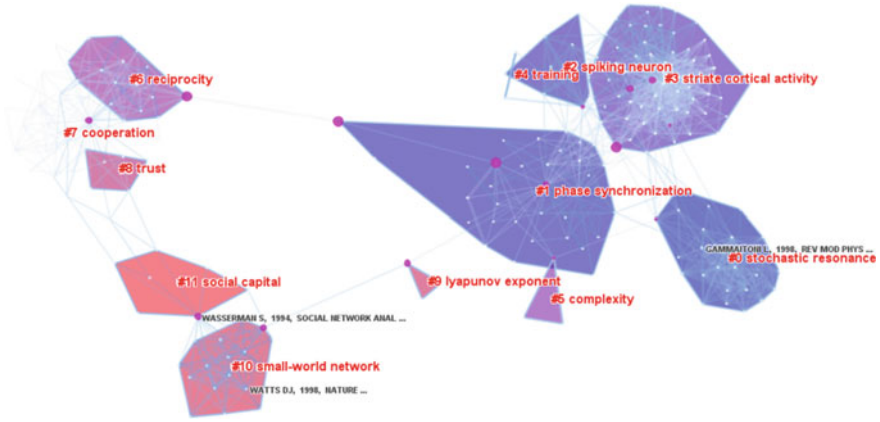
Fig. 3.4 The structure of the system before the publication of the ground breaking paper by Watts and Strogatz (1998)
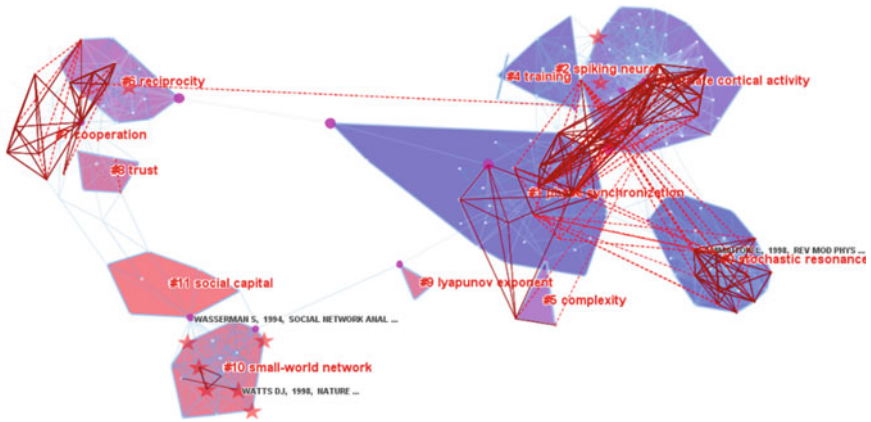

Fig. 3.5 The structure of the system after the publication of Watts and Strogatz (1998)

the top-15 articles measured by the centrality variation rate. Only major clusters' labels are shown in the figure. Dashed lines in red are novel connections made by Watts and Strogatz (1998) at the time of its publication. The article has the highest scores in Cluster Linkage and C_{KL} scores, 5.43 and 1.14, respectively. The figure offers a visual confirmation that the article was indeed making boundary-spanning connections. Recall that the data set was constructed by expanding the seed article based on forward citation links. These boundary-spanning links provide empirical evidence that the groundbreaking paper was connecting two groups of clusters. The emergence of Cluster #8 complex network was the consequence of the impact.

In this view, a network is a system of interconnected blocks. The most fundamental changes for such systems would be changes that alter how existing blocks

are connected as well as adding or eliminating participating blocks. Relatively speaking, changes that are essentially limited to the internal state of a block would not be considered as significant as changes that transform inter-block connections. At the article level, each pair of co-occurring semantic predications introduced in the article is potentially an agent of change or a perturbation signal. If the co-occurring connection falls within a single block, then it may generate a local impact without causing any global changes. In contrast, if the co-occurring connection links two blocks in an innovative way or in a surprising or unanticipated way, then it becomes likely that the new link may change not only the local structure but also how the existing continents are organized. In other words, we should particularly pay attention to the predications of the latter kind.

CiteSpace supports structural variable analysis. Given a set of scientific articles, these articles are separated by the year of their publication. For each article published in year Y, CiteSpace will compute all the changes introduced by the article with reference to a network that represents the state of the knowledge prior to year Y. The differences of the networks before and after the publication of the article are used to quantify the likelihood that the article is altering the global structure of the underlying network in a significant way.

Using MySQL Databases in CiteSpace

CiteSpace has a built-in interface with a MySQL database on your localhost. You can upload your data to the database and interact with your data directly as you would with any MySQL database. You can also interact with your data through special-purpose functions provided in CiteSpace (Fig. 3.6).

For each dataset uploaded to MySQL, you can perform some text analysis functions from the Data Processing Utilities interface. The text analysis functions here are slightly different from the network of co-occurring terms in the main interface of CiteSpace. The major difference is that functions here include a selection step based on log-likelihood ratio tests. In theory, the resultant graph visualization should represent the most important patterns of phrases.

VOSViewer and CitNetExplorer

VOSviewer is a popular science mapping software tool developed by Van Eck and Waltman (2010) at the Centre for Science and Technology Studies (CWTS) in Leiden, the Netherlands, for constructing and visualizing bibliometric networks. These networks may include journals, researchers, or individual publications, and they can be constructed based on co-citation, bibliographic coupling, or co-authorship relations. VOSviewer also offers text mining functionality that can be
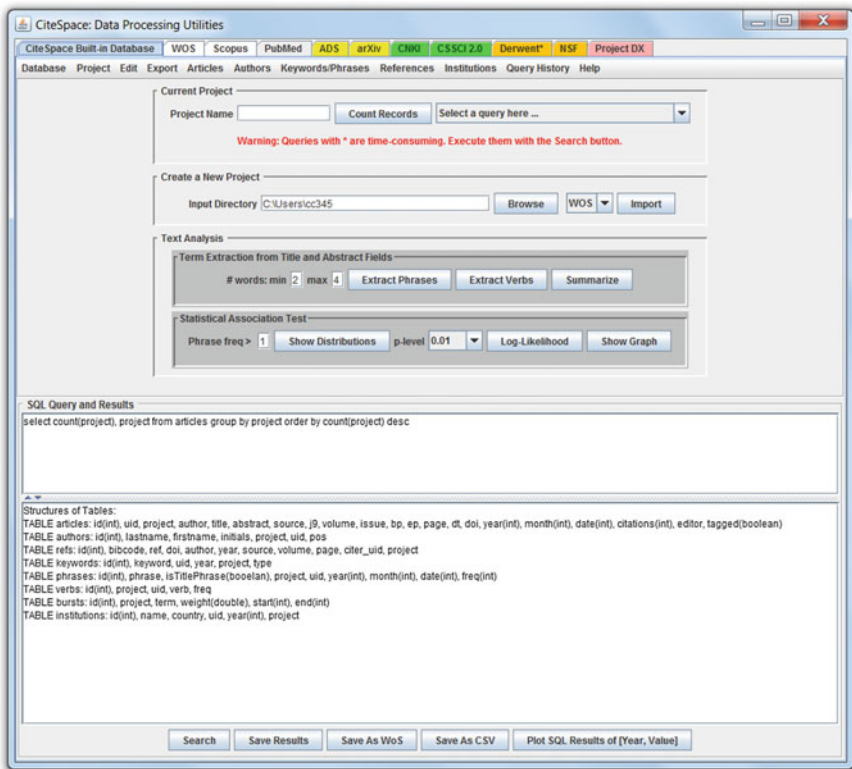
Fig. 3.6 An interface with MySQL in CiteSpace

used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature.

VOSviewer maintains a simple workflow from the data to visualization. It is relatively straightforward to generate a visualization from bibliographic records from the Web of Science, Scopus, and a few other sources. Figure 3.7 shows a density map of references cited in the Science Mapping dataset. Comparing with CiteSpace, a noticeable strength of VOSviewer is its nice and simplistic approach to visualizing scientific publications. On the other hand, the strength in occasions may become a weakness of VOSviewer if the analyst needs to conduct in-depth investigations beyond the initial visualization. Perhaps more importantly, to our knowledge, unlike CiteSpace, the visual design in VOSviewer is not driven by theories of scientific change. For example, VOSviewer does not support concepts such as intellectual turning points nor transformative potentials. Although VOSviewer supports the notion of clusters, it does not provide cluster labels. As a result, one has to rely heavily on the assistant of domain experts or on one's own domain knowledge when interpreting VOSviewer visualizations. In fact, the development of CitNetExplorer (Van Eck and Waltman 2014), by the same team of

Fig. 3.7 A density map visualization in VOSViewer of references cited in the science mapping dataset

VOSviewer, is primarily motivated by the aim to strengthen the relatively weak support of analytic functionality.

CitNetExplorer supports the visualization and exploration of direct citation networks (Van Eck and Waltman 2014). In a direct citation network, a link pointing from a node n_i to a node n_j represents that the article represented by n_i cites the article of n_j. Figure 3.8 shows an example of a direct citation network of articles in the Science Mapping dataset. Articles are arranged vertically based on the year of their publication with the earliest year on the top of the visualization and the latest year at the bottom. The directed citation link is shown vertically. The colors of nodes indicate their clusters.

CitNetExplorer provides more functions for exploring a visualized network, including the drill down function and the display of the shortest path between two nodes. Figure 3.9 shows the resultant network of performing the drill down function on the Science Mapping network. The user can explore the shortest path between two nodes.

There are an increasing number of computer software programs for analyzing scientific publications. Many of them are freely available. Apart from CiteSpace, VOSviewer, and CitNetExplorer, other widely known systems include HistCite,[1] sci2[2] developed at Indiana University, the growing set of programs developed by

---

[1] https://clarivate.com/products/web-of-science/.

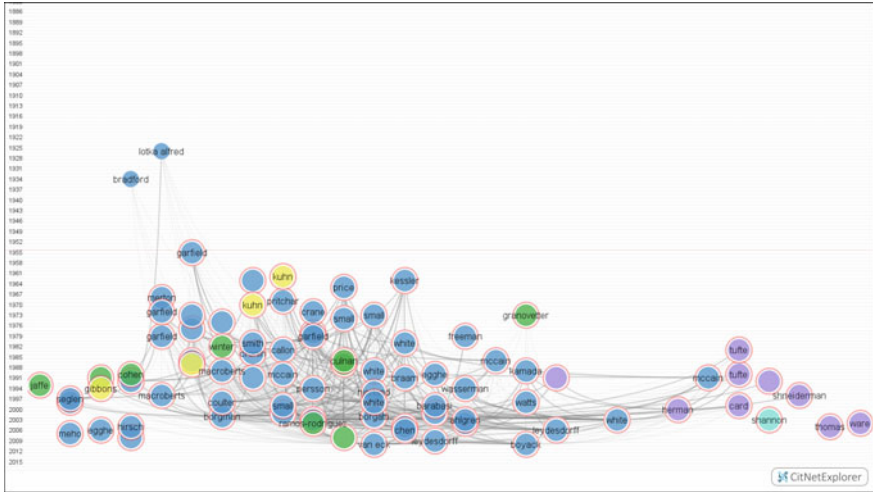[2] https://sci2.cns.iu.edu/user/index.php.

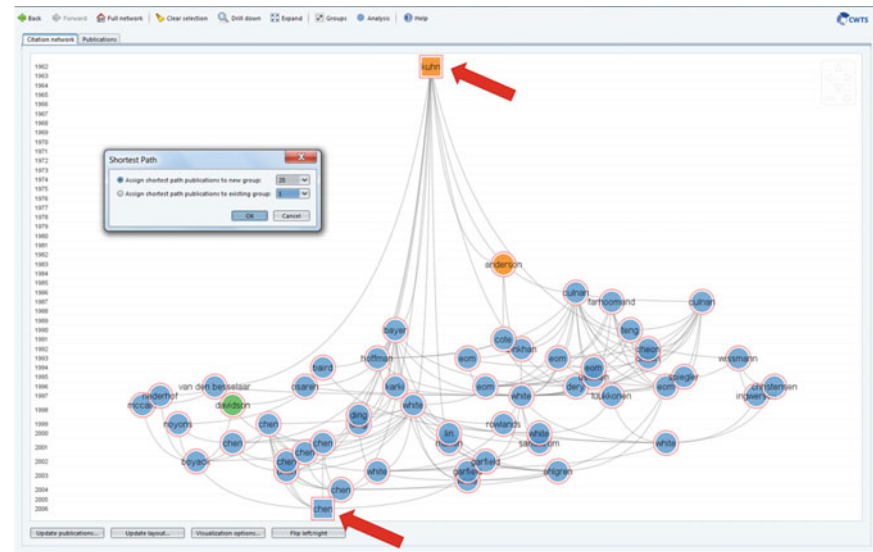**Fig. 3.8** A direct citation network visualized in CitNetExplorer

Fig. 3.9 Drill down and the shortest path between two nodes in CitNetExplorer

Loet Leydesdorff in Amsterdam,[3] Alluvial Generator,[4] KnowledgeMatrix Plus,[5] to name the few. A list of tools and resources is accessible from CiteSpace as well as on the web.[6]

Terrorism Research (1996–2003)

Our first example of exploring the knowledge structure of a research domain is the terrorism research (1996–2003). The source of the science citation data is the Web of Science. The dataset comes with the release of CiteSpace as the Demo 1 project for instructional purposes. Although everyone has probably heard of terrorists, terrorist attacks, and terrorism, many may still have no clear idea what terrorism as the subject of research may include. The lack of prior knowledge of the target domain is probably an accurate description of most of the users of the visual analytic procedure.

The data collection was based on a simple query in the Web of Science. If the term terrorist or the term terrorism appears in the title, the abstract, or the keyword list of an article, then the article is considered relevant and it will be included in the dataset to be analyzed further. This type of search is called topic search in the Web of Science. We used CiteSpace to visualize important patterns in the dataset so that we can explore the visualization and learn about the subject domain. At the end of the process, we should be able to obtain a good understanding of the domain in terms of its overall structure, key groups of publications, and critical works in the field.

Citation Bursts

A relatively simple but effective method is to identify publications in the domain that have drawn attention of the research community at various stages of the development. Burst detection is a reliable technique that enables us to accomplish this task (Kleinberg 2002). Given a sequence of frequency values, a burst is an abrupt elevation of the frequencies over a specific time interval. For example, the number of cars crossing a bridge connecting New Jersey and Pennsylvania everyday may experience bursts during rush hours and the number of cars crossing the bridge every month may experience bursts during holidays. As we have discussed, citations received by scientific publications may provide the first-order

---

[3]http://www.leydesdorff.net/software.htm.

[4]http://www.mapequation.org/apps/AlluvialGenerator.html.

[5]http://mirian.kisti.re.kr/km/km_pop_en.jsp.

[6]http://cluster.cis.drexel.edu/∼cchen/citespace/resources/.

indicator of scholarly impact. In contrast, bursts of citations provide a higher-order indicator of a scholarly impact in terms of the attention from the research community that is evidently above-and-beyond the normally expected level. As a result, publications with strong enough citation bursts during the course of the development of the domain are valuable landmarks for us to navigate the domain further.

Figure 3.10 lists 24 references with the strongest citation bursts between 1996 and 2003. The simplistic diagrams on the right depict the duration of a burst event in red. Overall, the periods of citation bursts drifted over time as new research topics move to the center of the stage. For example, COOPER1983, the second one on the list, has a strong citation burst weight of 5.916. Its citation burst lasted for four years from 1996 till 1999. At this point, although we may not know the specific role played by COOPER1983, we know that this article is evidently

Top 24 References with the Strongest Citation Bursts

References	Year	Strength	Begin	End	1996 - 2003
BRISMAR B, 1982, J TRAUMA, V22, P216	1982	3.6708	1996	1999	
COOPER GJ, 1983, J TRAUMA, V23, P955	1983	5.916	1996	1999	
ENDERS W, 1993, AM POLIT SCI REV, V87, P829	1993	2.2584	1996	1997	
KATZ E, 1989, ANN SURG, V209, P484	1989	4.0719	1996	2000	
PHILLIPS YY, 1986, ANN EMERG MED, V15, P1446	1986	2.3415	1997	1999	
HADDEN WA, 1978, BRIT J SURG, V65, P525	1978	2.8319	1997	1999	
HULLER T, 1970, ARCH SURG-CHICAGO, V100, P24	1970	2.8122	1997	1999	
SCHMID AP, 1988, ..., V, POLITICAL TERRORISM	1988	2.3624	1997	1998	
FRYKBERG ER, 1988, ANN SURG, V208, P569	1988	4.5731	1997	2000	
WATERWORTH TA, 1975, BRIT MED J, V2, P25	1975	2.3415	1997	1999	
MALLONEE S, 1996, JAMA-J AM MED ASSOC, V276, P382	1996	3.044	1998	2000	
SHALEV AY, 1992, J NERV MENT DIS, V180, P505	1992	2.4654	1998	2000	
MESELSON M, 1994, SCIENCE, V266, P1202	1994	2.3691	1999	2001	
*CDCP, 1999, MMWR-MORBID MORTAL W, V48, P69	1999	2.4063	1999	2001	
SIMON JD, 1997, JAMA-J AM MED ASSOC, V278, P428	1997	3.084	1999	2001	
CARTER A, 1998, FOREIGN AFF, V77, P80	1998	2.7523	1999	2001	
OKUMURA T, 1996, ANN EMERG MED, V28, P129	1996	2.4063	1999	2001	
FRANZ DR, 1997, JAMA-J AM MED ASSOC, V278, P399	1997	3.0233	1999	2003	
TOROK TJ, 1997, JAMA-J AM MED ASSOC, V278, P389	1997	2.5347	2000	2001	
ALIBEK K, 1999, BIOHAZARD, V, P	1999	2.3249	2000	2003	
RICHARDS CF, 1999, ANN EMERG MED, V34, P183	1999	3.1263	2000	2001	
INGLESBY TV, 1999, JAMA-J AM MED ASSOC, V281, P1735	1999	2.6297	2001	2003	
KOLAVIC SA, 1997, JAMA-J AM MED ASSOC, V278, P396	1997	2.4539	2001	2003	
INGLESBY TV, 2000, JAMA-J AM MED ASSOC, V283, P2281	2000	3.2797	2001	2003	

Fig. 3.10 Articles with citation bursts in terrorism research (1996–2003)

valuable, especially between 1996 and 1999. We can also tell from the simple depiction that during the same period of time, no other article reached the same level of citation bursts. If we want to invest our precious time on learning more about the research domain, this article should be on our list of a few landmark articles.

In addition to references with the strongest citation bursts, we should also pay attention to references that have the longest duration of citation burst or have the most recent periods of burst. Two of the references have the longest 5-year duration of citation burst, namely KATE1989 and FRANZ1997. We also notice that two of the three most recent bursts from 2001 were authored by INGLESBY in 1999 and 2000, respectively. This is an example of how to identify landmark articles without any prior knowledge of the target domain. This method has a few distinct advantages over using citation counts or altmetrics such as downloads. Usually citation counts are only available as a sum that is accumulated over all the years since the publication of an article. For example, the most cited publication in the Web of Science, the one at the very top of Mount Kilimanjaro, has passed its citation peak many years ago. Citation counts alone cannot tell us whether a highly cited article is still at the center of everyone's attention or its glory is really due to the credit it earned in its golden age that has long gone. Knowing when an article is particularly high performing in drawing the research community's attention is more useful than merely knowing that an article has a lot of citations.

More recent bibliographic records obtained from the Web of Science are likely to include DOIs of cited references. For those references with DOIs, the user can access the full text of a reference through its DOI link, which would be useful for exploring the literature.

Timeline Visualization

CiteSpace supports a few types of visualization, including a cluster view, a timeline view, and a timezone view. A cluster view depicts an overview of a network in a node-and-link diagram. A timeline view still displays the nodes and links but organizes them along multiple parallel timelines. A timeline visualization is intuitive. The analyst can obtain a good overview of the domain with a few simple steps.

Figure 3.11 shows a timeline visualization of the terrorism research (1996–2003). Each line from the left to the right represents a cluster of co-cited references, which in turn reveals the work of a distinct specialty. CiteSpace supports several other types of networks that can be derived from a set of bibliographic records. Here we focus on networks of co-cited references. Studies of this type of networks are called Document Co-Citation Analysis (DCA). Other types of studies include Author Co-Citation Analysis (ACA) and Collaborative Network Analysis.

The timelines are arranged by their size from the largest downwards. The label next to each cluster line summarizes the most likely context in which members of

Fig. 3.11 A timeline visualization of the terrorism research (1996–2003)

the corresponding cluster have been cited. The candidate words for the labels are drawn from articles that cited the members of the cluster. For example, the largest cluster #0 biological terrorism indicates that this cluster is essentially being cited by articles relevant to the topic of biological terrorism.

Note that the citing topics and the cited topics may not be the same necessarily. The difference is the one between the intellectual base and the reference front of the corresponding specialty. In other words, each specialty has two interconnected components: the intellectual base is where the specialty draws its inspiration from and the research front is where the specialty disseminates its new contributions. A good example is the cluster #2 terrorist attack. As we will see shortly, the specialty focuses on the topic of Post Traumatic Stress Disorder (PTDS) in the context of terrorism. The difference between its intellectual base and the research front underlines the nature of the specialty. More specifically, the intellectual base is essentially on PTDS prior to the September 11 terrorist attacks and the research front is mostly produced after. A key difference is that the research front takes a new turn by recognizing the possibility that was not on the radar of PTDS research, namely, people may develop PTSD symptoms even if they have never been on a trauma site physically.

The timeline visualization makes the life cycle of a specialty visible. For example, cluster #1 blast over-pressure has the longest active time—the entire duration of the observation. In contrast, the presence of cluster #6 counter terrorism is much short lived. The timeline visualization also makes it easier to identify active specialties—clusters with many items with circles in red—they have citation bursts.

The analyst can drill down by moving from one level of granularity to another. There are several ways to drill down. CiteSpace allows the user to apply the same analytic procedure repeatedly on a cluster of the network and then on a cluster of the cluster. A research front at the top level may turn out to have finer structures at a

lower level. The user can inspect the titles of those articles that cited a particular cluster and explore various aspects of the cluster. This can be done in CiteSpace with the Cluster Explorer function. Unless we note otherwise, the labels of the clusters are generated from the top 25% of the most cited citing articles for each cluster.

The largest cluster, #0 biological terrorism, has 61 cited references. It has a silhouette value of 0.658. The silhouette value measures the homogeneity of a cluster. Its value ranges between −1 and 1. The value 0.658 is strong enough to make the cluster meaningful, especially as the largest cluster, although we may expect to see higher silhouette values in some domains. The linkage between a citing article and the cluster can be measured in terms of the extent it cited members of the cluster. For cluster #0, there are six citing articles that each cited over 15% of the member references of the cluster (Table 3.1). The one with the strongest linkage, RICHARDS1999, cited 21% of them, is an article published in 1999 entitled "Emergency physicians and biological terrorism." The title of each of the six articles contains the term biological terrorism or bioterrorism, except HAIL1999. The median year of publication is 1999.

Similarly, we can inspect citing articles from the research front of the second largest cluster #1 blast over-pressure, containing 50 references. This cluster has a silhouette value of 0.862, much higher than that of the largest cluster. The median year of publication is 1985. We can also tell from the timeline view that overall this seems to be an older cluster than the largest one. The top three citing articles are shown in Table 3.2.

Two of them explicitly mentioned blast over-pressure. In fact, one mentioned blast over-pressure and the other mentioned blast overpressure-induced injury. Although the two semantically equivalent terms do not have the identical forms, they are grouped together because of the references they cited. This is an additional advantage of citation indexing, as opposed to approaches purely based on matching words or lexical patterns.

The next cluster is #2 terrorist attack, containing 47 references and an even higher silhouette value of 0.915. The median year of publication of the cluster is 1994. The first three strongest citing articles to this cluster all have the term terrorist attacks in their titles (Table 3.3). In fact, they all contain the longer phrase September 11th terrorist attacks in their titles. It is also clear that the central theme is to do with PTSD after the September 11th terrorist attacks. Each of the three articles has cited over 21% of the members.

In summary, by visualizing the citation patterns in articles published between 1996 and 2003, we have learned that the three most prominent areas of the research domain are bioterrorism, injuries caused by blast over-pressure, and PTSD caused by September 11th terrorist attacks. We have also found gateways that we could drill down as further as we like. We can check where those landmark articles are located and pinpoint articles that played critical roles in the course of the development of the complex domain. As we have seen, we can reach this macroscopic level of understanding of a domain that we knew little about. This is largely due to the way we tap into the domain expertise of numerous researchers through their publications. This procedure is generic. It is applicable to a wide range of scientific disciplines.

Table 3.1 Major citing articles of Cluster #0

Coverage	Citing article
0.21	Richards, CF (1999) Emergency physicians and biological terrorism
0.20	Atlas, RM (1999) Combating the threat of biowarfare and bioterrorism
0.16	Inglesby, TV (1999) Anthrax as a biological weapon—medical and public health management
0.16	Relman, DA (2001) Bioterrorism preparedness: what practitioners need to know
0.15	Dhawan, B (2001) Bioterrorism: a threat for which we are ill prepared
0.15	Hail, AS (1999) Comparison of noninvasive sampling sites for early detection of bacillus anthracis spores from rhesus monkeys after aerosol exposure

Table 3.2 Major citing articles of Cluster #1

Coverage	Citing article
0.58	Elsayed, NM (1997) Toxicology of blast over-pressure
0.24	Elsayed, NM (1997) A proposed biochemical mechanism involving hemoglobin for blast overpressure-induced injury
0.04	Stein, M (1999) Medical consequences of terrorism—the conventional weapon threat

Table 3.3 Major citing articles of Cluster #2

Coverage	Citing articles
0.28	Galea, S (2002) Posttraumatic stress disorder in manhattan, New York City, after the September 11th terrorist attacks
0.23	Vlahov, D (2002) Increased use of cigarettes, alcohol, and marijuana among manhattan, New York, residents after the September 11th terrorist attacks
0.21	Galea, S (2002) Psychological sequelae of the September 11 terrorist attacks in New York City

In the following example, we will look at the terrorism research again, but this time over a wider window, especially containing articles published between 1980 and 2017. We would like to see where the three prominent topic areas are located in the broader context. We would also like to see any major topic areas emerged since 2003.

Structural Variations

The goal of a structural variation analysis is to identify two types of links added to the current network representation of a domain's knowledge, namely, incremental links and transformative links (Chen 2012). Incremental links are within the boundary of a particular cluster, whereas transformative links connect different clusters. Thus, incremental links do not change the structure of the system at the cluster level, but transformative links do.
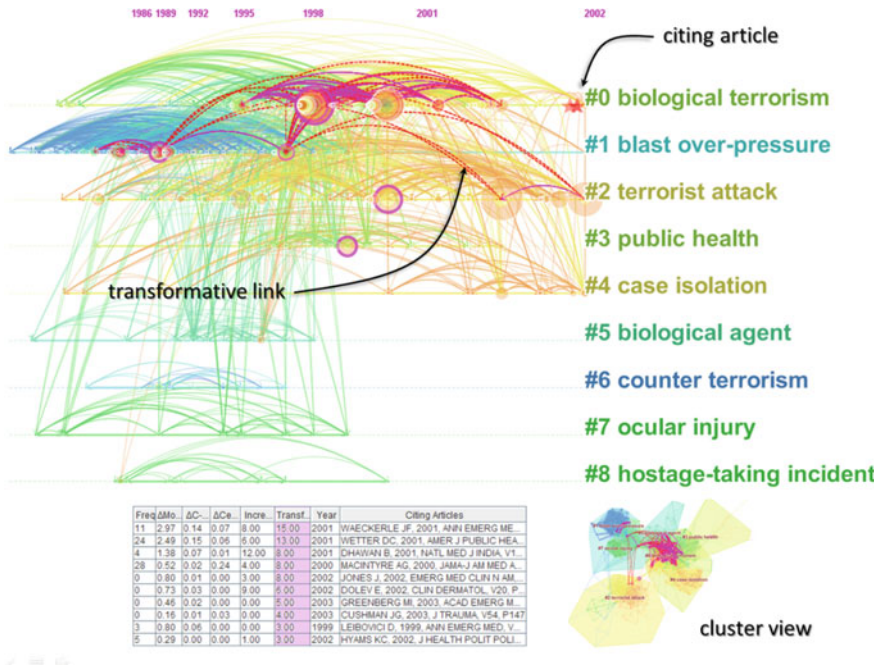
Fig. 3.12 Structural variation by transformative link count

Given a particular year Y, articles published in year Y will be examined against the structure of the network representation of the domain's knowledge over the last three years prior to Y. We refer this network as the baseline network for year Y. The analyst may choose the number of years prior to year Y to form the baseline network. The longer the baseline network extends back in time, the more accurate the structural variation measures would be because a network with a longer exposure time is likely to capture more links than a network with a shorter exposure time.

Figure 3.12 shows the footprints of top 10 articles with the largest modularity change rate. Dashed lines represent novel transformative links. Solid lines represent existing lines. Transformative links mostly connect the largest three clusters, namely, #0, #1, and #2.

Terrorism Research (1980–2017)

We used the same simplistic topic search in the Web of Science using broader terms of terrorist OR terrorism and limited to two types of publications: articles of original research and review articles. The new search found 14,656 relevant records. If one prefers to obtain additional articles that didn't use these topic search terms but may be relevant otherwise, one option is to use the citation expansion strategy to include

articles that cite this set of records. We did not perform the citation expansion for this particular case because it is adequate for our purpose to focus on the scope defined by the topic search.

We used the g-index to select articles for each time slice between 1980 and 2017. In addition, these articles must have received two or more citations themselves. Since we are dealing with a timespan of 38 years, imposing this minimum citation condition can filter out many publications that do not make sufficient impact on the research domain. Admittedly, this condition is likely to be relatively harsh for recently published articles, although the use of g-index may compensate the citation distribution to an extent. One remedy is to conduct a separate study using a lower threshold on articles published within the recent few years. We limited the Look Back Years to 5, which means we will ignore citations to references that are more than five years ago.

Figure 3.13 shows the cluster view visualization of the terrorism research (1980–2017). It depicts the largest connected component of the network of 908 cited references. The largest connected component contains 694 references (76% of the entire network). Each cluster is shown with a polygon colored to indicate the median of its citing articles' publication years. In this visualization, clusters located near the top are the oldest, whereas clusters near the bottom are the most recent ones.
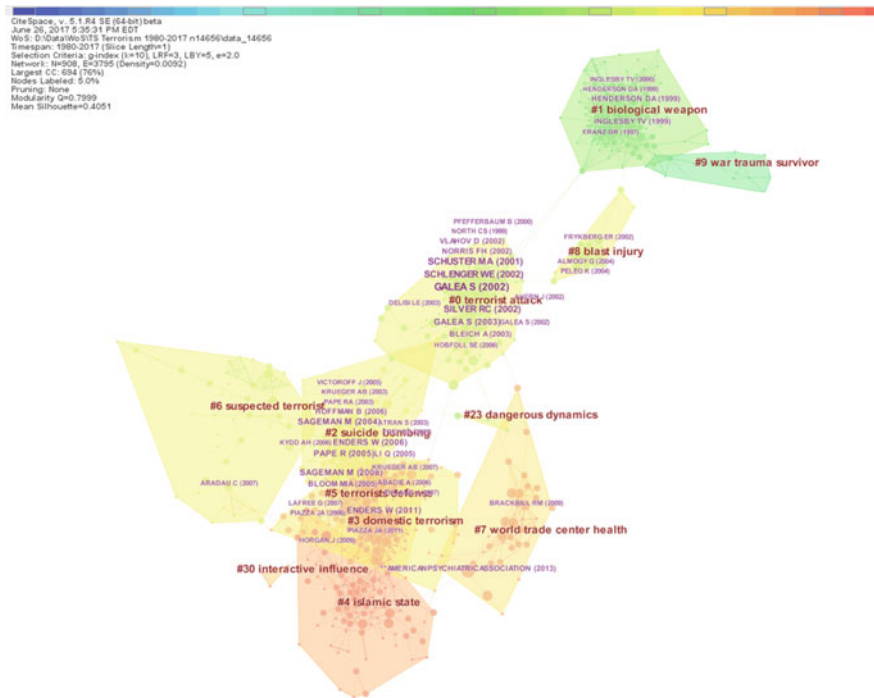
Fig. 3.13 A cluster-view visualization of the terrorism research (1980–2017). Node selection by g-index (k = 10)

The oldest cluster in the visualization is #9 war trauma survivor. Cluster #1 biological weapon is the second oldest, containing two articles authored by INGLESBY, which remind us the biological terrorism cluster identified in the terrorism research (1996–2003). Moving downwards, the cluster #8 blast injury is likely to be connected to the blast over-pressure cluster identified in the 1996–2003 study. The cluster #0 terrorist attack and references such as GALEA2002 indicate that this is the PTSD cluster identified before.

Moving further down, we encounter clusters such as #2 suicide bombing, #3 domestic terrorism, #7 word trade center health, and #4 islamic state. Evidently, many new topic areas emerged since 2003.

In Fig. 3.14, the network of the terrorism research in 1996–2003 is superimposed over the network in 1980–2017. This function is called a network overlay, which highlights the relationship between a subnetwork and a larger network.
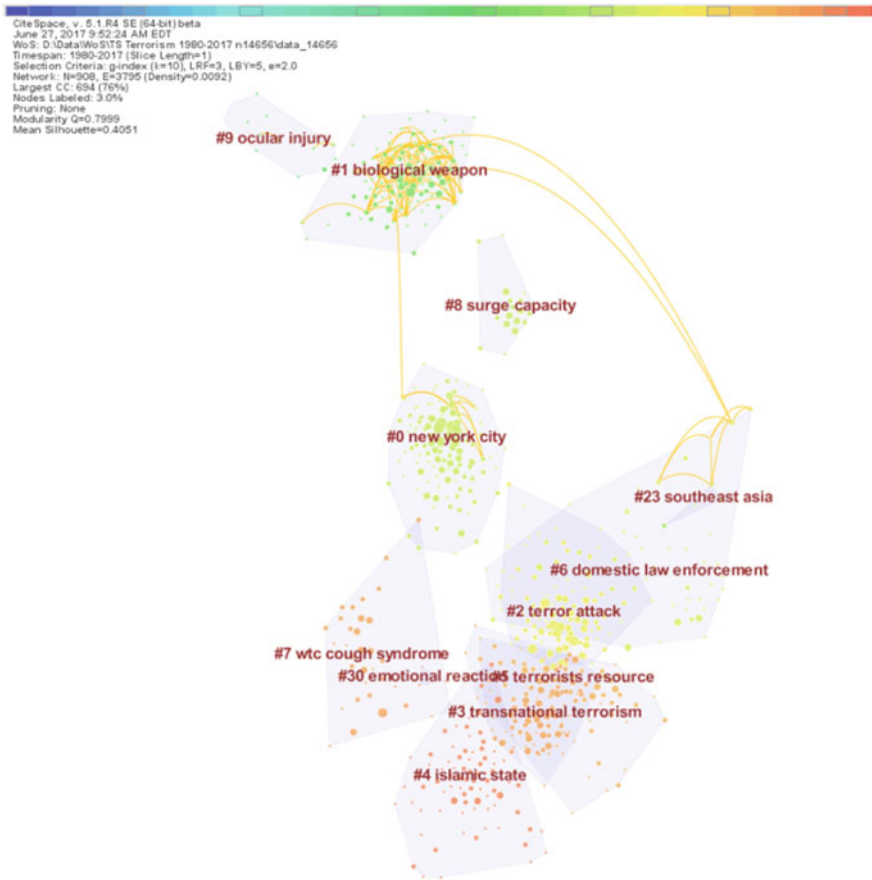
Fig. 3.14 A network overlay shows the 1996–2003 network in the context of the 1980–2017 network
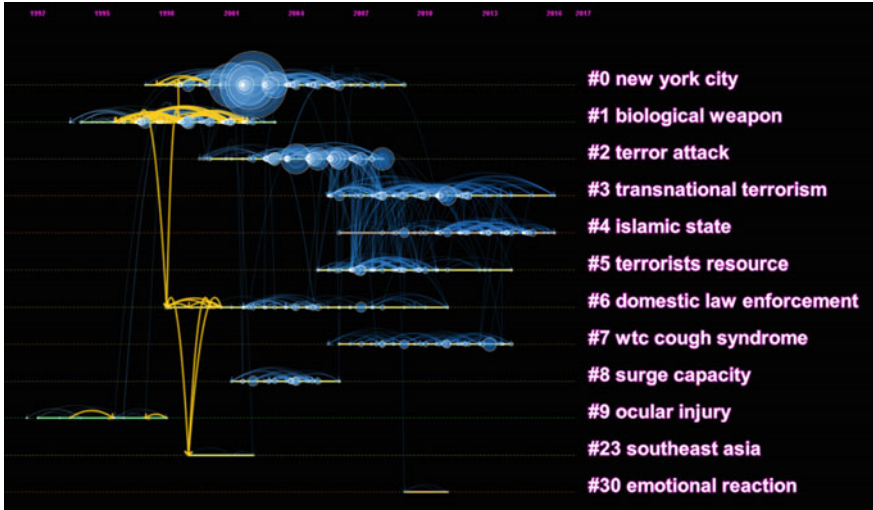
Fig. 3.15 A timeline visualization of terrorism research (1980–2017)

A new timeline visualization of the terrorism research (1980–2017) is shown in Fig. 3.15. The lines in yellow are from the earlier network (1996–2003). The timeline view shows a big picture of the terrorism research. The previously predominant topic areas such as bioterrorism and PTSD are no longer in active at this level of granularity, although there may be publications that are excluded by our selection criteria. In contrast, the two current lines of research are #3 transnational terrorism and #4 islamic state.

Using the Cluster Explorer function in CiteSpace, we can inspect the major clusters and see what the major topics are and how they may differ from their counterparts in the earlier visualization. The previously third largest cluster on PTSD now becomes the largest cluster of 127 references and an even higher silhouette value than before (0.962), which means that the specialty becomes more specialized. The median year of the cluster is 2003. As we can tell from the timeline view, the cluster remained to be active until about 2009. As shown in Table 3.4, the top five citing articles of the largest cluster are clearly related to the September 11 terrorist attacks. The major theme of PTDS and mental health in general continues the theme of the PTSD cluster identified in the previous study.

The second largest cluster #1 biological weapon has 116 member references with a very high silhouette value of 0.972. The median age of the cluster is 1999. As shown in the titles of the top five citing articles, this cluster is clearly about bioterrorism and biological weapons (Table 3.5). We notice that ATLAS1999 also appears in the biological terrorism cluster identified in the previous study. The timeline of the cluster stopped at 2003. On the other hand, there are some connections between this cluster and a few other clusters, notably #2 terror attack and #6 domestic law enforcement. It is possible that topics concerning bioterrorism may have transformed into research topics under other clusters. It is also possible, of course, that the topic of bioterrorism is no longer an active line of research.

Table 3.4 Major citing articles of Cluster #0 in Terrorism Research (1980–2017)

Coverage	Citing article
0.08	Boscarino, JA (2004) Mental health service use 1-year after the world trade center disaster: implications for mental health care
0.06	Boscarino, JA (2004) Mental health service and medication use in New York City after the September 11, 2001, terrorist attack
0.06	Adams, RE (2006) Alcohol use, mental health status and psychological well-being 2 years after the world trade center attacks in New York City
0.06	Boscarino, JA (2004) Adverse reactions associated with studying persons recently exposed to mass urban disaster
0.06	Pulcino, T (2003) Posttraumatic stress in women after the September 11 terrorist attacks in New York City

Table 3.5 Major citing articles of Cluster #1 in Terrorism Research (1980–2017)

Coverage	Citing articles
0.11	Atlas, RM (1999) Combating the threat of biowarfare and bioterrorism
0.09	Fidler, DP (1999) Facing the global challenges posed by biological weapons
0.09	Greenfield, RA (2002) Bacterial pathogens as biological weapons and agents of bioterrorism
0.09	Klietmann, WF (2001) Bioterrorism: implications for the clinical microbiologist
0.08	Atlas, RM (2001) Bioterrorism before and after September 11

The third largest cluster #2 terror attack has 95 references and a silhouette value of 0.835. This cluster is relatively younger than the first two clusters. Its median age is 2004. The relatively low silhouette value is perhaps reflected on the lack of a clear consensus among the top five citing articles' titles (Table 3.6). On the other hand, the timeline view shows that this cluster has connections with #3 transnational terrorism and #5 terrorist resource.

The cluster #3 transnational terrorism is a currently active area of research. With 85 references, it has a silhouette value of 0.834 and an even younger median age of 2010. Along with cluster #4 islamic state, this cluster represents essentially the current research focus of the terrorism research community. The titles of the top citing articles suggest that research in this cluster is concerned with questions concerning the causes of terrorism (Table 3.7).

Cluster #4 islamic state is the youngest one, containing 84 references with a silhouette value of 0.891 and the median age of 2012. Four of the top five citing articles of the cluster were published in 2016 (Table 3.8). The titles of the top 5 citing articles suggest that the cluster focuses on deeper reasons of terrorism.

Figure 3.16 shows a timeline visualization that is rendered with citation bursts in red circles, which correspond to the duration of citation burst. Almost every reference in the visualization had a citation burst. We have not seen this degree of citation burst in other domains we have analyzed. We will drill down one more level deeper to characterize each cluster's theme in more detail.

Table 3.6 Major citing articles of Cluster #2 in Terrorism Research (1980–2017)

Coverage	Citing articles
0.17	Gould, ED (2010) Does terrorism work?
0.12	Rosendorff, BP (2010) Suicide terrorism and the backlash effect
0.07	Czinkota, MR (2010) Terrorism and international business: a research agenda
0.07	Fielding, D (2010) 'An eye for an eye, a tooth for a tooth': political violence and counter-insurgency in egypt
0.07	Plumper, T (2010) The friend of my enemy is my enemy: international alliances and international terrorism

Table 3.7 Major citing articles of Cluster #3 in Terrorism Research (1980–2017)

Coverage	Citing articles
0.18	Gassebner, M (2011) Lock, stock, and barrel: a comprehensive assessment of the determinants of terror
0.18	Krieger, T (2011) What causes terrorism?
0.11	Berrebi, C (2011) Earthquakes, hurricanes, and terrorism: do natural disasters incite terror?
0.11	Chenoweth, E (2013) Terrorism and democracy
0.09	Freytag, A (2011) The origins of terrorism: cross-country estimates of socio-economic determinants of terrorism

Table 3.8 Major citing articles of Cluster #4 in Terrorism Research (1980–2017)

Coverage	Citing articles
0.07	Pearson, E (2016) The case of roshonara choudhry: implications for theory on online radicalization, ISIS women, and the gendered Jihad
0.06	Horgan, J (2016) Actions speak louder than words: a behavioral analysis of 183 individuals convicted for terrorist offenses in the united states from 1995 to 2012
0.06	Schuurman, B (2016) Rationales for terrorist violence in homegrown Jihadist groups: a case study from The Netherlands
0.05	Capellan, JA (2015) Lone wolf terrorist or deranged shooter? a study of ideological active shooter events in the united states, 1970–2014
0.05	Cold, JW (2016) Extremism, religion and psychiatric morbidity in a population-based sample of young men

Semantic Structures of Clusters

Titles of leading articles that cite a cluster and summarizing labels of a cluster provide a top-level characterization of the predominant theme of the cluster. Developing a deeper understanding of each cluster's theme is possible if we can construct an ontological structure of key concepts associated with the cluster.

The procedure consists of the following steps. First, we extract terms from articles that cite members of clusters such that extracted terms are representative to
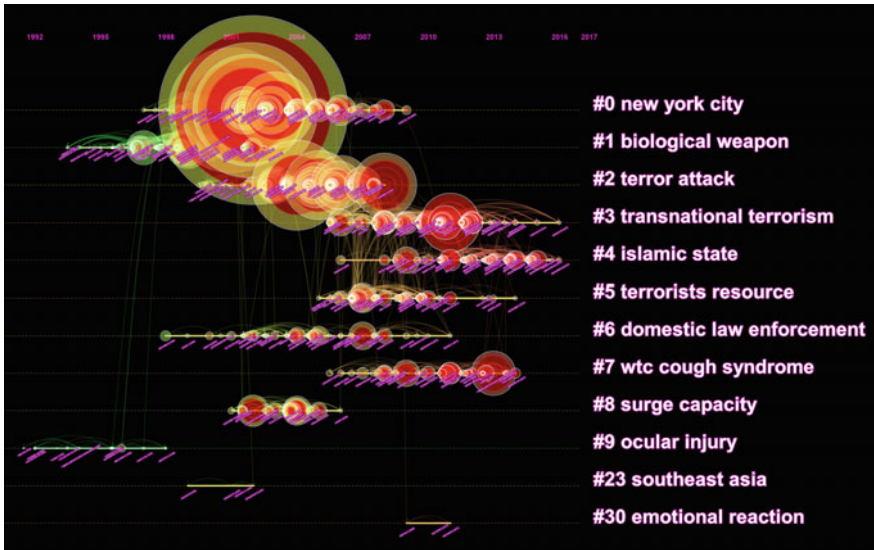
Fig. 3.16 A timeline visualization of citation bursts

individual clusters. Terms may come from titles, abstracts, and/or keywords of these articles. Next, the extracted representative terms are filtered by a given cluster so that we will only retain terms that actually appeared in the cluster. Then, co-occurrences of filtered terms within citing articles of the clusters are identified. These co-occurrences are used as the input to the construction of a hierarchical structure. Since co-occurring terms form a network, hierarchical relations between terms can be derived based on the concept of m-reachability. A term with a higher reachability is assigned to have a higher level position. The resultant hierarchical structure is finally visualized as a concept tree. A concept tree is a hierarchically organized set of concepts. Since our terms are representative of their own clusters, the resultant concept tree serves as a proxy of an ontological representation of the cluster's content.

Figure 3.17 shows a closer view of the largest cluster #0. It primarily focuses on PTSD resulted from the September 11 terrorist attacks in New York. Its member references are published between 1997 and 2009. The largest circles belong to SCHUSTER2001, GALEAS2002, and GALEAS2003. Figure 3.18 depicts the concept tree of terms extracted from the abstracts of the articles that cited the cluster. The root of the tree is on the left. The children of a term are placed on its right-hand side. Nodes that do not have any children nodes are called leave nodes. A path starting from a node to a leave node consists of all nodes along the way. The length of a path is the number of these nodes. The longest path in the concept tree of cluster #0 contains 5 nodes. In fact, three paths have the same length. They share the first four concepts: mental health treatment → terrorist attack → mental health status → New York City, then the path splits into three more specific concepts:
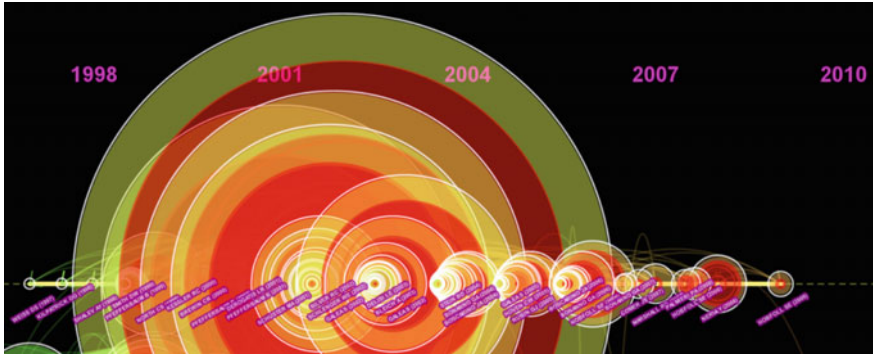
Fig. 3.17 Cluster # New York City (1997–2009)

Fig. 3.18 A concept tree of cluster #0 based on terms extracted from the abstracts of its citing articles, i.e. the research front

representative sample, probable PTSD, and mental health service. We may consider the longest path as the main path that characterizes the fundamental theme of a cluster. The term main path in network analysis has a special meaning. We address how to conduct a main path analysis in next section. Here we use the term main path for its meaning in its intuitive sense.

The mental health branch is the largest one for the cluster. It echoes what we have learned from the visualizations so far but now more specific contextual details will be very valuable for us to strengthen our understanding of the specialty. For example, the mental health dimension is particularly concerned with terrorist attack, which is further connected to terms such as world trade center. The post-traumatic stress disorder leads a branch of its own. Another branch is alcohol dependence, which is also related to mental health. The concept tree further clarifies the knowledge structure of the largest cluster in terrorism research.

Similarly, we obtained a concept tree for cluster #1 biological weapon, which has three branches (see Fig. 3.19). The largest branch starts with biological weapon, followed by civilian population, which leads to four leave nodes. Under the biological weapon node, we can see topics on bacillus anthracis, review article, growing concern, and mass destruction. The second branch consists of bioterrorist attack, bioterrorism preparedness, and public health.
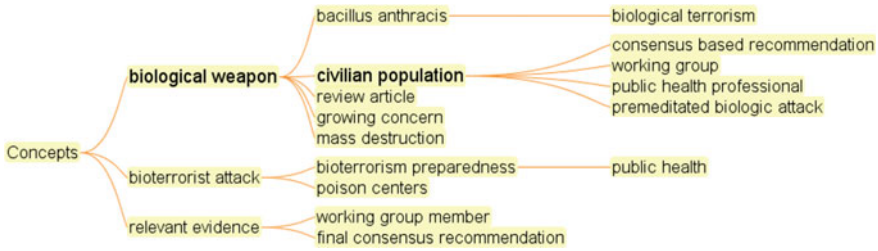
Fig. 3.19 A concept tree of Cluster #1 biological weapon
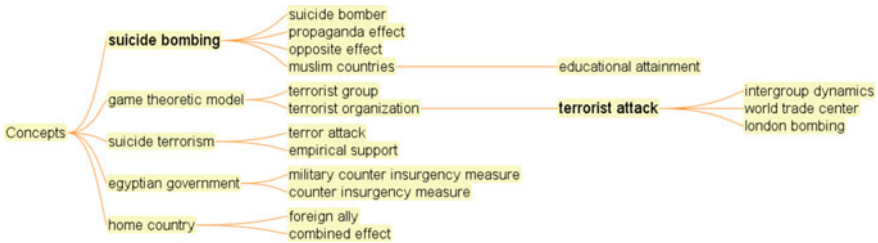
Fig. 3.20 A concept tree of Cluster #2 terror attack

The third largest cluster's concept tree contains branches on suicide bombing, game theoretical model, and three smaller branches (see Fig. 3.20). The suicide bombing branch appears to focus on individual suicide bombing, whereas the game theoretic model branch appears to focus on group dynamics and organizations of terrorist groups. Given that the overall cluster is labeled as terror attack, these branches further elaborate the research focus in these areas.

Recall that the two currently active specialties of research are associated with clusters #3 transnational terrorism and #4 islamic state. In cluster #3, transnational terrorism is a prominent concept, which leads to a few related concepts and a branch of several levels deep (Fig. 3.21). Although the information is still patchy, we can learn the most relevant vocabulary in the context of the cluster, including advanced democracies and domestic terrorism. Some contradicting terms such as democracies and nondemocratic countries appear to underline the role of democracy or the lack of it in understanding transnational terrorism. Similarly, the contrast between transnational terrorism and domestic terrorism can be observed in the concept tree as well. Closely related terms such as international terrorism and transnational terrorism invite further investigations on how these terms differ and how they are related. We will illustrate shortly how we can address these questions by exploring the actual contexts in which these concepts are discussed. We will construct a full-fledged concept tree of terms identified in the abstracts of citing articles. Furthermore, we can instantly reveal the instances of a given concept in their original contexts. The current discussion is at a higher level of granularity than

Fig. 3.21 A concept tree of Cluster #3 transnational terrorism based on terms extracted from abstracts of citing articles to the cluster

Fig. 3.22 A concept tree of Cluster #4 islamic state
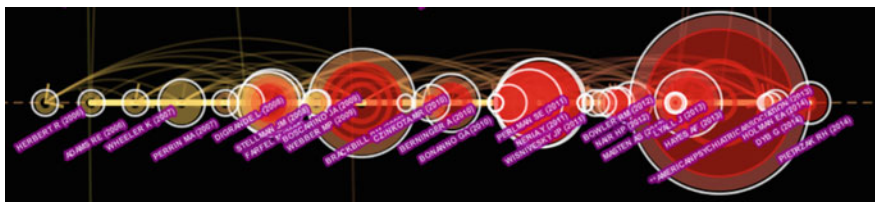
Fig. 3.23 Cluster #7—WTC cough syndrome

the full-fledged concept-to-concept exploration. We will drill down to the next level of granularity after our exploration at the current level. This is also a recommended search strategy. Instead of diving into a specific area first, seek a good understanding of the system at one level of granularity at a time.

Cluster #4 Islamic state is another cluster that is still active. The concept tree reveals some key concepts of this cluster such as religious extremism, psychiatric morbidity, and far right extremist (Fig. 3.22). We will also drill down this cluster deeper using a concept tree that is formed to reveal the concept-in-context details.

Figure 3.23 shows the timeline of cluster #7 WTC cough syndrome. Its publications range between 2006 and 2014. There are a few big and red circles, indicating that they are not only highly cited but also have strong citation bursts, namely, BRCKBILL2009, WISNIVESKY2011, and a 2013 publication of the American Psychiatric Association.

As shown in Fig. 3.24, this cluster's concept tree contains concepts concerning a particular population such as wtc disaster worker and wtc exposed firefighter, symptoms related concepts such as respiratory symptom, and PTSD related topics such as baseline PTSD symptom count.
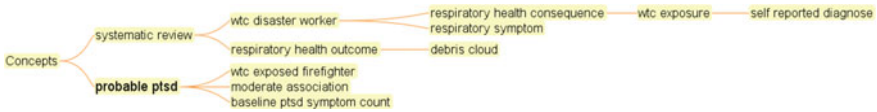
Fig. 3.24 A concept tree of cluster #7 world trade center cough syndrome
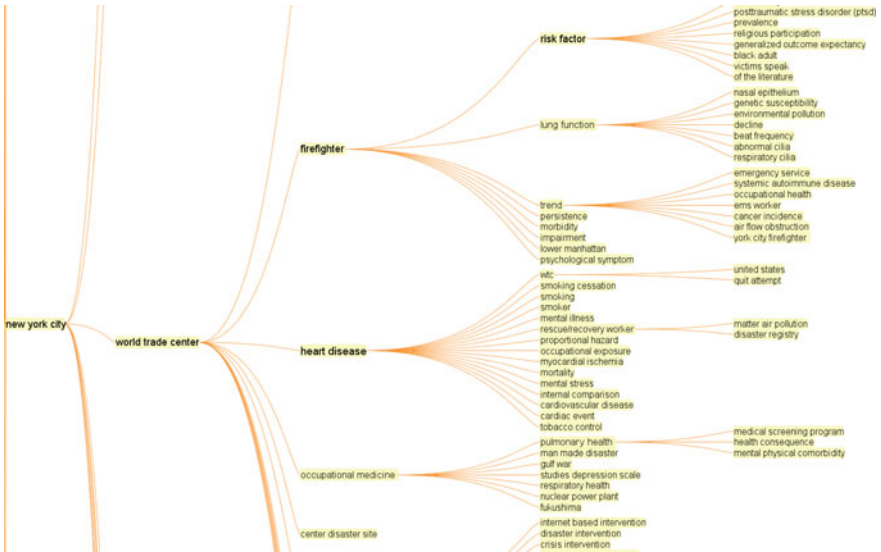
Fig. 3.25 An unfiltered concept tree of the WTC cough syndrome cluster (#7)

Figure 3.25 shows part of an unfiltered concept tree of the WTC cough syndrome cluster. It retains terms that are not deemed to be unique to the cluster, but the inclusion of such terms provides details that may be missing from the concept tree based on filtered terms. The hierarchical relations are easy to understand. For example, the New York city is the parent node of world trade center, which in turn has children nodes such as firefighter, heart disease, occupational medicine and a few other branches. Along the firefighter branch, the sub-branch of risk factor is prominent with many children nodes such as PTSD. The firefighter branch also includes a sub-branch of lung function with more specific terms such as nasal epithelium and respiratory cilia. In parallel to the firefighter branch, the heart disease branch is also prominent due to the number of its children nodes on smoking and mental illness. The contextual information provided by the concept tree is valuable as we can better plan our search strategy and explore the knowledge domain more effectively. Furthermore, the provision of such a concept tree serves the role of an organizing framework so that we can organize various concepts encountered in our search more easily, which tends to reduce the overall complexity of the task.

Concepts in Context

One way to learn about a subject domain is to explore how a concept is used in a
variety of contexts by researchers in this specialty. Although the concept tree shown
in Fig. 3.26 resembles those concept trees we have seen earlier, they differ in some
important ways. First of all, in this concept tree, children nodes represent attributes
of their parent node. For example, the concept of aid is divided into foreign aid and
military aid. In fact, children nodes are modifiers of their parent node in the original
text. If we organize the details in this way, we would be able to see the most
common features on the left and more specific features on the right, or further down
the tree structure. Secondly, the hierarchical relations differ in their semantics.
A parent-child hierarchical relation in the concept-in-context tree indicates that the
child node serves as a modifier of the parent node as in the aid-foreign example, the
term foreign modifies the term aid. In contrast, the parent-child hierarchical relation
in a concept tree in previous sections represents a broad-narrow relation as in New
York City → World Trade Center.

   We can interactively explore the original text not in the conventional way to read
the text in its original sequential order; instead, we can hop over the text and read
various contexts side by side. In the foreign aid example, as we hover over the
foreign-aid node with the mouse, a list of sentences will appear in a window. All
these sentences are about the concept foreign aid in the context of the transnational
terrorism cluster. We can see various topics concerning foreign aid, for example,
using foreign aid as a counterterrorism instrument and identifying sectors that have
been particularly effective as the target of foreign aid such as education and health.

   A common theme in cluster #3 is that democracies tend to experience more
terrorism than dictatorships, autocracies, or other non-democracies (see Fig. 3.27).
Much of these discussions are revolving around the democracy-autocracy divide,
for example, newly established democracies are more vulnerable to terrorism than
established democracies (1) and democracies experience more terrorism than
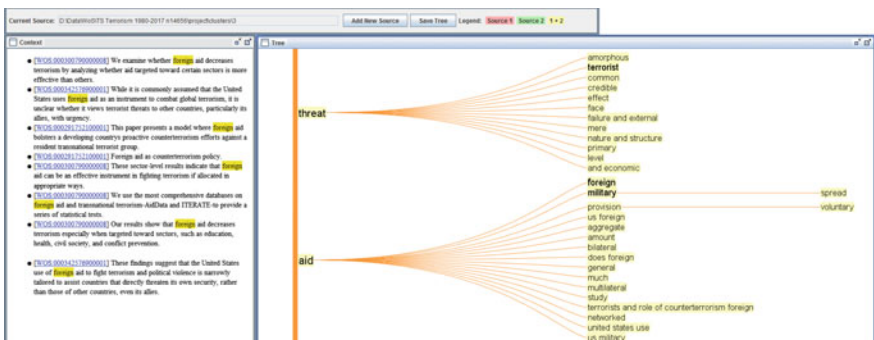non-democracies (2–4).

Fig. 3.26 The contexts of foreign aid in a concept-in-context tree of cluster #3 transnational
terrorism

- [WOS:000317343900006] One of the earlier empirical studies of the relationship between regime type and terrorism published in International Interactions determined that while established democracies were significantly less likely to experience terrorist attacks than were nondemocratic countries, newly established democracies were highly vulnerable to terrorism. **1**
- [WOS:000332371500009] The study finds that within-system groups are significantly less likely to emerge in democracies that have a proportional representation system and higher levels of district magnitude, while neither of these factors affects the emergence of anti-system groups.
- [WOS:000325495000011] Although empirical research has generally demonstrated that democracies experience more terrorism than autocracies, research suggests that this depends upon complex institutional differences that go beyond the democracy-autocracy divide. **2**
- [WOS:000319804100003] Research on terrorism in democracies borrows from the literature on civil war and rebellion to argue that more proportional representation decreases the likelihood of terrorist violence.
- [WOS:000284604200005] Conceiving democratic rule of law as the coexistence of effective and impartial judicial systems and citizens recognition of the law as legitimate, the author presents a causal explanation in which a high-quality rule of law is considered to dampen ordinary citizens opportunity and willingness to engage in political violence, protecting democracies from becoming victims of terrorism.
- [WOS:000321769600017] Intermediately wealthy and transitioning democracies with internally inconsistent institutions were more likely to experience domestic terrorism than advanced democracies and authoritarian regimes.
- [WOS:000341817900009] Approaching elections lead to an increase in the volume of attacks in democracies with low electoral permissiveness but not in others.
- [WOS:000332371500009] A wide range of studies find that democracies experience more terrorism than non-democracies. **3**
- [WOS:000273945700002] Why is terrorist activity more prevalent in democracies than in nondemocracies? **4**

Fig. 3.27 Sentences that mentioned the term democracies in abstracts of Cluster #3 transnational terrorism

- [WOS:000375999900005] From the outset, radicalization was conceived of as an intellectual process through which an individual would increasingly come under a spell of extremist ideas.
- [WOS:000375999900005] For more than a decade, radicalization has been a keyword in our understanding of terrorism. **1**
- [WOS:000356993800001] This Article examines the federal governments community engagement efforts with American Muslim communities as part of a larger infrastructure for policing radicalization and countering violent extremism (CVE).
- [WOS:000375999900005] Considering recent events, including the November 2015 Paris attacks, the present article sets out to reassess the above-mentioned intellectualist understanding of radicalization and come up with new suggestions as to how radicalization may be understood today.
- [WOS:000376226800002] It appears that radicalization is grounded on a vacuum of existential significance and by an ascending process, through the sharing of an ideology prone to violent actions, leads individuals to engage in terrorist goals for a collective one, greater than them.
- [WOS:000375999900005] In the second part of the article it is argued that although radicalization is often conceived of as an individual process, pathways towards terrorism are inherently social and political. **2**
- [WOS:000350893700005] In particular, little is known about the early processes and pathways to radicalization.
- [WOS:000352355500002] While the cases of Anders Behring Breivik and Mohamed Merah clearly demonstrate the impact of social networks and the role of the Internet and prison on the radicalization process, the killings in Norway and France in fact expose larger issues that exist within contemporary Europe, including profound identity crises manifesting as Islamist extremism in some others and far-right extremism in others. **3**
- [WOS:000289794000004] Transformative learning theory, developed from the sciences in education and rehabilitation, offers an interdisciplinary lens with which to study the processes of personal change associated with radicalization.
- [WOS:000375999900005] But the way we understand radicalization has specific consequences for the way we manage and fight the scourge of terrorism.
- [WOS:000281084600001] The aim is to take stock of the current state of research within this field and to answer the question: From an empirical point of view, what is known and what is not known about radicalization connected to militant Islamism in Europe?.

Fig. 3.28 Some of the contexts of radicalization in cluster # Islamic state

Radicalization is a key concept in cluster #4 (see Fig. 3.28). A common theme is the process, or the pathways, of radicalization. Radicalization is not a new topic (1). It is considered in connection with social and political influences as well as an individual process (2–3). In addition to the term radicalization, the term radicalisation, in British spelling, is used in several articles particularly from the UK's point of view.

The above examples have illustrated that one can develop an understanding of a scientific domain at multiple levels of granularity with a relatively low threshold of prior knowledge of the domain. The strategic exploration is essentially a top-down approach in the same spirit as Shneiderman's visual information search mantra advocates: overview first, zoom and filter, then details on demand. The key here is that scholarly significant patterns can be represented as prominent visual cues. Once we learn what the most common and critical patterns may look like, it usually takes little domain knowledge to recognize visually salient cues, which will lead us to the valuable information that we should concentrate on.

With little adjustments, we can transform the same methodology into a viable analytic approach to the analysis of a scientific domain at a finer level of granularity —semantic predications. Scientometric studies typically focus on structural and

dynamic patterns at the level of article or higher levels of granularity such as journals and groups of journals. Analyzing a scientific domain in terms of its semantic predications and their evolving patterns across disciplinary boundaries enables us to address research questions directly.

Main Path Analysis

Representing the scientific literature of a knowledge domain as a network lends us many analytic tools and methods to identify valuable patterns and trends. The main path analysis is a method that can simplify a usually complex network to a small number of paths that would characterize the major development of the underlying domain. Early studies of main paths include (Hummon and Doreian 1989; Carley et al. 1993; Batagelj 2003; Lucio-Arias and Leydesdorff 2008). More recent examples of main path studies include (Liu and Lu 2012; Liu and Kuan 2015). Here we illustrate how to perform a main path analysis of the terrorism research using a combination of CiteSpace and Pajek. Pajek is a computer program for processing large-scale networks, including visualizing a network and analyzing a network with a wide variety of algorithms. Pajek is freely available and it is very powerful. It can handle a network with millions of nodes (Batagelj and Mrvar 1998).

The procedure starts with the bibliographic data downloaded from the Web of Science. The next step is to generate a directed citation network from the bibliographic dataset. Each node in a directed citation network is an article. The article can cite other articles in the network. The article itself can be cited by other articles in the network as well. CiteSpace provides a function that takes the Web of Science records as the input and generates a directed citation network in the Pajek's .net format. Then we can use Pajek to generate main paths, which are a sub-network. First, open the directed citation network in Pajek (Fig. 3.29).

Next, retain the largest connected component of the network so that main paths can be selected from the largest connected component. This requires two steps in Pajek. First, identify the weakly connected components (Fig. 3.30). Then, select the largest one to retain (Fig. 3.31). A strongly connected component in a directed graph is defined as a sub-graph in which every node is reachable from every other node. A weakly connected component is similarly defined, except that we will have to ignore the directed links and consider them as undirected. For example, if we are at the end of a one-way street, we can reach the beginning of the one-way street only if we ignore the one-way restriction.

Pajek identified 243,964 components from the network of 330,662 article nodes. Most of them are singleton components that contain one article only. The largest connected component, cluster 2, contains 84,770 nodes, representing about 25% of the original loosely connected network (Fig. 3.31). Cluster 2 is the largest connected component to retain (Fig. 3.32).
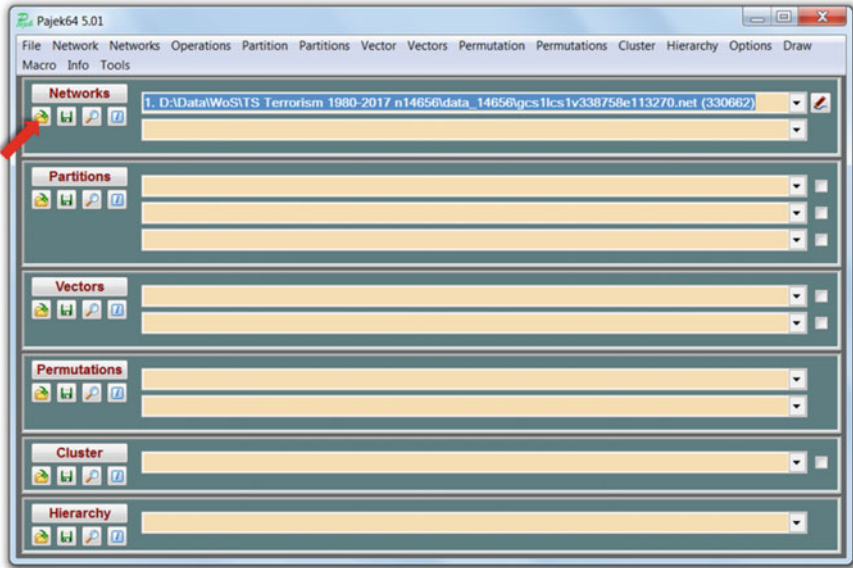
Fig. 3.29 Open the directed citation network in Pajek
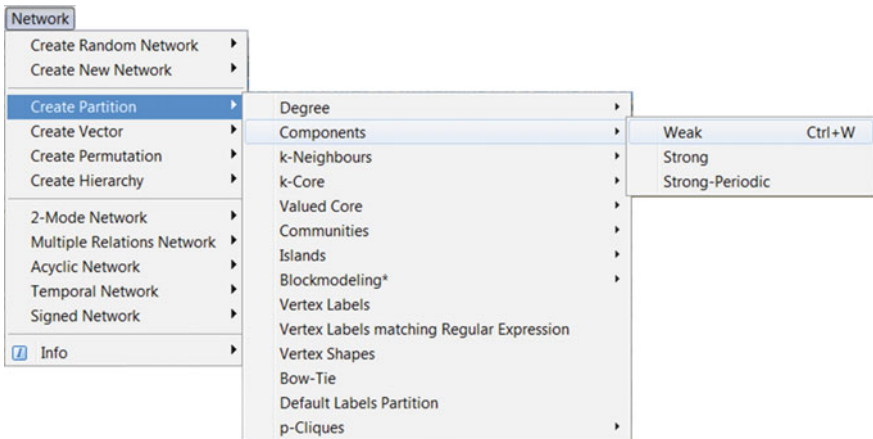
Fig. 3.30 Retain the largest connected component of the network

Main path analysis requires the target network is a directed acyclic network. If the largest connect component contains strongly connected components, such strongly connected components would violate the acyclic condition. If a few articles cite each other, they can be considered as a group. Thus the next step is to shrink strongly connected components as shown in Fig. 3.33.
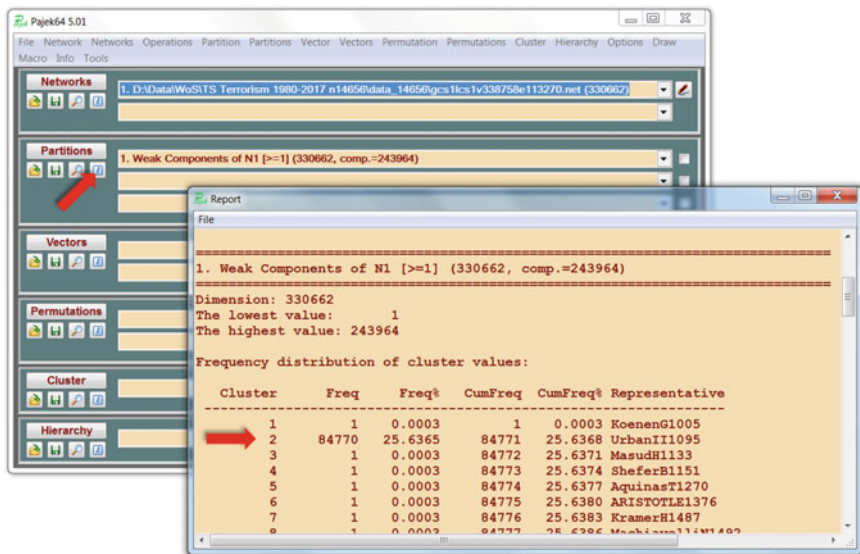
Fig. 3.31 Identify the largest connected component to retain
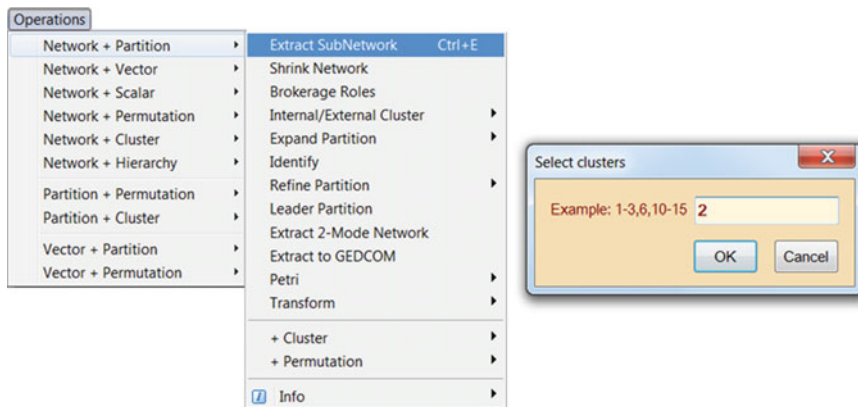
Fig. 3.32 Extract the largest connected component by selecting cluster 2 to retain

After removing loops from the remaining network (Fig. 3.34), we have obtained an acyclic network. There are several ways to compute traversal weights in order to identify main paths, including Search Path Count (SPC), Search Path Link Count (SPLC), and Search Path Node Pair (SPNP) (Fig. 3.35). The next step is to extract the main paths as a subgraph. The user has several options too. We illustrate an option called Key-Route, which presents a combination of a number of significant paths identified (Fig. 3.36).
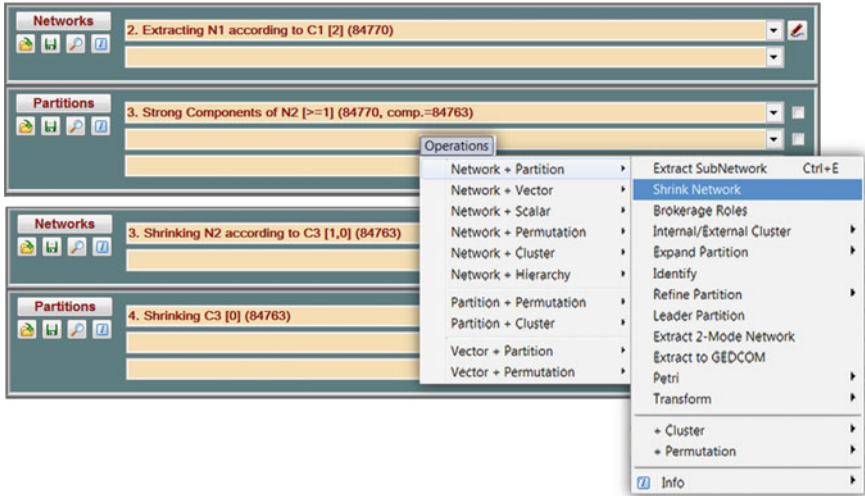
Fig. 3.33 Shrink strongly connected components
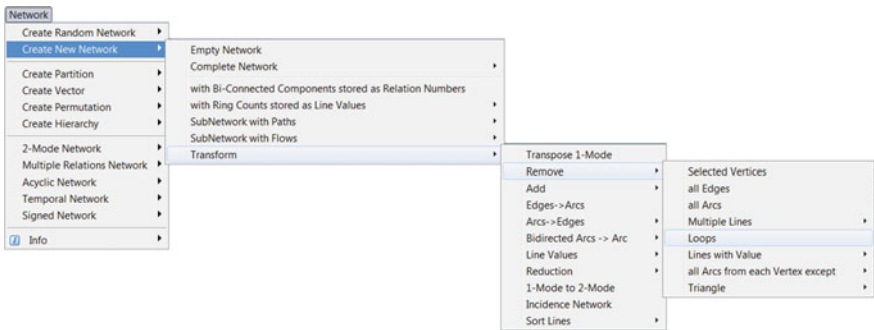
Fig. 3.34 Remove loops from the largest connected component
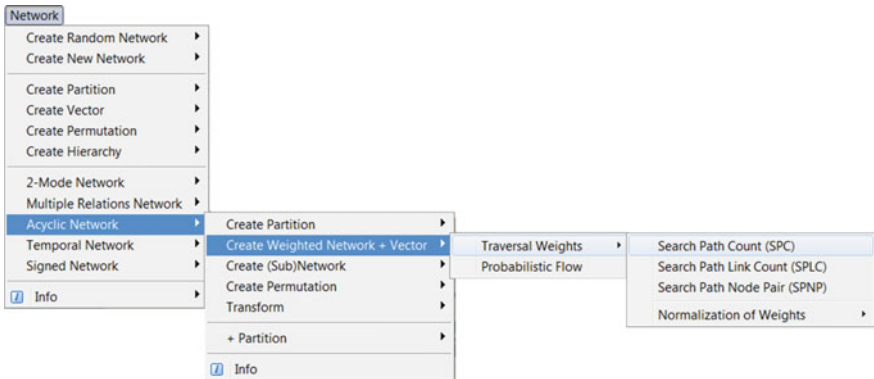
Fig. 3.35 Compute traversal weights along main paths
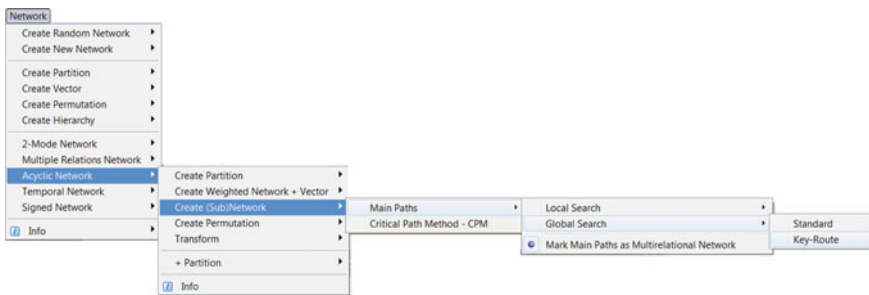
Fig. 3.36 Create the main paths by including multiple key routes, e.g. 1–10

Fig. 3.37 Use the LAYERS.MCR macro to draw the generated main paths

Finally, the extracted main paths can be displayed by using a macro from Pajek called LAYERS.MCR (Fig. 3.37). The resultant display of the main paths is shown in Fig. 3.38. The user can refine the display further, for example, by applying a community detection algorithm and color the nodes accordingly.

The main paths shown in Fig. 3.38 are arranged such that articles located at the top of the diagram are published earlier than articles below them. In other words, the top of the diagram is where the oldest publications are located. Therefore, the group of nodes in yellow at the top would be considered as the pioneers of the terrorism research. They are all cited by one node below them, GaleaS2002. We use the first author's last name, the initial, and the year of publication to identify the node. If the node is a citing article in the dataset, we will include a keyword from the article. In this case, the keyword posttraumatic stress disorder is selected from the article GaleaS2002. Let us trace the main paths by following the vertical lines that connect two nodes: the node at the higher end of the line is cited by the node at the lower end of the line.
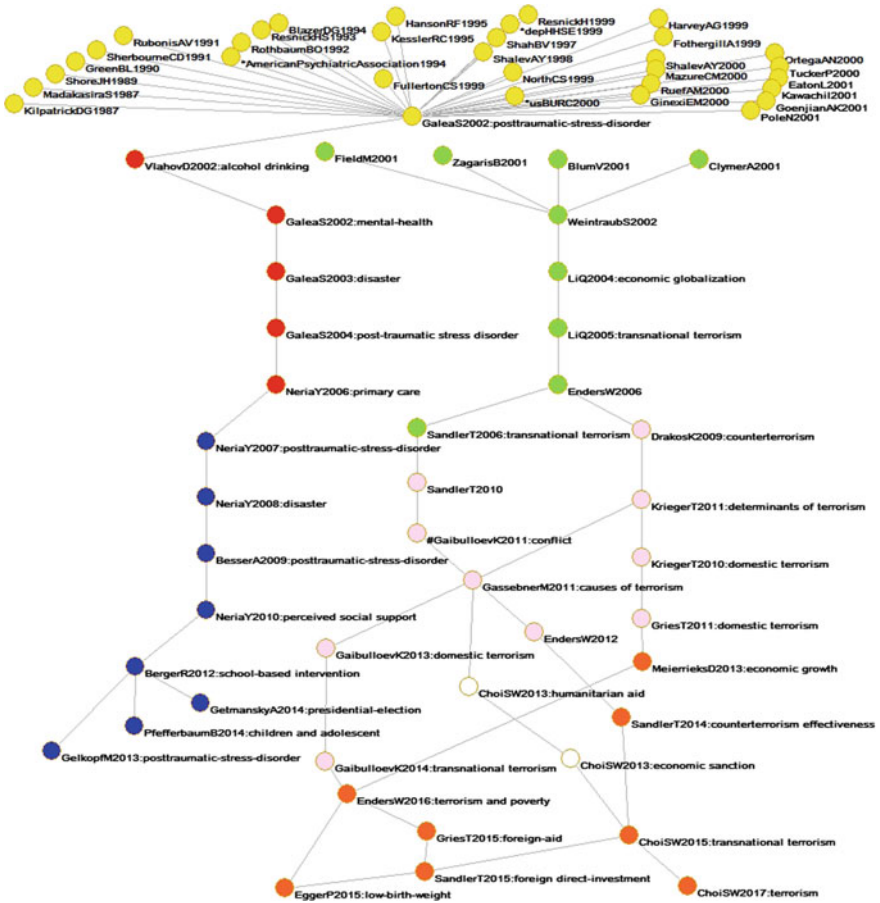
Fig. 3.38 Main paths derived from the direct citation network based on search path count (SPC)

GaleaS2002 leads to a chain of nodes in red, which means that these articles cited GaleaS2002 on PTSD. The red group is identified by a community detection algorithm. The first node of the group is VlahovD2002: alcohol drinking, followed by three more articles that appear to be authored by Galea between 2002 and 2004 on mental health, disaster, and PTSD. These articles are likely to have multiple authors, but in part due to the limitation of the data from the Web of Science, the complete authorship is not readily available. Besides our focus here is mainly on the course of evolution and less so on the authorship per se. The large node in the red group is by Neria Y. published in 2006 on primary care. As suggested by the common keywords, the red group is a line of research on mental health, especially on PTSD in relation to the September 11 terrorist attacks. Neria's 2006 article leads to a chain of nodes in blue published between 2007 and 2010. PTSD remains to be the most prominent keyword for this segment of the main path. The blue route ends with three articles that all cited BergerR2012 on school-based intervention.

The second main path starts with a group of nodes in green on the right-hand side of the diagram. More specifically, this main path begins with four articles published in 2001 followed by WeintraubS2002, which converged the multiple threads to a single-line path. The green segment of the path is characterized by keywords such as economic globalization and transnational terrorism. Then the path is split into two routes, which remained separated between 2006 and 2011. One route contains keywords such as counterterrorism, determinants of terrorism, and domestic terrorism. The other contains keywords such as transnational terrorism. However, these routes became increasingly interwoven into each other through articles such as GassebnerM2011 on causes of terrorism, EndersW2016 on terrorism and poverty, and ChoiSW2015 on transnational terrorism.

The main path analysis of the terrorism research (1980–2017) depicts a big picture that echoes what we have seen in the timeline views of co-cited references. The main path on PTSD corresponds to the largest cluster in the co-citation network, which is also clear that this line of research is no longer as active as before. The second and more complex main path on domestic and transnational terrorism corresponds to the two currently active clusters of co-cited references, which share a higher-level goal to identify the causes of terrorism from economic, social, political, and other dimensions. The questions to be answered are fundamentally significance because eliminating the environment that breeds terrorism would be more effective in a long run than focusing on dealing with aftermaths of terrorism alone.

Structural Variations

The structural variations of the structure of a domain's knowledge can be detected in terms of the changes of the modularity metric of the networks over time (Fig. 3.39). For the terrorism research (1980–2017), the domain has the lowest modularity in 2002. One interpretation would be that the overall connectivity of the network was the strongest in 2002 and a profound common theme made it hard to divide the network in 2002 into clearly separated parts. The predominant position of the PTSD specialty after the September 11 terrorist attacks was evident in both studies of the domain in 1996–2003 and 1980–2017. The September 11 terrorist attacks are likely to be the reason behind the low modularity.

The structural variation analysis of the terrorism research (1980–2017) reveals an interesting pattern: transformative links are all associated with the PTSD research (cluster #0). Since this once predominant cluster stopped its growth before 2010, the structural variation analysis did not find transformative connections in the two currently active lines of research on domestic terrorism and transnational terrorism. The lack of transformative links after the PTSD research could be explained as a sign of a period of normal science as in Kuhn's paradigmatic theory or Shneider's evolution theory. Researchers in these specialties have a clearly established conceptual framework to work with. It is less likely to observe transformative links during this period of time.
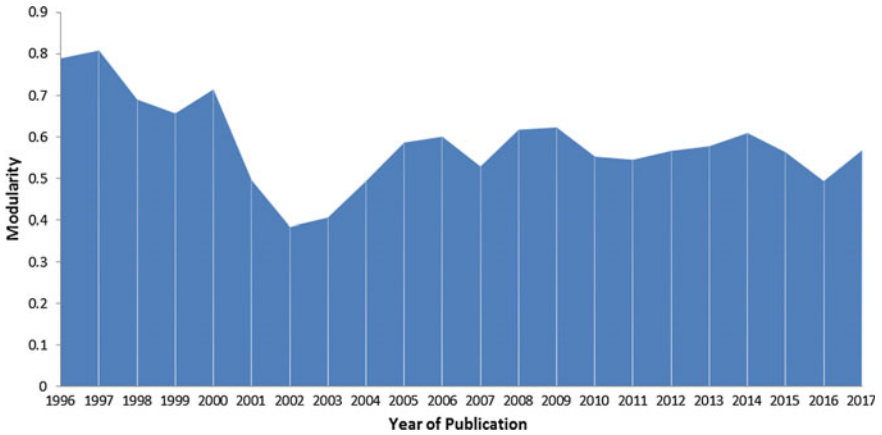
Fig. 3.39 The modularity of the baseline network changes over the years
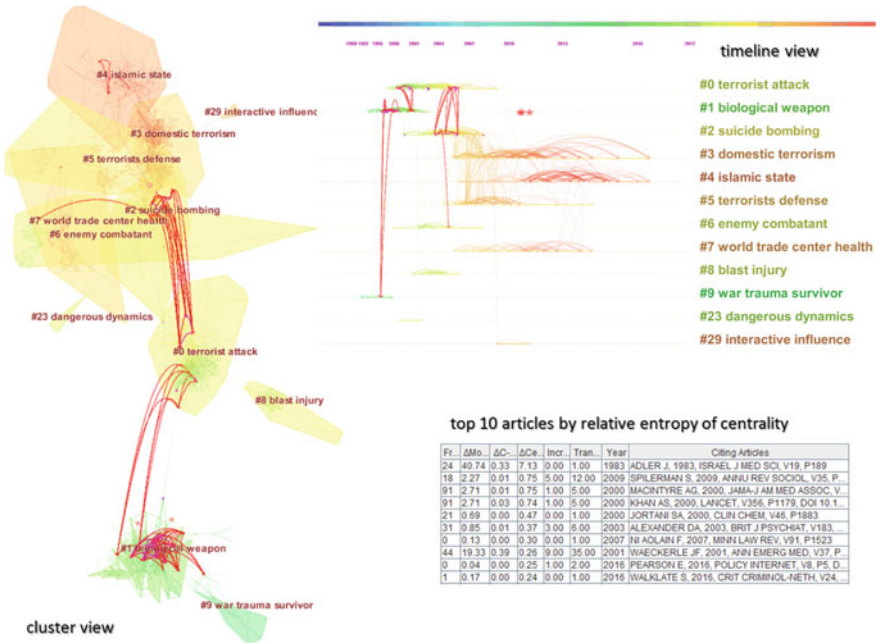
Fig. 3.40 Structural variations measured by the relative entropy of the distributions of betweenness centrality

Structural variations measured by the distribution of betweenness centrality reveal different patterns (Fig. 3.40). Earlier cross-cluster links were added between Cluster #1 biological weapon and Cluster #9 war trauma survivor and between Cluster #0 and Cluster #1. Articles identified with transformative potentials made

connections between Cluster #0 terrorist attack and Cluster #2 suicide bombing. The latter cluster was relatively recent. The timeline view shows that Cluster #2 has relatively strong connections with the two current clusters #3 and #4. None of the structural variation connections were added after 2007.

Science Mapping

Science mapping is a generic process of domain analysis and visualization. A science mapping study typically consists of several components, notably a body of scientific literature, a set of scientometric and visual analytic tools, metrics and indicators that can highlight potentially significant patterns and trends, and theories of scientific change that can guide the exploration and interpretation of visualized intellectual structures and dynamic patterns.

The Interplay Between Science and Theories of Science

Science mapping approaches typically aim to represent patterns and trends of the development of science at macroscopic levels such as disciplines and fields over a long period of time. Indeed, science mapping is a very promising combination of the domain analysis method originated in information science and visual analytics from computer science. In particular, science mapping provides a unique means to verifying the validity of individual theories of scientific change. In return, each of the macroscopic theories such as Kuhn's paradigm shifts, Fuchs' mutual dependencies and competition, Shneider's four-stage evolution provides a rich and yet potentially biased perspective that may guide us to interpret how a scientific field is unfolding in front of us.

Commonly used sources of scientific literature include the Web of Science, Scopus, Google Scholar, and PubMed. Scientometric methods include author co-citation analysis (ACA) (White and McCain 1998; Chen 1999), document co-citation analysis (DCA) (Small 1973; Chen 2006), co-word analysis (Callon et al. 1983), and many other variations. Visualization techniques include graph or network visualization (Herman et al. 2000), visualizations of hierarchies or trees (Johnson and Shneiderman 1991), visualizations of temporal structures (Morris et al. 2003), geospatial visualizations, and coordinated views of multiple types of visualizations. Metrics and indicators of research impact include citation counts (Garfield 1955), the h-index (Hirsch 2005) and its numerous extensions, and a rich set of altmetrics on social media (Thelwall et al. 2013).

Theories of scientific change include the paradigmatic views of scientific revolutions, scientific advances driven by competitions, and evolutionary stages of a scientific discipline. In order to conduct a science mapping study, researchers need to develop a good understanding of each of the categories of skills and knowledge

outlined above. Furthermore, each of these categories is a current and active research area in its own right, for instance, the current research on finding the optimal field normalization method and the debates over how various potentially conflicting theories of scientific change may be utilized to reveal the underlying mechanisms of how science advances.

The complexity of science mapping is shared by many research fields. We will illustrate the process of a systematic review based on a series of visual analytic functions implemented in CiteSpace (Chen 2004, 2006; Chen et al. 2010). We demonstrate the steps of preparing a representative dataset, how to generate visualizations that can guide our review, and how to identify salient patterns at various levels of granularity.

Characterizing the Field of Study

The dataset to represent the research field is collected through multiple topic search queries to the Web of Science. The rationale of the query construction is as follows. First, we would like to ensure that currently widely used science mapping tools such as VOSViewer, CiteSpace, HistCite, SciMAT, and Sci2 are covered by our topic search query. The inclusion of software tools is based on the characterization of Shneider's second evolutionary stage. Thus, publications that mention any of these software tools in their titles, abstracts, and/or keyword lists will be included. This query generates 135 records as Set #1 (Fig. 3.41).

Second, since the goal of science mapping is to identify the intellectual structure of a scientific domain, the second query focuses on the object of science mapping, including topic terms such as intellectual structure, scientific change, research front, invisible college, and domain analysis. This query is motivated by the first evolutionary stage in Shneider's evolution model. The query may also capture major paradigms because these concepts are fundamental to the research. As we will see later on, terms such as domain analysis may be ambiguous as they are also used in other contexts that are irrelevant to science mapping. In practice, one should defer the assessment of relevance until the analysis stage. This query produces 13,242 records as Set #2.

The third query focuses on scientometric and visual analytic techniques that are potentially relevant to science mapping. Topic terms include science mapping, knowledge domain visualization, information visualization, citation analysis, co-citation analysis. Some of these techniques are enabling techniques developed elsewhere in fields such as computer science. This query would capture the development and application of these techniques. This query leads to 4772 records.

The queries #4–#10 aim to retrieve bibliographic records on the common data sources for science mapping, including Scopus (6782 records), the Web of Science (15,401 records), Google Scholar (5170 records), Pubmed (46,760 records), and MEDLINE (61,405 records).

Set	Results	
		Save History / Create Alert Open Saved History
# 15	350	#13 AND #2 *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 14	17,731	#13 OR #2 *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 13	4,839	#3 OR #1 *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 12	111	(TS=("literature-based discovery")) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 11	23,275	#8 OR #7 OR #6 OR #5 OR #4 *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 10	61,405	(TS=Medline) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 9	46,760	(TS=Pubmed) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 8	354	(TS=("PubMed Central")) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 7	24	(TS=("Microsoft Academic Search")) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 6	5,170	(TS=("Google Scholar")) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 5	15,401	(TS=("web of science" OR "web of knowledge" OR "science citation index")) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 4	6,782	(TS=(Scopus)) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 3	4,772	(TS=("science mapping" OR "science mapping analysis" OR "scientometric study" OR "scientometric analysis" OR "scientometric review" OR "knowledge domain visualization" OR "knowledge visualization" OR "bibliometric mapping" OR "bibliographic network visualization" OR "historiographic analysis" OR "research impact assessment" OR "information visualization" OR "domain visualization" OR "network visualization" OR "citation analysis" OR "co-citation analysis" OR "cocitation analysis" OR "co-word analysis" OR "coword analysis" OR "author co-citation" OR "document co-citation" OR "journal co-citation")) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 2	13,242	(TS=("intellectual structure*" OR "scientific change" OR "scientific revolutions" OR "scientific progress" OR "research trend" OR "conceptual evolution" OR "domain analysis" OR "research front" OR "knowledge diffusion" OR "research programme*" OR "scientific programme*" OR "invisible college" OR "scientific frontiers" OR "literature dynamics")) *AND* **LANGUAGE:** (English) *AND* **DOCUMENT TYPES:** (Article OR Review) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*
# 1	135	TS=(VOSViewer OR CiteSpace OR HistCite OR SciMAT OR Sci2) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*

Fig. 3.41 Topic search queries used for data collection

The final dataset is Set #14, containing 17,731 bibliographic records of the types of Article or Review in English (Fig. 3.42). This query formation strategy is generic enough to be applicable to a science mapping study unless of course one has access to the entire database.

Patents and research grants are other types of data sources one may consider, but for this particular review, we are limited to the scientific literature indexed by the Web of Science.

Visual Analysis of the Literature

We visualize and analyze the dataset with CiteSpace. CiteSpace takes a set of bibliographic records as its input and models the intellectual structure of the underlying domain in terms of a synthesized network based on a time series of networks derived from each year's publications. CiteSpace has been continuously developed for more than a decade. CiteSpace supports several types of bibliometric studies, including collaboration network analysis, co-word analysis, author co-citation analysis, document co-citation analysis, text and geospatial visualizations. In this case, we focus on the document co-citation analysis within the period of time between 1995 and 2016 (Fig. 3.43).
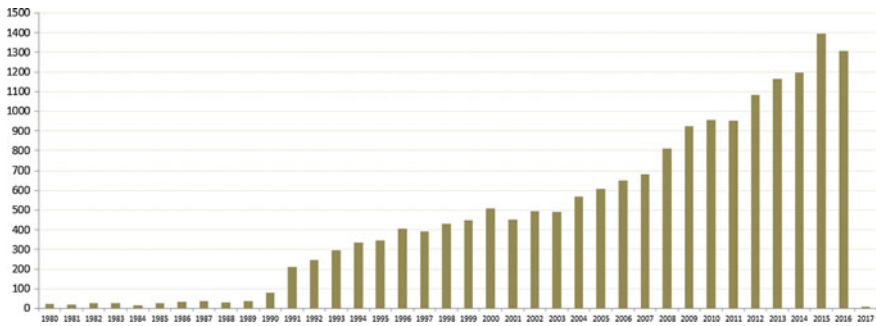
Fig. 3.42 The distribution of the bibliographic records in Set #14
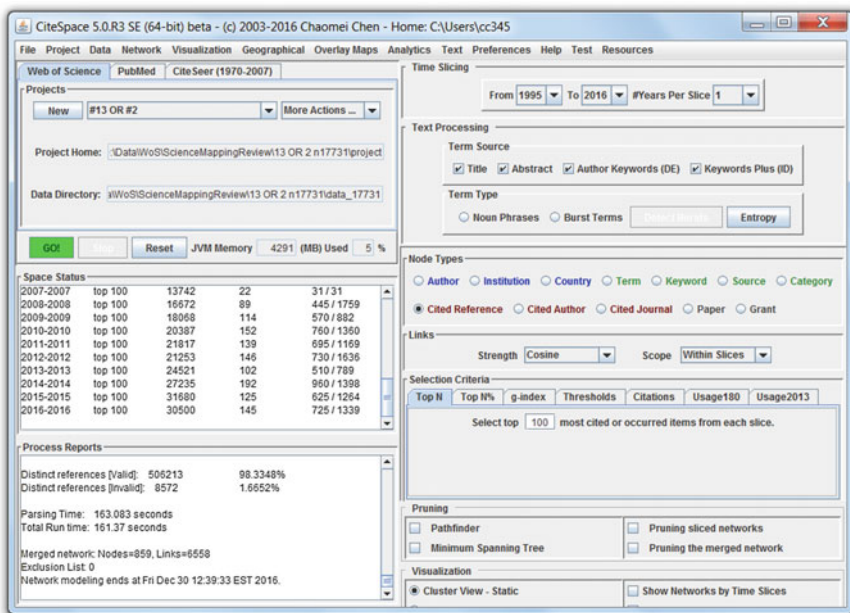
Fig. 3.43 The main user interface of CiteSpace

The Set #14 contains 16,250 records published in the range of 1980–2017. These records collectively cited 515,026 references. The document co-citation analysis function in CiteSpace constructs networks of cited references. Connections between references represent co-citation strengths. CiteSpace uses a time slicing technique to build a time series of network models over time and synthesize these individual networks to form an overview network for the systematic review of the relevant literature.

The synthesized network is divided into co-citation clusters of references. Citers to these references are considered as the research fronts associated with these clusters. Each cluster represents the intellectual base of the underlying specialty. According to Shneider's four stage model, the intellectual base of a specialty and the corresponding research fronts provide valuable insights into the current stage of the specialty as well as the intellectual milestones in the evolution of the specialty.

Our first step in the review is to make sense of the nature of major clusters and characteristics that may inform us about the stage of the underlying specialties. In this study, we consider a cluster as the embodiment of an underlying specialty. Thus, science mapping consists of multiple specialties that contribute to various aspects of the domain.

In each cluster, we focus on cluster members that are identified by structural and temporal metrics of research impact and evolutionary significance. A commonly used structural metric is the betweenness centrality of a node in a network. Studies have shown that nodes with high betweenness centrality values tend to identify boundary spanning potentials that may lead to transformative discoveries (Chen et al. 2009). Burst detection is a computational technique that has been used to identify abrupt changes of events and other types of information (Kleinberg 2002). In CiteSpace, the sigma score of a node is a composite metric of the betweenness centrality and the citation burstness of the node, i.e. the cited reference. CiteSpace represents the strength of these metrics through the design of visual encoding such that articles that are salient in terms of these metrics will be easy to see in the visualizations. For example, the citation history of a node is depicted as a number of tree rings and each tree ring represents the number of citations received in the corresponding year of publication. If a citation burst is detected for a cited reference, the corresponding tree ring will be colored in red. Otherwise, tree rings will be colored by a spectrum that ranges from cold colors such as blue to warm colors such as orange.

The nature of a cluster is identified from the following aspects: a hierarchy of key terms in articles that cite the cluster (Tibély et al. 2013), the prominent members of the cluster as the intellectual milestones in its evolution and as the intellectual base of the specialty, recurring themes in the citing articles to the cluster to reflect the interrelationship between the intellectual base and the research fronts. In particular, we will pay attention to indicators of the evolutionary stages of a specialty such as the original conceptualization, research instruments, applications, and routinization of the domain knowledge of the specialty.

In addition to the study of citation-based patterns, we will demonstrate the concept of citation trajectories in the context of distinct clusters. According to the theory of structural variation, the transformative potential of an article may be reflected by the extent to which it varies the existing intellectual structure (Chen 2012). For example, if an article adds many inter-cluster links, it may alter the overall structure. If the structural change is subsequently accepted and reinforced by other researchers, then transformative changes of the knowledge become significant in a socio-cognitive view of the domain.

Visualizing the Field

A dual-map overlay of the science mapping literature represents the entire dataset in the context of a global map of science generated from over 10,000 journals indexed in the Web of Science (Chen and Leydesdorff 2014). The dual-map overlay in Fig. 3.44 shows that science mapping papers are published in almost all major disciplines. Publications in the discipline of information science (shown in the map as curves in cyan) are built on top of at least four disciplines on the right-hand side of the map.

A hierarchical visualization of index terms, i.e. keywords, is generated to represent the coverage of the dataset (Fig. 3.45). Five semantic types of nodes are annotated in the visualized hierarchy:

What: a fundamental phenomenon of a specialty and the object of a study, for example, the intellectual structure or the dynamics of a research field.
How: methodologies, procedures, and processes of science mapping, for example, author co-citation analysis, bibliometric mapping, and co-citation analysis.
Abstraction: computational models of an underlying phenomenon identified from the bibliographic data, representations such as Pathfinder networks, metrics and indicators such as the h-index and the g-index.
Tools: computational techniques, algorithms and software tools for visualization and ranking scholarly publications.
Data: data sources used by science mapping studies, for example, Scopus and Google Scholar.

These semantic types will be also used to identify the evolutionary stage of a specialty. For example, if a cluster contains several articles that report the development of software tools, then the underlying specialty is considered as a specialty that has reached at least Stage II. If the methodologies appear in a cluster of knowledge domains external to information science, such as regenerative medicine

Fig. 3.44 A dual-map overlay of the science mapping literature
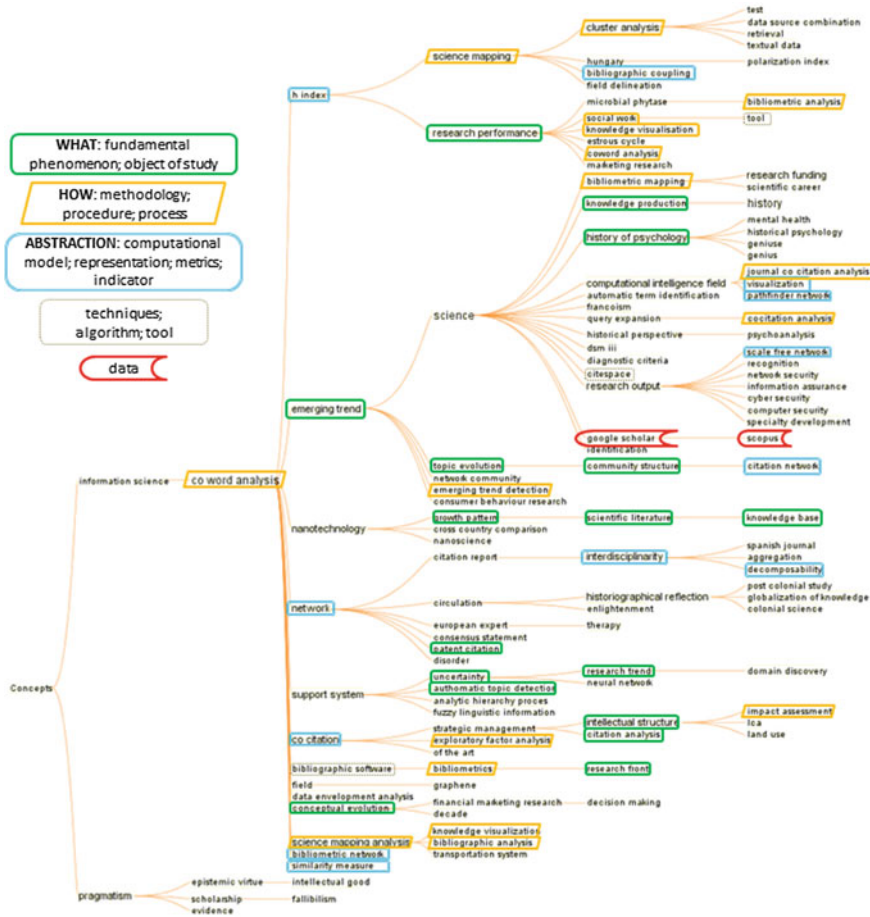
Fig. 3.45 A hierarchy of indexing terms derived from Set #14

and strategic management research, then we will consider the specialty has reached Stage III—tools developed by the specialty are applied to other subject domains. In the following analysis, we will use the terms in the hierarchy as the primary source of our vocabulary to identify the role of the contributions made by a scientific publication to a specialty.

Major milestones in the development of science mapping can be identified from the list of references that have strong citation bursts between 1995 and 2016 (Fig. 3.46). References with strong values in the Strength column tend to be significant milestones for the science mapping research. We label such references with high-level concepts. For example, the first milestone paper in the study is a landmark ACA study of information science (White and McCain 1998). The next milestone is a major collection of seminal papers in information visualization by

	References	Year	Strength	Begin	End	1995 - 2016
	MALIK M, 1992, J AM COLL CARDIOL, V20, P127	1992	10.4711	1995	1999	
	MANCINI DM, 1993, CIRCULATION, V87, P1083	1993	5.8964	1996	2000	
	CAMM AJ, 1996, CIRCULATION, V93, P1043	1996	46.4658	1997	2004	
	TAFLOVE A, 1995, COMPUTATIONAL ELECTR, V, P	1995	23.0034	1997	2002	
	CAMM AJ, 1996, EUR HEART J, V17, P354	1996	5.4362	1998	2004	
	MACROBERTS MH, 1996, SCIENTOMETRICS, V36, P435, DOI	1996	7.9414	1999	2004	
Author Co-Citation Analysis	WHITE HD, 1998, J AM SOC INFORM SCI, V49, P327, DOI	1998	40.1581	1999	2006	
	WHITE HD, 1997, ANNU REV INFORM SCI, V32, P99	1997	10.5983	1999	2003	
Information Visualization	CARD SK, 1999, READINGS INFORMATION, V, P	1999	25.7498	2000	2006	
	MAY RM, 1997, SCIENCE, V275, P793, DOI	1997	4.9544	2000	2004	
	BALDI S, 1998, AM SOCIOL REV, V63, P829, DOI	1998	4.8083	2001	2005	
	ALTSCHUL SF, 1997, NUCLEIC ACIDS RES, V25, P3389, DOI	1997	7.217	2001	2005	
	GOODRUM AA, 2001, INFORM PROCESS MANAG, V37, P661, DOI	2001	7.2483	2002	2009	
	CRONIN B, 2001, J INFORM SCI, V27, P1	2001	12.5667	2002	2008	
Citation Mapping	SMALL H, 1999, J AM SOC INFORM SCI, V50, P799, DOI	1999	15.8489	2002	2006	
Domain Analysis	HJORLAND B, 2002, J DOC, V58, P422, DOI	2002	11.7233	2003	2010	
	BORGMAN CL, 2002, ANNU REV INFORM SCI, V36, P3	2002	13.2442	2003	2010	
Graph Visualization	HERMAN I, 2000, IEEE T VIS COMPUT GR, V6, P24, DOI	2000	15.7512	2003	2008	
	TAFLOVE A, 2000, COMPUTATIONAL ELECTR, V, P	2000	11.9635	2004	2009	
	ADAM D, 2002, NATURE, V415, P726	2002	8.4895	2004	2009	
	WHITE HD, 2003, J AM SOC INF SCI TEC, V54, P423, DOI	2003	10.5968	2004	2010	
Domain Visualization	BORNER K, 2003, ANNU REV INFORM SCI, V37, P179, DOI	2003	15.4015	2004	2011	
	CHEN CM, 2002, J AM SOC INF SCI TEC, V53, P678, DOI	2002	4.5481	2004	2008	
	AKSNES DW, 2003, SCIENTOMETRICS, V56, P235, DOI	2003	3.3291	2005	2010	
	KING DA, 2004, NATURE, V430, P311, DOI	2004	6.6773	2005	2010	
	SAHA S, 2003, J MED LIBR ASSOC, V91, P42	2003	5.414	2005	2010	
	TUFTE ER, 2001, VISUAL DISPLAY QUANT, V, P	2001	4.598	2005	2009	
	KEIM DA, 2002, IEEE T VIS COMPUT GR, V8, P1, DOI	2002	6.2489	2005	2010	
Mapping Scientific Frontiers	CHEN C, 2003, MAPPING SCI FRONTIER, V, P	2003	6.6665	2005	2010	
	GLANZEL W, 2002, SCIENTOMETRICS, V53, P171, DOI	2002	7.5022	2005	2010	
	AHLGREN P, 2003, J AM SOC INF SCI TEC, V54, P550, DOI	2003	16.1924	2006	2011	
UCINET	BORGATTI S P, 2002, UCINET WINDOWS SOFTW, V, P	2002	9.3193	2006	2010	
Visual Analytics	THOMAS J J, 2005, ILLUMINATING PATH RE, V, P	2005	9.9577	2006	2011	
	GMUR M, 2003, SCIENTOMETRICS, V57, P27, DOI	2003	6.6503	2006	2010	
	RAMOS-RODRIGUEZ AR, 2004, STRATEGIC MANAGE J, V25, P981, DOI	2004	12.7007	2006	2012	
	WHITE HD, 2003, J AM SOC INF SCI TEC, V54, P1250, DOI	2003	9.6883	2006	2011	
	ALAHBABI MN, 2005, J OPT SOC AM B, V22, P1321, DOI	2005	4.291	2006	2012	
Cytoscape	SHANNON P, 2003, GENOME RES, V13, P2498, DOI	2003	16.4941	2007	2011	
h-index	HIRSCH JE, 2005, P NATL ACAD SCI USA, V102, P16569, DOI	2005	40.6435	2007	2013	
	MOED H F, 2005, CITATION ANAL RES EV, V, P	2005	17.9615	2007	2012	
	MEHO LI, 2007, J AM SOC INF SCI TEC, V58, P2105, DOI	2007	3.315	2008	2013	
CiteSpace I	CHEN CM, 2004, P NATL ACAD SCI USA, V101, P5303, DOI	2004	6.8456	2008	2012	
	BOYACK KW, 2005, SCIENTOMETRICS, V64, P351, DOI	2005	11.8903	2009	2013	
CiteSpace II	CHEN CM, 2006, J AM SOC INF SCI TEC, V57, P359, DOI	2006	6.7057	2009	2014	
	SINGH J, 2005, MANAGE SCI, V51, P756, DOI	2005	4.6233	2009	2013	
	GARFIELD E, 2006, JAMA-J AM MED ASSOC, V295, P90, DOI	2006	7.1782	2009	2013	
	CHIU WT, 2007, SCIENTOMETRICS, V73, P3, DOI	2007	4.6127	2010	2016	
	WUCHTY S, 2007, SCIENCE, V316, P1036, DOI	2007	7.1104	2011	2016	
	XIE SD, 2008, SCIENTOMETRICS, V77, P113, DOI	2008	4.1387	2011	2016	

Fig. 3.46 49 references with citation bursts of at least 5 years

Card et al. (1999). Other major milestones include visual analytics (Thomas and Cook 2005), and the h-index (Hirsch 2005).

Landscape View

The landscape view in Fig. 3.47 is generated based on publications between 1995 and 2016. Top 100 most cited publications in each year are used to construct a network of references cited in that year. Then individual networks are synthesized. The synthesized network contains 3145 references. The network contains 603
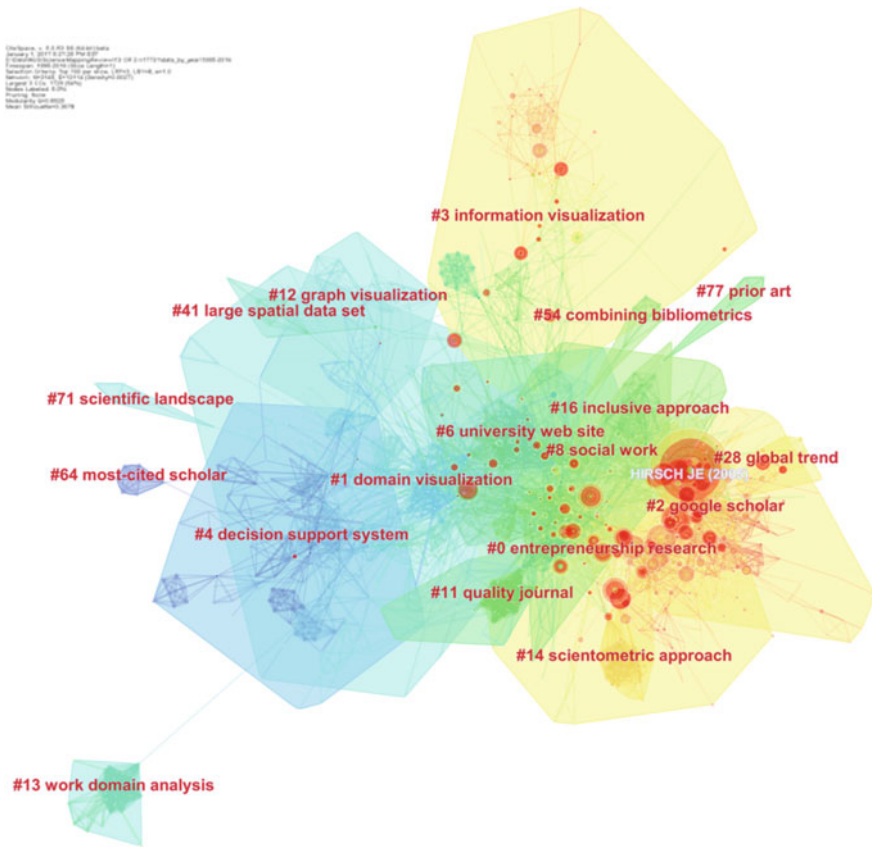
Fig. 3.47 A landscape view of the co-citation network, generated by top 100 per slice between 1995 and 2016 (LRF = 3, LBY = 8, and e = 1.0)

co-citation clusters. The three largest connected components include 1729 nodes, which account for 54% of the entire network. The network has a modularity of 0.8925, which is considered as very high, suggesting that the specialties in science mapping are clearly defined in terms of co-citation clusters. The average silhouette score of 0.3678 is relative low mainly because of the numerous small clusters. The major clusters that we will focus on in the review are sufficiently high.

The areas of different colors indicate the time when co-citation links in those areas appeared for the first time. Areas in blue were generated earlier than areas in green. Areas in yellow were generated after the green areas and so on. Each cluster can be labeled by title terms, keywords, and abstract terms of citing articles to the cluster. For example, the yellow-colored area at the upper right quadrant is labeled as #3 information visualization, indicating that Cluster #3 is cited by articles on information visualization. The largest node is the paper that introduces the h-index. Other nodes with tree rings in red are references with citation bursts.

Timeline View

A timeline visualization in CiteSpace depicts clusters along horizontal timelines (Fig. 3.48). Each cluster is displayed from left to right. The legend of the publication time is shown on top of the view. The clusters are arranged vertically in the descending order of their size. The largest cluster is shown at the top of the view. The colored curves represent co-citation links added in the year of the corresponding color. Large-sized nodes or nodes with red tree rings are of particular interest because they are either highly cited or have citation bursts or both. Below each timeline the three most cited references in a particular year are displayed. The label of the most cited reference is placed at the lowest position. References published in the same year are placed so that the less cited references are shifted to the left. The new version of CiteSpace supports the function to generate labels of a cluster year by year based on terms identified by Latent Semantic Indexing (LSI) (Deerwester et al. 1990). The year-by-year labels can be displayed in a table or above the corresponding timeline. Users may control the displays interactively.

Clusters are numbered from 0, i.e. Cluster #0 is the largest cluster and Cluster #1 is the second largest one. As shown in the timeline overview, the sustainability of a specialty varies. Some clusters sustain a period over 20 years, whereas some clusters are relatively short-lived. Some clusters remain active until the 2015, the most recent year of publication for a cited reference in this study.
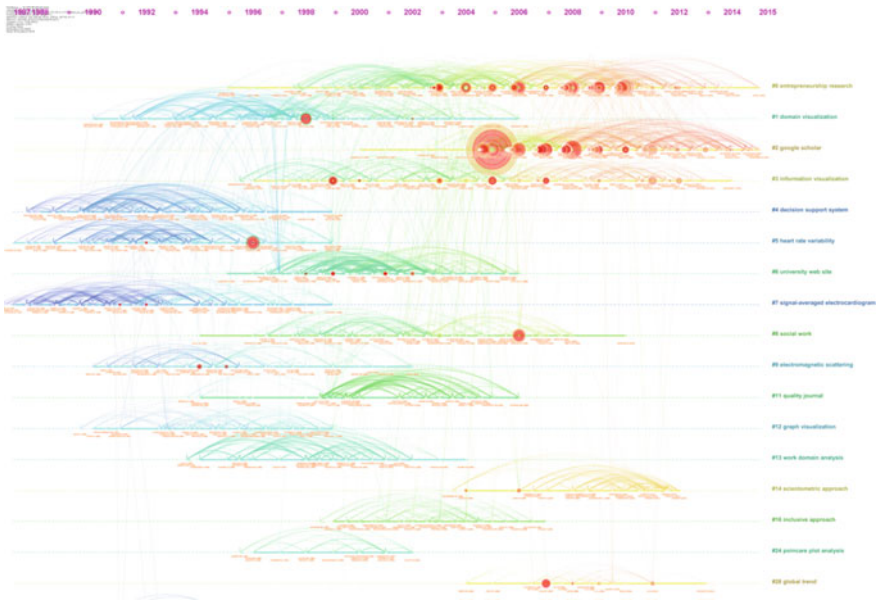
Fig. 3.48 A timeline visualization of the largest clusters of the total of 603 clusters

Table 3.9 The five largest clusters of co-cited references of the network of 3145 references

Cluster	Size	Mean (Year)	Silhouette	% of the network	Accumulated % of network	% of top 3 LCCs	Accumulated % of LCCs
0	214	2006	0.748	4.5	4.5	8.1	8.1
1	209	1997	0.765	2.3	6.7	4.1	12.2
2	190	2009	0.845	3.3	10.0	6.0	18.2
3	160	2005	0.954	2.9	12.9	5.3	23.5
4	152	1992	0.890	1.7	14.6	3.0	26.5

The largest three connected components include 1729 of the references

As shown in Table 3.9, each of the largest five clusters has over 150 members. The largest cluster's homogeneity in terms of the silhouette score is slightly lower than that of the smaller clusters. The largest cluster represents 4.5% of the references from the entire network and 8.1% of the largest three connected components of the network (LCCs). In this study, our review will primarily focus on the largest five clusters.

The duration of a cluster is particularly interesting (see Table 3.10). The largest cluster lasts 21 years and it is still active. Cluster #3 spans a 19-year period and also remains to be active. In contrast, Cluster #6 on webometrics ends by 2006, but as we will see, relevant research finds its way in new specialties, notably in the form of altmetrics.

Major Specialties

In the following discussion, we will particularly focus on the five largest clusters. A research programme, or a paradigm, in a field of research can be characterized by its intellectual base and research fronts. The intellectual base is the collection of scholarly works that have been cited by the corresponding research community, whereas research fronts are the works that are inspired by the ones of the intellectual base. A variety of research fronts may rise from a common intellectual base.

Cluster #0—Science Mapping

Cluster #0 is the largest cluster, containing 214 references across a 21-year period from 1995 till 2015. The median year of all references in this cluster is 2006, but the median year of the 20 most representative citing articles to this cluster is 2010. This cluster's silhouette value of 0.748 is the lowest among the major clusters, but this is generally considered a relatively high level of homogeneity.

Table 3.10 Temporal properties of major clusters

Cluster ID	Size	Silhouette	From	To	Duration	Median	Sustainability	Activeness	Theme
0	214	0.748	1995	2015	21	2006	++++++	Active	Science mapping
1	209	0.765	1990	2006	17	1997	++	Inactive	Domain analysis
2	190	0.845	2000	2015	16	2009		Active	Research evaluation
3	160	0.954	1996	2014	19	2005	++++	Active	Information visualization/visual analytics
4	152	0.890	1988	1999	12	1993		Inactive	Applications of ACA
6	125	0.925	1995	2006	12	2001		Inactive	Webometrics
8	93	0.882	1994	2010	17	2002	++	Inactive	Bibliometric studies of social work in health
11	48	0.965	1994	2006	13	2000		Inactive	Bibliometric studies of management research
12	44	0.966	1990	1999	10	1996		Inactive	Graph visualization
16	29	0.977	1999	2007	9	2003		Inactive	Bibliometric studies of information systems
28	15	0.995	2004	2013	10	2008		Inactive	Global trend; water resources

Fig. 3.49 A hierarchy of key concepts selected from citing articles of Cluster #0 by log-likelihood ratio test

The primary focus of the large and currently active cluster is on the intellectual structure of a scientific discipline, a field of research, or any sufficiently self-contained domain of scientific inquiry. Key concepts identified from the titles of citing articles to this cluster can be algorithmically organized according to hierarchical relations derived from co-occurring concepts (Fig. 3.49). The largest branch of a such hierarchy typically reflects the core concepts of scholarly publications produced by the specialty behind the cluster. For example, concepts such as intellectual structure, co-citation analysis, co-authorship network underline the primary interest of this specialty.

We can use a simple method to classify various terms into two broad categories: domain-intrinsic or domain-extrinsic. Domain-intrinsic terms belong to the research field that aims to advance the conceptual and methodological capabilities of science mapping, for example, intellectual structure and co-citation analysis. In contrast, domain-extrinsic terms belong to the domain to which science mapping techniques are applied. In other words, they belong to the domain that is the object of a science mapping study. For example, stem cell research per se may not directly influence the advance of a specialty that is mainly concerned with how to identify the intellectual structure of a research field from scientific literature. Information science has a unique position. On the one hand, it is the discipline that hosts a considerable number of fields relevant to science mapping. On the other hand, it is the most frequent choice of a knowledge domain to test drive newly developed techniques and methods.

The timeline visualization reveals three periods of its development (Fig. 3.50). The first period is from 1995 to 2002. This period is relatively uneventful without high-profile references in terms of citation counts or bursts. Two visualization-centric domain analysis articles, Boyack2002 and Chen2002, preluded the subsequent wave of high-impact studies appeared in the second period. This period also features a social network analysis tool UCINET Borgatti2002.

The second period is from 2003 to 2010. Unlike the first period, the second period is full of high impact contributions—large citation tree rings and periods of citation bursts colored in red. Several types of high impact contributions appeared in this period, notably

- literature reviews—Börner 2003
- software tools—CiteSpace (Chen 2004), CiteSpace II (Chen 2006), CiteSpace III (Chen et al. 2010), VOSViewer (Van Eck and Waltman 2010)
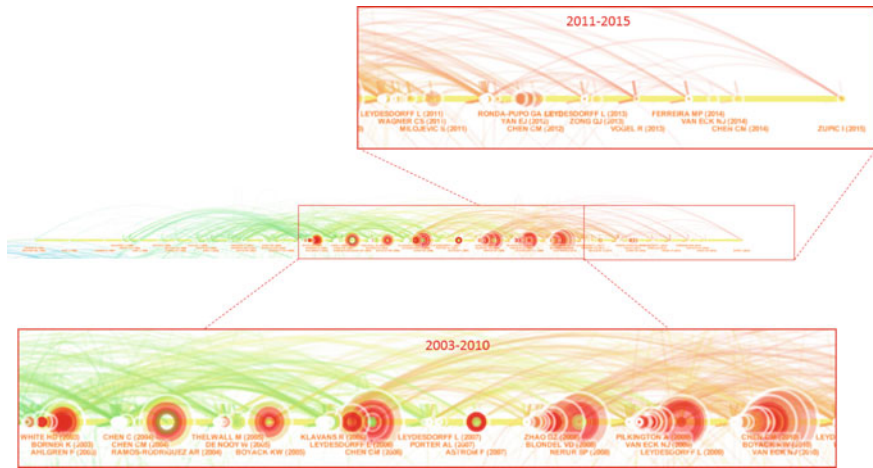
Fig. 3.50 High impact members of Cluster #0

- science mapping applications—visualization of information science White 2003, mapping the backbone of science Boyack et al. 2005 and a global map of science based on ISI subject categories Leydesdorff and Rafols 2009
- metrics and indicators—a critique on the use of Pearson's correlation coefficients as co-citation similarities—a previously common practice in ACA studies Ahlgren 2003
- applications to other domains—a bibliometric study of strategic management research Ramos-Rodriguez 2004 and another ACA of strategic management research Nerur 2008

The third period is from 2010 to 2015. Although no citation bursts were detected so far in this period, the themes of this period sheds additional insights into the more recent developmental status of the specialty. Most cited publications in this period include a study of the cognitive structure of library and information science —Milojević 2011 and a few studies that focus on domains with no apparent overlaps with computer and information science, for example regenerative medicine (Chen et al. 2012, 2014a, b) and strategic management—Vogel 2013.

A specialty may experience the initial conceptualization stage, the growth of research capabilities through the flourish of research tools, the expansion stage when researchers apply their methods to subject domains beyond the original research problems, and the final stage of decay (Shneider 2009). The largest cluster is dominated by an overwhelming number of tool-related references. As shown in Fig. 3.51, the top 20 most cited members of the cluster include several software tools such as CiteSpace, UCINET, VOSviewer, and global maps of science. If we follow Shneider's four-stage evolution model, the high concentration of software tools seems to suggest that the specialty behind this cluster evidently reached the second stage of its evolution by 2010. However, the several types of high-impact articles in this cluster, especially in the second period, suggest a far more complex picture.

Freq	Burst	Centrality	Σ	PageRank	...	Author	Year	Tit..	Source	Vol	Page	HalfLife	Cluster
80	24.90	0.02	1.81	0.00	...	Chen CM	2006	...	J AM SOC I...	V57	P359	6	0
39	12.90	0.04	1.69	0.00	...	White HD	2003	...	J AM SOC I...	V54	P423	5	0
67	19.84	0.01	1.25	0.00	...	Leydesdorff L	2009	...	J AM SOC I...	V60	P348	3	0
63	22.61	0.01	1.20	0.00	...	Boyack KW	2005	...	SCIENTOM...	V64	P351	6	0
73	19.33	0.01	1.17	0.00	...	Nerur SP	2008	...	STRATEG ...	V29	P319	6	0
53	18.11	0.01	1.16	0.00	...	Ahlgren P	2003	...	J AM SOC I...	V54	P550	6	0
38		0.01	1.16	0.00	...	Van Eck NJ	2009	...	J AM SOC I...	V60	P1635	5	0
32	10.78	0.01	1.13	0.00	...	White HD	2003	...	J AM SOC I...	V54	P1250	5	0
82		0.00	1.12	0.00	...	Van Eck NJ	2010	...	SCIENTOM...	V84	P523	5	0
27	11.59	0.01	1.11	0.00	...	Borgatti S P	2002	...	UCINET WI...	V	P	6	0
16	6.06	0.02	1.10	0.00	...	Boyack KW	2002	...	J AM SOC I...	V53	P764	6	0
51	16.88	0.00	1.08	0.00	...	Borner K	2003	...	ANNU REV...	V37	P179	7	0
16	5.54	0.01	1.08	0.00	...	White HD	2001	...	J AM SOC I...	V52	P87	7	0
25		0.01	1.07	0.00	...	De Nooy W	2005	...	EXPLORAT...	V	P	4	0
58	20.32	0.00	1.06	0.00	...	Chen CM	2010	...	J AM SOC I...	V61	P1386	4	0
54	18.09	0.00	1.06	0.00	...	Ramos-rodriguez...	2004	...	STRATEGI...	V25	P981	7	0
20	9.13	0.01	1.06	0.00	...	Gmur M	2003	...	SCIENTOM...	V57	P27	6	0
10	4.22	0.01	1.06	0.00	...	Ding Y	2000	...	SCIENTOM...	V47	P55	6	0
30	10.77	0.00	1.05	0.00	...	Rafols I	2009	...	J AM SOC I...	V60	P1823	4	0
18	6.82	0.01	1.05	0.00	...	Morris SA	2003	...	J AM SOC I...	V54	P413	6	0

Fig. 3.51 Top 20 most cited references in the largest cluster

The cluster includes several author co-citation studies of disciplines and research areas such as information science and strategic management. White 2003 revisits the intellectual structure of information science. Instead of using multidimensional scaling technique as they did in a previous study of the domain, the new study applied the Pathfinder network scaling technique and demonstrated the advantages of the technique. Pathfinder network scaling was first introduced to author co-citation analysis in (Chen 1999). The studies of strategic management research can be seen as applications outside the original specialty of author co-citation analysis. Furthermore, as we can see here, the application of ACA to a new target domain was made by researchers from the target domain several years after the analytic procedure was developed in information science. The techniques evidently spread to domains beyond information science. Fuchs's theory explains the speed of such diffusion in terms of the density of scientists' social network. Information travels faster in tightly coupled networks than loosely connected ones.

According to Shneider's evolution model, the application of tools to a new target should mark the beginning of the third stage. However, it seems we are seeing a considerable overlap between the second stage and the third stage. On the one hand, the development of new tools appears to be strengthening. There is no obvious sign that this trend would slow down anytime soon. On the other hand, the application of science mapping techniques to subject domains beyond information science appears to be a gradual process. As new tools have been developed, their applications are likely to follow. This particular example seems to suggest that techniques may be transferred in waves and that the speed of transfer is influenced by the structure of the networks of the researchers at the providing and the receiving ends.

Articles that cited members of the cluster convey additional information for us to understand the dynamics of the specialty (Fig. 3.52). The top 20 citing articles ranked by the bibliographic overlap with the cluster reveal similar types of contributions, namely software tools and techniques (1, 2, 5, 8, 14), new methods (9, 11, 16, 19, 20), surveys and reviews (3, 10, 13), and applications of bibliometric studies (6, 12, 17).

1. 0.11 van Eck, NJ (2010) software survey: vosviewer, a computer program for bibliometric mapping
2. 0.09 Rafols, I (2010) science overlay maps: a new tool for research policy and library management
3. 0.09 Bar-Ilan, J (2008) informetrics at the beginning of the 21st century - a review
4. 0.09 van Eck, NJ (2010) a comparison of two techniques for bibliometric mapping: multidimensional scaling and vos
5. 0.08 Vargas-Quesada, B (2010) showing the essential science structure of a scientific domain and its evolution
6. 0.07 Dolfsma, W (2010) the citation field of evolutionary economics
7. 0.07 Leydesdorff, L (2010) mapping the geography of science: distribution patterns and networks of relations among cities and institutes
8. 0.07 Quirin, A (2010) graph-based data mining: a new tool for the analysis and comparison of scientific domains represented as scientograms
9. 0.07 Waltman, L (2010) a unified approach to mapping and clustering of bibliometric networks
10. 0.06 Morris, SA (2008) mapping research specialties
11. 0.06 Takeda, Y (2010) tracking modularity in citation networks
12. 0.06 Uysal, OO (2010) business ethics research with an accounting focus: a bibliometric analysis from 1988 to 2007
13. 0.05 Borner, K (2003) visualizing knowledge domains
14. 0.05 Chen, CM (2010) the structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis
15. 0.05 Hicks, CC (2010) interdisciplinarity in the environmental sciences: barriers and frontiers
16. 0.05 Minguillo, D (2010) toward a new way of mapping scientific fields: authors' competence for publishing in scholarly journals
17. 0.04 Hofer, KM (2010) conference proceedings as a matter of bibliometric studies: the academy of international business 2006-2008
18. 0.04 Kurtz, MJ (2010) usage bibliometrics
19. 0.04 Leydesdorff, L (2013) interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations
20. 0.04 Zhao, DZ (2008) information science during the first decade of the web: an enriched author cocitation analysis

Fig. 3.52 Major citing articles to the largest cluster

The timeline visualization suggests that the specialty represented by the largest cluster has cumulated sufficient research techniques and tools by the end of the third period. It is likely that the specialty is ready for a larger scale of applications to subject domains rather than information science. According to Shneider's four-stage model, this is also the stage in while researchers may encounter anomalies that could lead to new discoveries and even the emergence of a new field.

At a more pragmatic level, one may monitor the further development of the specialty by tracking research fronts that are building on the early stages of the specialty. One can monitor emerging trends and patterns in terms of the major dimensions in the latent semantic space spanned by each year's publications connected to this particular cluster. For example, the growing number of domain-extrinsic terms such as nanotechnology, case study, and solar cell, suggest an expansion of the research scope—a hallmark of a third-stage specialty.

In summary, taken all the characteristics into account, the specialty seems to have a sustained second stage while clearly showing characteristics of the third stage in terms of Shneider's evolutionary model. Fuchs' theory provides a framework that one may pursue the diffusion of techniques from the origin of their developers to their users. In particular, one may trace the paths of the diffusion in the context of social networks of the researchers involved. Shneider's theory provides the most concrete account of how a specialty develops. Fuchs' theory provides the mid-range framework to embed the development of techniques in the context of social networks. Kuhn's theory seems to capture the dynamics at the highest level of abstraction. It is more likely that one would find evidence of a paradigm shift between distinct clusters than within the same cluster.

Cluster #1—Domain Analysis

Cluster #1 is the second largest cluster, containing 209 references that range a 17-year duration from 1990 to 2006. The cluster, or its underlying specialty, is
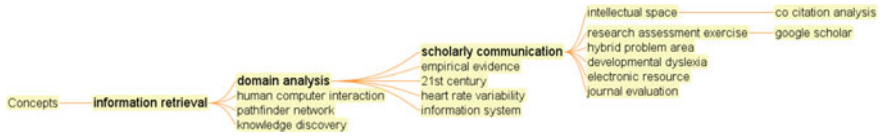
Fig. 3.53 A hierarchy of key concepts in Cluster #1
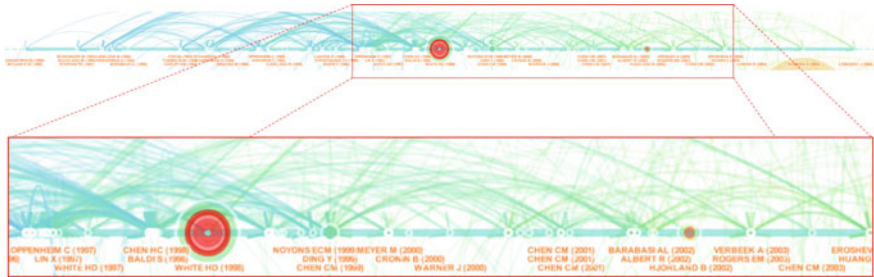
Fig. 3.54 Key members of Cluster #1

largely inactive with reference to the resolution of this study. This cluster is dominated by representative terms such as information retrieval, domain analysis, scholarly communication, and intellectual space (Fig. 3.53). Although information retrieval is the root node in the hierarchy of key terms in this cluster, domain analysis underlines the conceptual foundation of this cluster, as we will see shortly.

Two outstanding references from the timeline visualization of this cluster have strong citation burstness (Fig. 3.54). One is a domain analysis of information science (White and McCain 1998), in which the multidimensional scaling of an author co-citation space was utilized to visualize the intellectual structure of the domain. The other is a study of major approaches to domain analysis—Hjørland 2002. In early 1990s, Hjørland developed a domain-analytic approach, also known as sociological-epistemological approach or a socio-cognitive view, as a methodological alternative to the then methodological individualism and cognitive perspective towards information science that largely marginalized the social, historical, and cultural roles in understanding a domain of scientific knowledge. Hjørland's another article published in 1997 on domain analysis is also a member of the cluster.

The sigma score of a cited reference reflects its structural and temporal significance. In addition to the author co-citation analysis of information science (White and McCain 1998), two more author co-citation studies are ranked highly by their sigma scores, namely an author co-citation study of information retrieval—Ding 1999, and an author co-citation study of hypertext—Chen 1999 (Fig. 3.55).

The review article by White and McCain (1997) on visualization of literatures is an important member of the cluster, whereas Tabah's (1999) review of the study of literature dynamics is a citing article to the cluster. Although the term domain

Freq	Burst	Centrality	Σ	PageRank	...	Author	Year	Tit.	Source	Vol	Page	HalfLife	Cluster
71	26.53	0.02	1.57	0.00	...	White HD	1998	...	J AM SOC I...	V49	P327	5	1
18	6.70	0.02	1.17	0.00	...	Ding Y	1999	...	J INF SCI	V25	P67	4	1
14	5.94	0.03	1.17	0.00	...	Noyons ECM	1999	...	J AM SOC I...	V50	P115	3	1
26	10.32	0.01	1.08	0.00	...	White HD	1997	...	ANNU REV ...	V32	P99	5	1
18		0.01	1.08	0.00	...	Chen CM	1999	...	INFORM P...	V35	P401	4	1
14	5.50	0.01	1.07	0.00	...	Lin X	1997	...	J AM SOC I...	V48	P40	5	1
12	4.76	0.01	1.07	0.00	...	Baldi S	1998	...	AM SOCIOL...	V63	P829	6	1
10	3.93	0.01	1.05	0.00	...	Luukkonen T	1997	...	SCIENTOM...	V38	P27	5	1
12	4.76	0.01	1.04	0.00	...	Oppenheim C	1997	...	J DOC	V53	P477	6	1
13		0.01	1.03	0.00	...	Barabasi AL	2002	...	PHYSICA A	V311	P590	7	1
11		0.01	1.03	0.00	...	Chen C	1999	...	INFORMATI...	V	P	4	1
9	3.57	0.01	1.03	0.00	...	Kostoff RN	1998	...	SCIENTOM...	V43	P27	4	1
7	3.32	0.01	1.03	0.00	...	Garfield E	1998	...	LIBRI	V48	P67	2	1
15	6.50	0.00	1.02	0.00	...	Albert R	2002	...	REV MOD P...	V74	P47	6	1
10	3.93	0.00	1.02	0.00	...	Small H	1997	...	SCIENTOM...	V38	P275	5	1
9	3.82	0.01	1.02	0.00	...	Noyons ECM	1998	...	J AM SOC I...	V49	P68	2	1
8		0.01	1.02	0.00	...	Chen CM	2001	...	J AM SOC I...	V52	P315	2	1
7	3.60	0.00	1.02	0.00	...	Borgman CL	1992	...	J AM SOC I...	V43	P397	6	1
10		0.00	1.01	0.00	...	Hjorland B	1997	...	INFORMATI...	V	P	6	1
8	3.73	0.00	1.01	0.00	...	Ingwersen P	1997	...	J AM SOC I...	V48	P205	4	1

Fig. 3.55 Key members of Cluster #1, sorted by sigma

1. 0.65 Wilson, CS (1999) informetrics
2. 0.23 Ingwersen, P (1999) cognitive information retrieval
3. 0.18 Ding, Y (2000) bibliometric information retrieval system (birs): a web search interface utilizing bibliometric research results
4. 0.15 Boyack, KW (2002) domain visualization using vxinsight (r) for science and technology management
5. 0.15 Ding, Y (2000) incorporating the results of co-word analyses to increase search variety for information retrieval
6. 0.14 Borner, K (2003) visualizing knowledge domains
7. 0.14 Boyack, KW (2002) information visualization, human-computer interaction, and cognitive psychology: domain visualizations
8. 0.13 Cointet, JP (2007) how realistic should knowledge diffusion models be?
9. 0.13 Tabah, AN (1999) literature dynamics: studies on growth, diffusion, and epidemics
10. 0.11 Roth, C (2007) empiricism for descriptive social network models
11. 0.1 Borgman, CL (2002) scholarly communication and bibliometrics
12. 0.1 Ellis, D (1999) information science and information systems: conjunct subjects disjunct disciplines
13. 0.1 Ingwersen, P (2000) applying diachronic citation analysis to research program evaluations
14. 0.1 Rowlands, I (1999) patterns of scholarly communication in information policy: a bibliometric study
15. 0.08 Chen, CM (2002) visualizing and tracking the growth of competing paradigms: two case studies
16. 0.08 Jarvelin, K (2000) a user-oriented interface for generalised informetric analysis based on applying advanced data modelling techniques
17. 0.08 White, HD (2001) authors as citers over time
18. 0.07 Li, X (2007) patent citation network in nanotechnology (1976-2004)
19. 0.07 Song, M (2000) visualization in information retrieval: a three-level analysis
20. 0.06 Cronin, B (2000) semiotics and evaluative bibliometrics

Fig. 3.56 Citing articles to Cluster #1

analysis was not used consistently during the period of this cluster, the contributions consistently focus on holistic views of a knowledge domain. As Hjørland argued, domain analysis serves a fundamental role in information science because its goal is to understand the subject matter from a holistic view of sociological, cognitive, historical, and epistemological dimensions.

Citing articles to Cluster #1 include some of the earliest attempts to integrate information visualization techniques to the methodology of a domain analysis—Börner 2003, Boyack 2002, Chen 2002 (Fig. 3.56). Interestingly, some of these citing articles appear as cited references in Cluster #0. In other words, the downturn of Cluster #1 does not mean that researchers lost their interest in the domain analysis approaches. Rather, they shifted their focus to explore a new generation of domain analysis with the support of a variety of computational and visualization techniques. As a result, the specialty underline Cluster #0 continues the vision conceived in the works of Cluster #1. The citers of Cluster #1 identify the group of researchers who would be the core members of the specialty of the new generation of domain analysis.

Author co-citation analysis (ACA) plays an instrumental role in the development of the domain analysis specialty embodied in Cluster #1. It is not only a bibliometric method that has been adopted by researchers beyond information science, but also a research instrument that helps to reveal challenges that the next generation of domain analysis must deal with.

In their 1998 ACA study of information science, White and McCain masterfully demonstrated the power and the potential of what one may learn from a holistic view of the intellectual landscape of a discipline. They utilized the multidimensional scaling technique as a vehicle for visualization and tapped into their encyclopedic knowledge of the information science discipline in an intellectually rich guided tour across the literature. In an attempt to enrich and enhance the conventional methodology of ACA, Chen (1999) introduced the Pathfinder network scaling technique. Using Pathfinder networks brings several advantages to the methodology of ACA, including the ability to identify and preserve salient structural patterns and algorithmically derived visual cues to assist the navigation and interpretation of resultant visualizations. White (2003) revisited the ACA study of information science with Pathfinder network scaling. A fast algorithm to compute Pathfinder networks is published in 2008 (Quirin et al. 2008).

The re-introduction of the network thinking opens up a wider variety of computational techniques to an ACA study, notably network modeling and visualization. Furthermore, technical advances resulted from the improvement of ACA have been applied to a broader range of bibliometric studies, notably document co-citation analysis (DCA). As we will see shortly, the adaptation of network modeling and information visualization techniques in general results from a Stage III specialty of information visualization and visual analytics.

Cluster #2—Research Evaluation

Cluster #2 is the third largest cluster with 190 cited references and a silhouette value of 0.845, which is slightly higher than the previous two larger clusters #0 and #1, suggesting a higher homogeneity. In other words, one would consider this specialty a more specialized than the previously identified specialties. This cluster is active over a 16-year period from 2000 till 2015. It represents an active specialty.

The overarching theme of the cluster is suggested by the two major branches shown in the hierarchy of key terms of this cluster: the *information visualization* branch and the much larger branch of *research evaluation* (Fig. 3.57). The information visualization branch highlights the recurring themes of intellectual structure and co-citation analysis. The research evaluation branch highlights numerous concepts that are central to measuring scholarly impact, notably h-index, bibliometric ranking, bibliometric indicator, sub-field normalization, web indicator, citation distribution, social media metrics, and alternative metrics.

The 6-year period from 2005 through 2010 is a highly active period of the cluster (Fig. 3.58). The most prominent contributions in this period include the
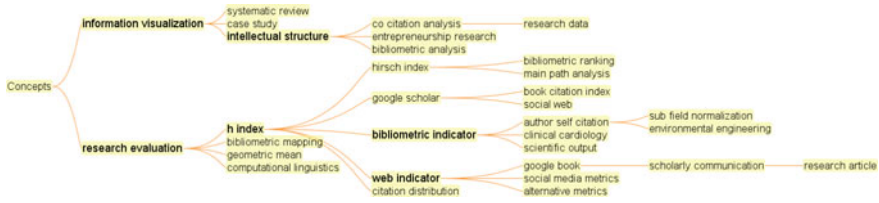
Fig. 3.57 A hierarchy of key concepts in Cluster #2
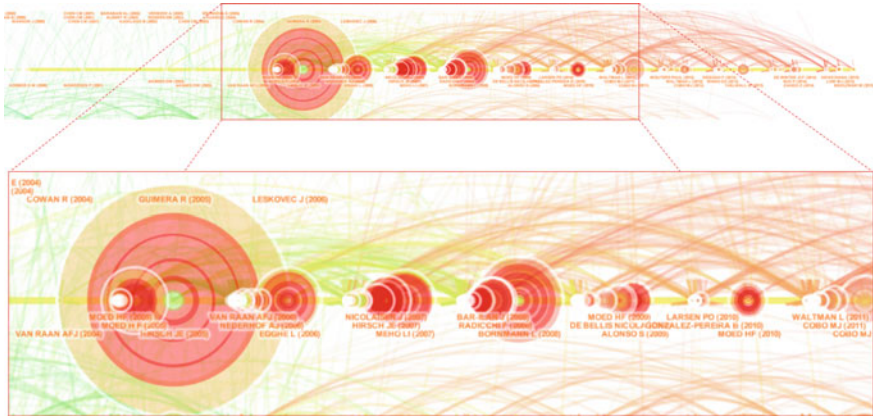
Fig. 3.58 High impact members of Cluster #2

original article that introduces the now widely known h-index (Hirsch 2005), the subsequent introduction of g-index as a refinement by taking citations into account (Egghe 2006), a 2007 study that compares the impact of using the Web of Science, Scopus, and Google Scholar on citation-based ranking—Meho 2007, a 2008 review entitled "What do citation counts measure?"—Bornmann 2008, and a study of the universality of citation distributions (Radicchi et al. 2008). These papers are also among the top sigma ranked members of this cluster because of their structural centrality as well as the strength of their citation burstness (Fig. 3.59).

The top 20 citing articles of the cluster reveal a considerable level of thematic consistency (Fig. 3.60). The overarching theme of research evaluation is evidently behind all these articles with popular title terms identified by latent semantic indexing such as citation impact, scientific impact, impact measures, bibliometric indicators, research evaluation, and web indicators.

Some of the more recent and highly cited members in Cluster #2 include a comparative study of 11 altmetrics and counterpart articles matched in the Web of Science (Thelwall et al. 2013) and the Leiden manifesto for research metrics (Hicks et al. 2015).

Freq	Burst	Centrality	Σ	PageRank	...	Author	Year	Tit..	Source	Vol	Page	HalfLife	Cluster
221	82.97	0.00	1.34	0.00	...	Hirsch JE	2005	...	P NATL AC...	V102	P16569	6	2
89	22.91	0.01	1.16	0.00	...	Bornmann L	2008	...	J DOC	V64	P45	5	2
68	17.76	0.01	1.13	0.00	...	Meho LI	2007	...	J AM SOC I...	V58	P2105	5	2
54	16.13	0.01	1.12	0.00	...	Radicchi F	2008	...	P NATL AC...	V105	P17268	5	2
17		0.01	1.12	0.00	...	Batista PD	2006	...	SCIENTOM...	V68	P179	4	2
49	12.52	0.01	1.11	0.00	...	Bar-ilan J	2008	...	SCIENTOM...	V74	P257	4	2
49	17.16	0.01	1.09	0.00	...	Moed HF	2010	...	J INFORME...	V4	P265	3	2
74	23.20	0.00	1.07	0.00	...	Egghe L	2006	...	SCIENTOM...	V69	P131	5	2
39	12.38	0.01	1.07	0.00	...	Nederhof AJ	2006	...	SCIENTOM...	V66	P81	6	2
48	12.54	0.01	1.07	0.00	...	Nicolaisen J	2007	...	ANNU REV ...	V41	P609	5	2
50		0.00	1.05	0.00	...	Cobo MJ	2011	...	J AM SOC I...	V62	P1382	4	2
32		0.00	1.05	0.00	...	Moed HF	2009	...	ARCH IMM...	V57	P13	5	2
42	14.70	0.00	1.04	0.00	...	Alonso S	2009	...	J INFORME...	V3	P273	5	2
68	25.47	0.00	1.04	0.00	...	Moed H F	2005	...	CITATION A...	V	P	5	2
32	10.20	0.00	1.04	0.00	...	Harzing A W K	2008	...	ETHICS SC...	V8	P61	5	2
19		0.00	1.03	0.00	...	Kousha K	2007	...	J AM SOC I...	V58	P1055	2	2
48	12.52	0.00	1.03	0.00	...	Hirsch JE	2007	...	P NATL AC...	V104	P19193	6	2
10	4.34	0.01	1.02	0.00	...	Egghe L	2006	...	SCIENTOM...	V69	P121	7	2
30	11.30	0.00	1.02	0.00	...	Van Raan AFJ	2006	...	SCIENTOM...	V67	P491	5	2
10		0.00	1.02	0.00	...	Bauer K	2005	...	D LIB MAGA...	V11	P	3	2

Fig. 3.59 High impact members of Cluster #2

1. 0.16 Egghe, L (2010) the hirsch index and related impact measures
2. 0.12 Zhang, L (2011) the diffusion of h-related literature
3. 0.12 Franceschet, M (2010) a comparison of bibliometric indicators for computer science scholars and journals on web of science and google scholar
4. 0.12 Haddow, G (2010) citation analysis and peer ranking of australian social science journals
5. 0.12 Lacasse, JR (2011) evaluating the productivity of social work scholars using the h-index
6. 0.1 Wu, H (2010) scientific impact at the topic level: a case study in computational linguistics
7. 0.09 Bornmann, L (2011) a multilevel modelling approach to investigating the predictive validity of editorial decisions: do the editors of a high profile journal select manuscripts that are highly cited after publication?
8. 0.09 Mas-Bleda, A (2014) do highly cited researchers successfully use the social web?
9. 0.09 Thelwall, M (2015) web indicators for research evaluation. part 2: social media metrics
10. 0.09 Waltman, L (2016) a review of the literature on citation impact indicators
11. 0.08 Thelwall, M (2016) are there too many uncited articles? zero inflated variants of the discretised lognormal and hooked power law distributions
12. 0.08 Thelwall, M (2016) the precision of the arithmetic mean, geometric mean and percentiles for citation data: an experimental simulation modelling approach
13. 0.07 Boell, SK (2010) journal impact factors for evaluating scientific performance: use of h-like indicators
14. 0.07 Mingers, J (2015) a review of theory and practice in scientometrics
15. 0.07 Saad, G (2010) applying the h-index in exploring bibliometric properties of elite marketing scholars
16. 0.07 Serenko, A (2010) the development of an ai journal ranking based on the revealed preference approach
17. 0.07 Thelwall, M (2015) web indicators for research evaluation. part 1: citations and links to academic articles from the web
18. 0.07 Thelwall, M (2016) interpreting correlations between citation counts and other indicators
19. 0.07 Thelwall, M (2016) the discretised lognormal and hooked power law distributions for complete citation data: best options for modelling and regression
20. 0.06 Bornmann, L (2010) from black box to white box at open access journals: predictive validity of manuscript reviewing and editorial decisions at atmospheric chemistry and physics

Fig. 3.60 Citing articles of Cluster #2

Cluster #3—Information Visualization and Visual Analytics

Cluster #3 is the fourth largest cluster. Its duration ranges from 2004 through 2014. The topic hierarchy has two branches: information visualization and heart rate variability (Fig. 3.61). The heart rate variability does not belong to the domain analysis in the context of information science. In fact, its inclusion in the original results of the topic search was due to the ambiguity of the term domain analysis across multiple disciplines. Pragmatically it is easier and more efficient to simply skip an irrelevant branch than keep refining the original topic search query until all noticeable irrelevant topics are eliminated. This is one of the fundamental challenges for information retrieval and this is where domain analysis has an instrumental role to play (Hjørland 2002).
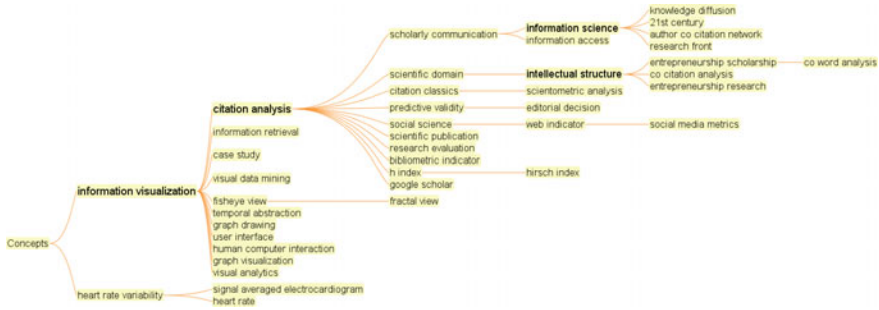
Fig. 3.61 A hierarchy of key concepts in Cluster #3

Fig. 3.62 High impact members of Cluster #3

The information visualization branch includes a mixture of information visualization techniques such as fisheye view, group drawing, graph visualization, and visual analytics and topics that are center to information science such as citation analysis, information retrieval. The mixture is a sign of attempts to apply information visualization and visual analytic techniques to bibliometric approaches to the study of intellectual structure of a research domain. The vision of information visualization is to identify insightful patterns from abstract information (Card et al. 1999). The subsequently emerged visual analytics emphasizes the critical and more specific role of sense-making and analytic reasoning in accomplishing such goals (Thomas and Cook 2005) (See Fig. 3.62).

High-impact contributions in Cluster #3 include the collection of seminal works in information visualization—Card 1999, a survey of graph visualization techniques —Herman 2000, Cytoscape—a widely used software tool for visualizing biomolecular interaction networks—Shannon 2003, the ground breaking work of visual analytics (Thomas and Cook 2005), Many Eyes—the popular web-based visualization platform—Viégas 2007, and a framework of seven types of interaction techniques in information visualization—Yi 2007 (Fig. 3.63).

In addition to the above high-impact contributions, this cluster features information visualization tools such as the InfoVis toolkit—Fekete 2004, NodeTrix—Henry 2007, Jigsaw—a visual analytic tool—Stasko 2008, and D3—Bostock 2011. The most widely used information visualization tools such as Many Eyes and D3 became available between 2007 and 2011. Figure 3.64 shows a list of citing articles of Cluster #3.

Freq	Burst	Centrality	Σ	PageRank	...	Author	Year	Tit.	Source	Vol	Page	HalfLife	Cluster
51	14.89	0.03	1.45	0.00	...	Thomas J J	2005	...	ILLUMINATI...	V	P	4	3
46		0.01	1.18	0.00	...	Bostock M	2011	...	IEEE T VIS ...	V17	P2301	4	3
46	18.80	0.01	1.14	0.00	...	Shannon P	2003	...	GENOME R...	V13	P2498	6	3
48	19.93	0.01	1.12	0.00	...	Card SK	1999	...	READINGS ...	V	P	4	3
19	7.26	0.01	1.07	0.00	...	Tufte ER	2001	...	VISUAL DIS...	V	P	5	3
40		0.00	1.07	0.00	...	Ware C	2004	...	INFORM VI...	V	P	6	3
31		0.00	1.07	0.00	...	Lam H	2012	...	IEEE T VIS ...	V18	P1520	2	3
44	11.47	0.00	1.05	0.00	...	Yi JS	2007	...	IEEE T VIS ...	V13	P1224	5	3
9		0.01	1.05	0.00	...	Henry N	2006	...	IEEE T VIS ...	V12	P677	2	3
31	11.71	0.00	1.04	0.00	...	Herman I	2000	...	IEEE T VIS ...	V6	P24	6	3
19	7.80	0.01	1.04	0.00	...	Henry N	2007	...	IEEE T VIS ...	V13	P1302	5	3
10		0.01	1.04	0.00	...	Watts DJ	1998	...	NATURE	V393	P440	7	3
17		0.01	1.03	0.00	...	Shneiderman B	2006	...	IEEE T VIS ...	V12	P733	3	3
12		0.00	1.03	0.00	...	Fekete J	2004	...	P IEEE S IN...	V.	P	5	3
14		0.00	1.02	0.00	...	Plaisant C	2004	...	P WORK C ...	V.	P	5	3
24	10.30	0.00	1.01	0.00	...	Keim DA	2002	...	IEEE T VIS ...	V8	P1	6	3
19	8.26	0.00	1.01	0.00	...	Singh J	2005	...	MANAGE SCI	V51	P756	7	3
9	3.33	0.00	1.01	0.00	...	Stasko J	2008	...	INFORM VI...	V7	P118	6	3
14		0.00	1.01	0.00	...	Keim D A	2010	...	MASTERIN...	V	P	4	3
11		0.00	1.01	0.00	...	Liu ZC	2010	...	IEEE T VIS ...	V16	P999	6	3

Fig. 3.63 Key members of Cluster #3

1. 0.14 Liu, SX (2014) a survey on information visualization: recent advances and challenges
2. 0.14 Morone, P (2004) knowledge diffusion dynamics and network properties of face-to-face interactions
3. 0.14 Morone, P (2004) small world dynamics and the process of knowledge diffusion: the case of the metropolitan area of greater santiago de chile
4. 0.12 Blanch, R (2007) browsing zoomable treemaps: structure-aware multi-scale navigation techniques
5. 0.1 Sedlmair, M (2014) visual parameter space analysis: a conceptual framework
6. 0.08 Cowan, R (2004) network structure and the diffusion of knowledge
7. 0.08 Gleicher, M (2011) visual comparison for information visualization
8. 0.08 von Landesberger, T (2011) visual analysis of large graphs: state-of-the-art and future research challenges
9. 0.07 Elmqvist, N (2010) hierarchical aggregation for information visualization: overview, techniques, and design guidelines
10. 0.07 Elmqvist, N (2011) fluid interaction for information visualization
11. 0.07 Lee, B (2006) treeplus: interactive exploration of networks with enhanced tree layouts
12. 0.05 Gelernter, J (2007) visual classification with information visualization (infoviz) for digital library collections
13. 0.05 Heer, J (2007) animated transitions in statistical data graphics
14. 0.05 Henry, N (2007) nodetrix: a hybrid visualization of social networks
15. 0.05 Hornbaek, K (2007) untangling the usability of fisheye menus
16. 0.05 Hornbaek, K (2011) the notion of overview in information visualization
17. 0.05 Liu, ZC (2014) ploceus: modeling, visualizing, and analyzing tabular data as networks
18. 0.05 Wong, PC (2006) generating graphs for visual analytics through interactive sketching
19. 0.04 Brandes, U (2011) asymmetric relations in longitudinal social networks
20. 0.04 Chen, M (2014) visual multiplexing

Fig. 3.64 Citing articles of Cluster #3

According to Shneider's four stage model, the information visualization and visual analytics specialty in the context of domain analysis and literature visualization has demonstrated properties of a Stage IV specialty. For example, in the most recent few years of the cluster, researchers reflect on empirical evaluations of information visualization in various scenarios—Lam 2012, revisit taxonomic organizations of abstract visualization tasks—Brehmer 2013, and synthesize and codify domain knowledge in the forms of textbooks—Munzner 2014.

Trajectories of Citations Across Cluster Boundaries

Cluster analysis helps us to understand the major specialties associated with science mapping. Now we turn our attention to the trajectories of several leading contributors in the landscape of these clusters. We are interested in what we may learn from citation links made in publications of a scholar, especially those links bridging distinct clusters.

Trajectories of Prolific Authors

The first example is the citation trajectory of Howard White (Fig. 3.65 left). He is the author of several seminal papers featured in several clusters. His citation trajectories move across the citation landscape from the left to the center, ranging from #4 decision support system (applications of ACA), #1 domain visualization (domain analysis), and #8 social work (another cluster of bibliometric studies).

The second example is the citation trajectory of Mike Thelwall (Fig. 3.65 right). He is a prolific researcher who contributed to webometrics and altmetrics among other areas of bibliometrics. An overlay of his citation trajectories on a citation landscape view shows that his trajectories spanning clusters such as #6 university websites (webometrics) and # google scholar (research evaluation).

In both examples of citation trajectories, we have observed that their citation trajectories span across a wide area over the citation landscape. Monitoring the movement of citation trajectories in such a way provides an intuitive insight into the evolution of the underlying specialties and the context in which high-impact researchers make their contributions.

Articles with Transformative Potentials

It is widely known that a major limitation of any citation-based indicators is their reliance on citations accumulated over time. Thus, citation-based indicators are likely to overlook newly published articles. An alternative method is to focus on the extent to which a newly published article brings to the conceptual structure of the knowledge domain of interest (Chen 2012). The idea is to identify the potential of an article to make extraordinary or unexpected connections across distinct clusters.
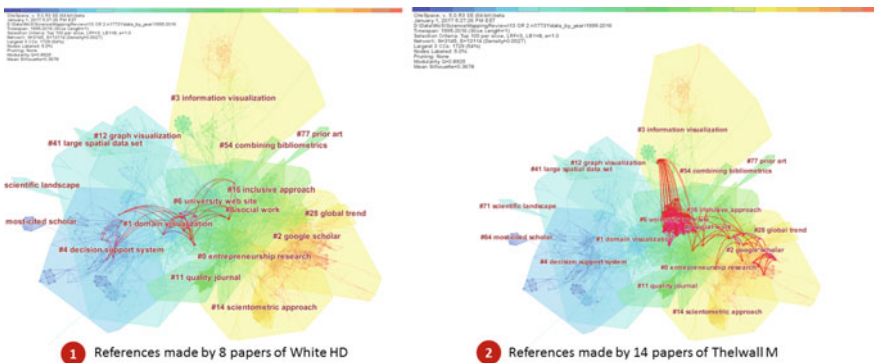
Fig. 3.65 Novel co-citations made by 8 papers of White HD (left) and by 14 papers of Thelwall M (right)

According to theories of scientific discovery, many significant contributions are resulted from boundary spanning ideas.

Table 3.11 lists three articles each year for the last five years. These articles have the highest geometric mean of three structural variation variables generated by CiteSpace. For example, in 2016, the highest score goes to the review of citation impact indicators—Waltman 2016, followed by two bibliometric analyses—one contrasts two closely related but distinct domains and the other studies the research over a 20-year span. In 2015, two bibliometric studies followed by a review of theory and practice in scientometrics (Mingers and Leydesdorff 2015).

Table 3.11 Potentially transformative papers published in recent years (2012–2016)

Year	ΔM	ΔCL_w	C_{KL}	Geometric mean	GC	Title	References
2016	6.0541	0.0152	0.0251	0.1322	5	A review of the literature on citation impact indicators	Waltman (2016)
2016	0.9235	0.0019	0.3407	0.0842	0	How are they different? A quantitative domain comparison of information visualization and data visualization (2000–2014)	Kim et al. (2016)
2016	0.8207	0.0017	0.0640	0.0447	2	A bibliometric analysis of 20 years of research on software product lines	Heradio et al. (2016)
2015	1.7498	0.0073	0.0380	0.0786	0	Global ontology research progress: A bibliometric analysis	Zhu et al. (2015)
2015	1.9873	0.0052	0.0397	0.0743	9	Bibliometric Methods in Management and Organization	Zupic (2015)
2015	1.9906	0.0029	0.0238	0.0516	13	A review of theory and practice in scientometrics	Mingers and Leydesdorff (2015)
2014	1.6240	0.0087	0.0434	0.0850	3	Research dynamics: Measuring the continuity and popularity of research topics	Yan (2014)
2014	1.1837	0.0031	0.0463	0.0554	1	Making a Mark: A computational and visual analysis of one researcher's intellectual domain	Skupin (2014)

Table 3.11 (continued)

Year	ΔM	ΔCL_w	C_{KL}	Geometric mean	GC	Title	References
2014	0.4462	0.0024	0.0270	0.0307	12	The Knowledge Base and Research Front of Information Science 2006–2010: An Author Co citation and Bibliographic Coupling Analysis	Zhao and Strotmann (2014)
2013	2.5398	0.0112	0.0643	0.1223	13	Analysis of bibliometric indicators for individual scholars in a large data set	Radicchi and Castellano (2013)
2013	1.0781	0.0065	0.2180	0.1152	6	A visual analytic study of retracted articles in scientific literature	Chen et al. (2013)
2013	1.7978	0.0064	0.0542	0.0854	24	Quantitative evaluation of alternative field normalization procedures	Li et al. (2013)
2012	3.6274	0.0107	0.0811	0.1466	29	SciMAT: A new science mapping analysis software tool	Cobo et al. (2011)
2012	3.4380	0.0248	0.0259	0.1302	15	A forward diversity index	Carley and Porter (2012)
2012	1.0719	0.0032	0.0321	0.0479	11	Visualizing and mapping the intellectual structure of information retrieval	Rorissa and Yuan (2012)

These highly ranked articles represent a few types of studies that may serve as predictive indicators, namely review papers (Mingers and Leydesdorff 2015; Waltman 2016), applications of bibliometric studies to specific domains, software tools for science mapping (Cobo et al. 2011), new metrics and indicators (Li et al. 2013), and visual analytic studies of unconventional topics—retractions (Chen et al. 2013). Figure 3.66 shows the trajectories of three articles with high modularity change rates.

The Emergence of a Specialty

The emergence of a specialty is determined by two factors: the intellectual base and the research fronts associated with the intellectual base. The intellectual base is what the specialty cites, whereas the research fronts are what the specialty is
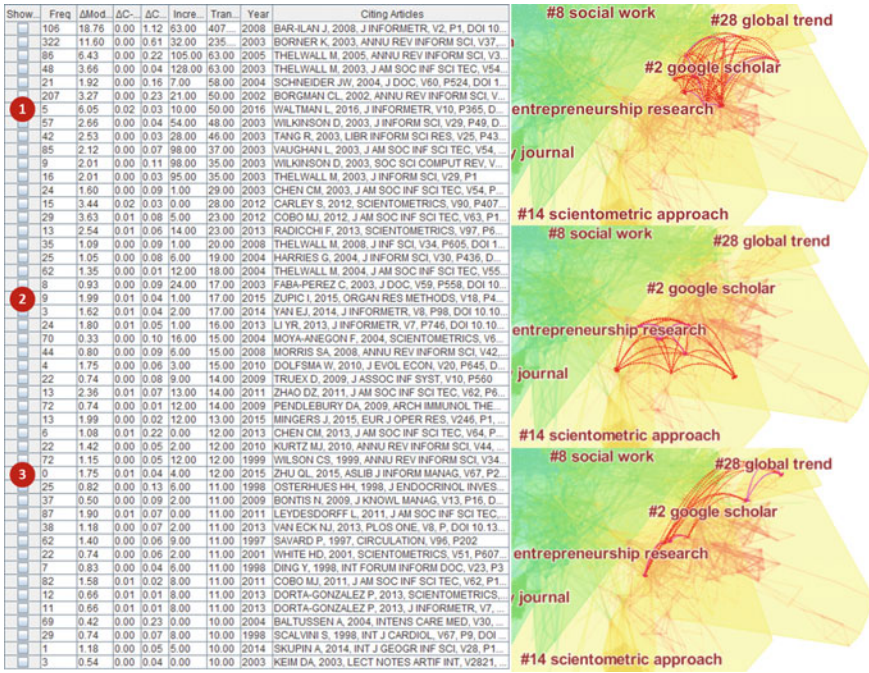
Show…	Freq	ΔMod…	ΔC-…	ΔC…	Incre…	Tran…	Year	Citing Articles
	106	18.76	0.00	1.12	63.00	407…	2008	BAR-ILAN J, 2008, J INFORMETR, V2, P1, DOI 10…
	322	11.60	0.00	0.61	32.00	235…	2003	BORNER K, 2003, ANNU REV INFORM SCI, V37,…
	86	6.43	0.00	0.22	105.00	63.00	2005	THELWALL M, 2005, ANNU REV INFORM SCI, V3…
	48	3.66	0.00	0.04	128.00	63.00	2003	THELWALL M, 2003, J AM SOC INF SCI TEC, V54…
	21	1.92	0.00	0.16	7.00	58.00	2004	SCHNEIDER JW, 2004, J DOC, V60, P524, DOI 1…
	207	3.27	0.00	0.23	21.00	50.00	2002	BORGMAN CL, 2002, ANNU REV INFORM SCI, V…
(1)	5	6.05	0.02	0.00	10.00	50.00	2016	WALTMAN L, 2016, J INFORMETR, V10, P365, D…
	57	2.66	0.00	0.04	54.00	48.00	2003	WILKINSON D, 2003, J INFORM SCI, V29, P49, D…
	42	2.53	0.00	0.03	28.00	46.00	2003	TANG R, 2003, LIBR INFORM SCI RES, V25, P43…
	85	2.12	0.00	0.07	98.00	37.00	2003	VAUGHAN L, 2003, J AM SOC INF SCI TEC, V54,…
	9	2.01	0.00	0.11	98.00	35.00	2003	WILKINSON D, 2003, SOC SCI COMPUT REV, V…
	16	2.01	0.00	0.03	95.00	35.00	2003	THELWALL M, 2003, J INFORM SCI, V29, P1
	24	1.60	0.00	0.09	1.00	29.00	2003	CHEN CM, 2003, J AM SOC INF SCI TEC, V54, P…
	15	3.44	0.02	0.03	0.00	28.00	2012	CARLEY S, 2012, SCIENTOMETRICS, V90, P407…
	29	3.63	0.01	0.08	5.00	23.00	2012	COBO MJ, 2012, J AM SOC INF SCI TEC, V63, P1…
	13	2.54	0.01	0.06	14.00	23.00	2013	RADICCHI F, 2013, SCIENTOMETRICS, V97, P6…
	35	1.09	0.00	0.09	1.00	20.00	2008	THELWALL M, 2008, J INF SCI, V34, P605, DOI 1…
	25	1.05	0.00	0.08	6.00	19.00	2004	HARRIES G, 2004, J INFORM SCI, V30, P436, D…
	62	1.35	0.00	0.01	12.00	18.00	2004	THELWALL M, 2004, J AM SOC INF SCI TEC, V55…
	8	0.93	0.00	0.09	24.00	17.00	2003	FABA-PEREZ C, 2003, J DOC, V59, P558, DOI 10…
(2)	9	1.99	0.01	0.04	1.00	17.00	2015	ZUPIC I, 2015, ORGAN RES METHODS, V18, P4…
	3	1.62	0.01	0.04	2.00	17.00	2014	YAN EJ, 2014, J INFORMETR, V8, P98, DOI 10.10…
	24	1.80	0.01	0.05	1.00	16.00	2013	LI YR, 2013, J INFORMETR, V7, P746, DOI 10.10…
	70	0.33	0.00	0.10	16.00	15.00	2004	MOYA-ANEGON F, 2004, SCIENTOMETRICS, V6…
	44	0.80	0.00	0.09	6.00	15.00	2008	MORRIS SA, 2008, ANNU REV INFORM SCI, V42…
	4	1.75	0.00	0.06	3.00	15.00	2010	DOLFSMA W, 2010, J EVOL ECON, V20, P645, D…
	22	0.74	0.00	0.08	9.00	14.00	2009	TRUEX D, 2009, J ASSOC INF SYST, V10, P560
	13	2.36	0.01	0.07	13.00	14.00	2011	ZHAO DZ, 2011, J AM SOC INF SCI TEC, V62, P6…
	72	0.74	0.00	0.01	12.00	14.00	2009	PENDLEBURY DA, 2009, ARCH IMMUNOL THE…
	13	1.99	0.00	0.02	12.00	13.00	2015	MINGERS J, 2015, EUR J OPER RES, V246, P1…
	6	1.08	0.01	0.22	0.00	12.00	2013	CHEN CM, 2013, J AM SOC INF SCI TEC, V64, P…
	22	1.42	0.00	0.05	2.00	12.00	2010	KURTZ MJ, 2010, ANNU REV INFORM SCI, V44…
(3)	72	1.15	0.00	0.05	12.00	12.00	1999	WILSON CS, 1999, ANNU REV INFORM SCI, V34
	0	1.75	0.01	0.04	4.00	12.00	2015	ZHU QL, 2015, ASLIB J INFORM MANAG, V67, P2…
	25	0.82	0.00	0.13	6.00	11.00	1998	OSTERHUES HH, 1998, J ENDOCRINOL INVES…
	37	0.50	0.00	0.09	2.00	11.00	2009	BONTIS N, 2009, J KNOWL MANAG, V13, P16, D…
	87	1.90	0.01	0.07	0.00	11.00	2011	LEYDESDORFF L, 2011, J AM SOC INF SCI TEC,…
	38	1.18	0.00	0.07	2.00	11.00	2013	VAN ECK NJ, 2013, PLOS ONE, V8, P, DOI 10.13…
	62	1.40	0.00	0.06	9.00	11.00	1997	SAVARD P, 1997, CIRCULATION, V96, P202
	22	0.74	0.00	0.04	2.00	11.00	2001	WHITE HD, 2001, SCIENTOMETRICS, V51, P607…
	7	0.83	0.00	0.04	6.00	11.00	1998	DING Y, 1998, INT FORUM INFORM DOC, V23, P3
	82	1.58	0.01	0.02	8.00	11.00	2011	COBO MJ, 2011, J AM SOC INF SCI TEC, V62, P1…
	12	0.66	0.01	0.01	8.00	11.00	2013	DORTA-GONZALEZ P, 2013, SCIENTOMETRICS,…
	11	0.66	0.01	0.01	8.00	11.00	2013	DORTA-GONZALEZ P, 2013, J INFORMETR, V7,…
	69	0.42	0.00	0.23	0.00	10.00	2004	BALTUSSEN A, 2004, INTENS CARE MED, V30,…
	29	0.74	0.00	0.07	8.00	10.00	1998	SCALVINI S, 1998, INT J CARDIOL, V67, P9, DOI…
	1	1.18	0.00	0.05	5.00	10.00	2014	SKUPIN A, 2014, INT J GEOGR INF SCI, V28, P1…
	3	0.54	0.00	0.04	0.00	10.00	2003	KEIM DA, 2003, LECT NOTES ARTIF INT, V2821,…

#8 social work #28 global trend #2 google scholar entrepreneurship research journal #14 scientometric approach

Fig. 3.66 Three examples of articles with high modularity change rates: (1) Waltman (2016), (2) Zupic (2015), and (3) Zhu et al. (2015)

currently addressing. As we have seen, on the one hand, a research front may remain in the same co-citation cluster as in the case of Cluster #2 Research Evaluation. On the other hand, a research front may belong to a different specialty and become the intellectual base of a new specialty as in the case of Cluster # Domain Analysis and Cluster #0 Bibliometric Mapping.

The citation trajectories of a researcher's publications and the positions of these publications as cited references can be simultaneously shown by overlaying trajectories (dashed lines for novel links or solid lines for existing links) and citing papers as stars if they also appear in a co-citation cluster as cited references. For example, the series of stars in the visualization shown in Fig. 3.67 tell us two things: First, the author is connecting topics in two clusters (Cluster #0 Science Mapping and Cluster 2 Research Evaluation) and second, the author belongs to the specialty of science mapping.

The example in Fig. 3.68 illustrates the citation trajectories of Howard White's publications and their own positions in the timelines of clusters. His publications appear in the early stage of the science mapping cluster (#0) and make novel connections between science mapping and domain analysis (Cluster #1), domain analysis (Cluster #1) and applications of ACA (Cluster #4), domain analysis (Cluster #1) and webometrics (Cluster #6).
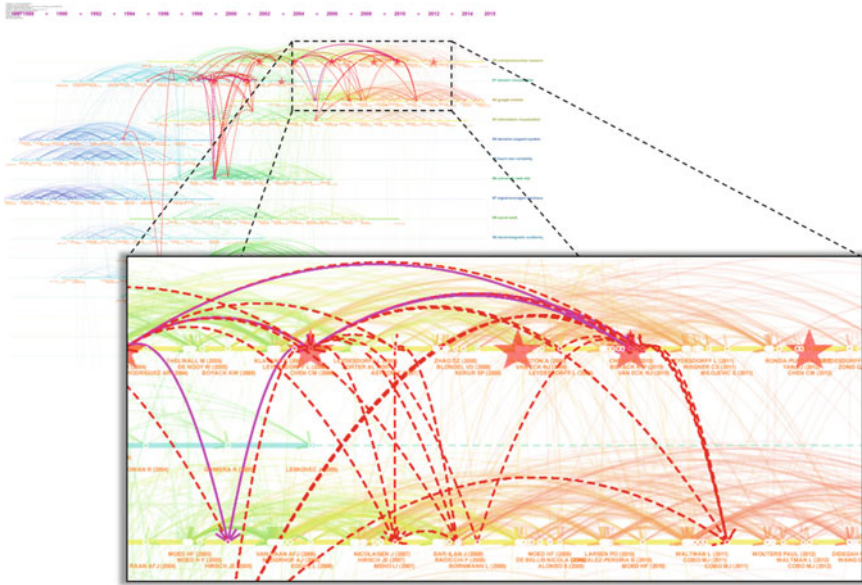
Fig. 3.67 Stars indicate articles that are both cited and citing articles. Dashed lines indicate novel co-citation links. Illustrated based on 15 papers of the author's own publications
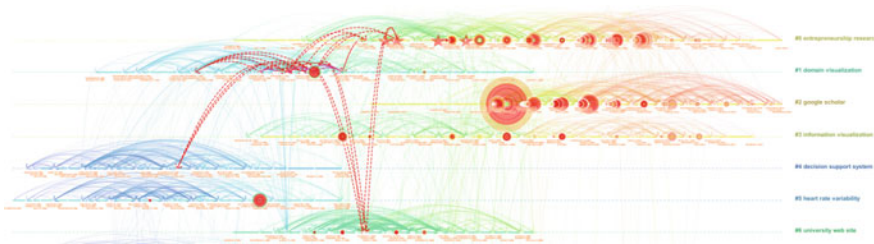

Fig. 3.68 Citation trajectories of Howard White's publications and their own locations

The next example in Fig. 3.69 depicts the novel co-citation links made by a review paper of informetrics (Bar-Ilan 2008). These novel links include within-cluster links as well as between-cluster links. It should be easy to tell that the scope of the review is essentially limited to research papers published about 6–7 years prior to the time of the review. Furthermore, we can see that the review systematically emphasizes the diversity of topics instead of tracing to the origin of any particular specialty.
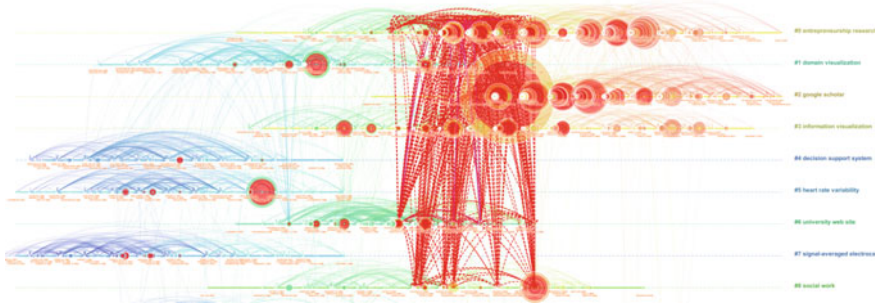
Fig. 3.69 Novel links made by a review paper of informetrics (Bar-Ilan 2008)

Summary

We present three examples of visually exploring the scientific literature of a field of study. Our intention is twofold. First, our goal is to demonstrate the depth of a systematic review that one can reach by applying a science mapping approach to terrorism research and the science mapping domain itself. The first example of terrorism research is based on publications between 1996 and 2003. The second example of terrorism research is based on a much longer timespan between 1980 and 2017, with particular interests in how the visual analytic approach is sensitive to the latent changes over the years. The third example is the science mapping field itself.

In addition to the application of computational functions available in the CiteSpace software, we also enrich the procedure of producing a systematic review of a knowledge domain by incorporating evolutionary models of a scientific specialty—especially the four-stage model of a scientific discipline into the interpretation of the identified specialties. Our interpretation not only identifies thematic milestones of major streams of science mapping research, but also characterizes the developmental stages of the underlying specialties and the dynamics of transitions from one specialty to another.

Second, our goal is to provide a reliable historiographic survey of the science mapping research. The survey identifies the major clusters in terms of their high-impact members and citing articles that form new research fronts. We also demonstrate new insights that one can intuitively obtain through an inspection of citation trajectories and the positions of citing papers. The enhanced science mapping procedure introduced in this article is applicable to the analysis of other domains of interest. Researchers can utilize these visual analytic tools to perform timely surveys of the literature as frequently as they wish and find relevant publications more effectively.

References

Abt HA (1998) Why some papers have long citation lifetimes. Nature 395:756–757

Bar-Ilan J (2008) Informetrics at the beginning of the 21st century—a review. J Informetrics 2 (1):1–52

Batagelj V (2003) Efficient algorithms for citation network analysis. arXiv:cs/0309023v1

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

Boyack KW, Klavans R (2010) Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? J Am Soc Inform Sci Technol 61(12):2389–2404. doi:10.1002/asi.21419

Boyack KW, Klavans R, Börner K (2005) Mapping the backbone of science. Scientometrics 64 (3):351–374

Burt RS (1992) Structural holes: the social structure of competition. Harvard University Press, Cambridge, MA

Bush V (1945) As we may think. Atlantic Monthly 176(1):101–108

Callon M, Courtial JP, Turner WA, Bauin S (1983) From translations to problematic networks—an introduction to co-word analysis. Soc Sci Inf Sci Soc 22(2):191–235

Card S, Mackinlay DJ, Shneiderman B (1999) Readings in information visualization: using vision to think. Interactive technologies. Morgan Kaufmann Publisher, San Francisco, CA USA

Carley KM, Hummon NP, Harty M (1993) Scientific influence—an analysis of the main path structure in the journal of conflict-resolution. Knowledge-Creation Diffus Utilization 14 (4):417–447

Carley S, Porter AL (2012) A forward diversity index. SCIENTOMETRICS 90(2):407–427. doi:10.1007/s11192-011-0528-1

Chen C (2004) Searching for intellectual turning points: progressive knowledge domain visualization. Proc Natl Acad Sci U S A 101 (Suppl):5303–5310

Chen C (1999) Visualising semantic spaces and author co-citation networks in digital libraries. Inf Process Manage 35(2):401–420

Chen C (2006) CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inform Sci Technol 57(3):359–377

Chen C (2012) Predictive effects of structural variation on citation counts. J Am Soc Inform Sci Technol 63(3):431–449. doi:10.1002/asi.21694

Chen C (2014) The fitness of information: quantitative assessments of critical evidence. Wiley

Chen C (2017) Science mapping: a systematic review of the literature. J Data Inform Sci 2(2):1–40

Chen C, Leydesdorff L (2014) Patterns of connections and movements in dual-map overlays: a new method of publication portfolio analysis. J Assoc Inform Sci Technol 65(2):334–351

Chen C, Chen Y, Horowitz M, Hou H, Liu Z, Pellegrino D (2009) Towards an explanatory and computational theory of scientific discovery. J Informetrics 3(3):191–209

Chen C, Ibekwe-SanJuan F, Hou J (2010) The structure and dynamics of co-citation clusters: a multiple-perspective co-citation analysis. J Am Soc Inform Sci Technol 61(7):1386–1409

Chen C, Hu Z, Liu S, Tseng H (2012) Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. Exp Opin Biol Ther 12(5):593–608

Chen C, Hu Z, Milbank J, Schultz T (2013) A visual analytic study of retracted articles in scientific literature. Journal of the American Society for Information Science and Technology 64 (2):234–253

Chen C, Dubin R, Kim MC (2014a) Emerging trends and new developments in regenerative medicine: a scientometric update (2000–2014). Expert Opin Biol Ther 14(9):1295–1317

Chen C, Dubin R, Kim MC (2014b) Orphan drugs and rare diseases: a scientometric review (2000–2014). Exp Opin Orphan Drugs 2(7):709–724

Cobo MJ, Lopez-Herrera AG, Herrera-Viedma E, Herrera F (2011) Science mapping software tools: review, analysis, and cooperative study among tools. J Am Soc Inform Sci Technol 62 (7):1382–1402. doi:10.1002/asi.21525

Deerwester S, Dumais ST, Landauer TK, Furnas GW, Harshman RA (1990) Indexing by latent semantic analysis. J Am Soc Inform Sci 41(6):391–407

Egghe L (2006) Theory and practise of the g-index. Scientometrics 69(1):131–152. doi:10.1007/s11192-006-0144-7

Garfield E (1955) Citation indexes for science: a new dimension in documentation through association of ideas. Science 122(3159):108–111

Heer J (2007) The prefuse visualization toolkit. http://prefuse.org/

Heradio R, Perez-Moragoa H, Fernandez-Amorosa D, Cabrerizoa FJ, Herrera-Viedmab E (2016) A bibliometric analysis of 20 years of research on software product lines. Information and Software Technology 72:1–15. doi:10.1016/j.infsof.2015.11.004

Herman I, Melançon G, Marshall MS (2000) Graph visualization and navigation in information visualization: a survey. IEEE Trans Visual Comput Graph 6(1):24–44

Hicks D, Wouters P, Waltman L, Rijcke Sd, Rafols I (2015) Bibliometrics: the Leiden Manifesto for research metrics. Nature 520(7548):429–431. doi:10.1038/520429a

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 102(46):16569–16572. doi:10.1073/pnas.0507655102

Hjørland B (2002) Epistemology and the socio-cognitive perspective in information science. J Am Soc Inform Sci Technol 53(4):257–270

Hummon NP, Doreian P (1989) Connectivity in a citation network—the Development of DNA Theory. Soc Netw 11(1):39–63

Johnson B, Shneiderman B (1991) Tree-maps: a space filling approach to the visualization of hierarchical information structures. IEEE Vis 91(Oct 1991):284–291

Kim MC, Zhu Y, Chen C (2016) How are they different? A quantitative domain comparison of information visualization and data visualization (2000-2014). SCIENTOMETRICS 107:123

Kleinberg J (2002) Bursty and hierarchical structure in streams. In: Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining, pp 91–101

Leydesdorff L, Rafols I (2009) A global map of science based on the ISI subject categories. J Am Soc Inform Sci Technol 60(2):348–362

Li Y, Radicchi F, Castellano C, Ruiz-Castillo J (2013) Quantitative evaluation of alternative field normalization procedures. J Informetrics 7(3):746–755

Liu JS, Kuan C-H (2015) A new approach for main path analysis: Decay in knowledge diffusion. J Assoc Inform Sci Technol. doi:10.1002/asi.23384

Liu JS, Lu LYY (2012) An integrated approach for main path analysis: Development of the Hirsch Index as an example. J Am Soc Inform Sci Technol 63(3):528–542

Lucio-Arias D, Leydesdorff L (2008) Main-path analysis and path-dependent transitions in HistCite-based historiograms. J Am Soc Inform Sci Technol 59(12):1948–1962

Mingers J, Leydesdorff L (2015) A review of theory and practice in scientometrics. Eur J Oper Res 246(1):1–19

Morris SA, Yen G, Wu Z, Asnake B (2003) Timeline visualization of research fronts. J Am Soc Inform Sci Technol 55(5):413–422

Price DD (1965) Networks of scientific papers. Science 149:510–515

Quirin A, Cordon O, Santamaria J, Vargas-Quesada B, Moya-Anegon F (2008) A new variant of the Pathfinder algorithm to generate large visual science maps in cubic time. Inform Process Manag 44(4):1611–1623. doi:10.1016/j.ipm.2007.09.005

Radicchi F, Castellano C (2013) Analysis of bibliometric indicators for individual scholars in a large data set. SCIENTOMETRICS 97(3):627–637. doi:10.1007/s11192-013-1027-3

Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: toward an objective measure of scientific impact. PNAS 105(45):17268–17272

Rorissa A, Yuan X (2012) Visualizing and mapping the intellectual structure of information retrieval. Information Processing & Management 48(1):120–135. doi:10.1016/j.ipm.2011.03.004

Schvaneveldt RW (ed) (1990) Pathfinder associative networks: studies in knowledge organization. Ablex series in computational sciences. Ablex Publishing Corporations, Norwood, New Jersey

Shneider AM (2009) Four stages of a scientific discipline: four types of scientists. Trends Biochem Sci 34(5):217–223

Shneiderman B (1996) The eyes have it: a task by data type taxonomy for information visualization. In: IEEE workshop on visual language, Boulder, CO, Sept 3–6, 1996. IEEE Computer Society Press, pp 336–343

Skupin A (2014) Making a Mark: a computational and visual analysis of one researcher's intellectual domain. International Journal of Geographical Information Science 28(6):1209–1232. doi:10.1080/13658816.2014.906040

Smalheiser NR, Swanson DR (1998) Using ARROWSMITH: a computer-assisted approach to formulating and assessing scientific hypotheses. Comput Meth Programs Biomed 57(3):149–153

Small H (1973) Co-citation in the scientific literature: a new measure of the relationship between two documents. J Am Soc Inform Sci 24:265–269

Tabah AN (1999) Literature dynamics: studies on growth, diffusion, and epidemics. Annu Rev Inform Sci Technol 34:249–286

Thelwall M, Haustein S, Larivière V, Sugimoto CR (2013) Do altmetrics work? Twitter and ten other social web services. PLoS ONE 8(5):e64841

Thomas JJ, Cook AK (2005) Illuminating the path: the research and development agenda for visual analytics. IEEE Press

Tibély G, Pollner P, Vicsek T, Palla G (2013) Extracting tag-hierarchies. PLoS ONE 8(12):e84133

Van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84(2):523–538

Van Eck NJ, Waltman L (2014) CitNetExplorer: a new software tool for analyzing and visualizing citation networks. J Informetrics 8(4):802–823

van Raan A (2000) On growth, ageing, and fractal differentiation of science. Scientometrics 47(2):347–362

Waltman L (2016) A review of the literature on citation impact indicators. J Informetr 10(2):365–391. doi:10.1016/j.joi.2016.02.007

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442

White HD, McCain KW (1997) Visualization of Literatures. Annu Rev Inform Sci Technol 32:99–168

White HD, McCain KW (1998) Visualizing a discipline: an author co-citation analysis of information science, 1972–1995. J Am Soc Inform Sci 49(4):327–356

Yan E (2014) Research dynamics: Measuring the continuity and popularity of research topics. Journal of Informetrics 8(1):98–110. doi:10.1016/j.joi.2013.10.010

Zhu Q, Kong X, Hong S, Li J, He Z (2015) Global ontology research progress: a bibliometric analysis. Aslib Journal of Information Management 67(1):27–54. doi:10.1108/AJIM-05-2014-0061

Zupic I (2015) Bibliometric methods in management and organization. Organizational Research Methods 18(3):429–472