Visualizing and Tracking the Growth of Competing Paradigms: Two Case Studies

Chaomei Chen
College of Information Science and Technology, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104-2875. E-mail: chaomei.chen@cis.drexel.edu

Timothy Cribbin, Robert Macredie, and Sonali Morar
The VIVID Research Center, Department of Information Systems and Computing, Brunel University, Uxbridge UB8 3PH, UK. E-mail: {timothy.cribbin, robert.macredie, sonali.morar}@brunel.ac.uk

In this article we demonstrate the use of an integrative approach to visualizing and tracking the development of scientific paradigms. This approach is designed to reveal the long-term process of competing scientific paradigms. We assume that a cluster of highly cited and cocited scientific publications in a cocitation network represents the core of a predominant scientific paradigm. The growth of a paradigm is depicted and animated through the rise of citation rates and the movement of its core cluster towards the center of the cocitation network. We study two cases of competing scientific paradigms in the real world: (1) the causes of mass extinctions, and (2) the connections between mad cow disease and a new variant of a brain disease in humans—vCJD. Various theoretical and practical issues concerning this approach are discussed.

Introduction

Government agencies, research funding authorities, international companies, as well as individual scientists are under increasing pressure to ensure that they can benefit from the latest scientific and technological developments and improve the competitiveness of their operations. In this context, knowledge tracking and technology monitoring tools have become increasingly desirable. The rapid advances of information visualization in the past few years have highlighted several promising areas in relation to knowledge tracking and technology monitoring (Card, Mackinlay, & Shneiderman, 1999; Chen, 1999a; Spence, 2001).

A considerable amount of work in information visualization has focused on depicting salient semantic structures of words and documents, such as Bead (Chalmers, 1992), VIBE (Olsen, Korfhage, & Sochats, 1993), VR-VIBE (Benford, Snowdon, Greenhalgh, Ingram, Knox, & Brown, 1995), and ThemeScape (Wise, Thomas, Pennock, Lantrip, Pottier, Schur, & Crow, 1995). In this type of information visualization, salient semantic structures are typically derived from patterns of the word-frequency distribution in a collection of documents. In ThemeScape, a peak in a three-dimensional information terrain model, indicates the presence of a theme—a high concentration of the use of certain words.

The tradition of deriving higher level structures by analyzing word-occurrence patterns in text can be traced back to the co-word analysis method developed in 1980s (Callon, Courtial, Turner, & Bauin, 1983; Callon, Law, & Rip, 1986). Co-word analysis has become a well-established camp in scientometrics—the field of study that concerns with indicators and metrics of the dynamics of science and technology at large. Co-word analysis was motivated by a profound theory about the dynamics of science and technology—the actor-network theory (ANT). The primary purpose of the co-word analysis methodology developed in the 1980s was to help researchers analyze the dynamics of science and technology. One of the fundamental premises of the ANT theory is that scientists use scientific publications as a vehicle to put forward their ideas, to shape the invisible socio-technological web, and eventually to influence our everyday life in the society. The outcome of co-word analysis was typically depicted as a network of concepts. The term *leximappe* was used to refer to such a concept map. More recently, researchers have incorporated artificial neural network techniques such as self-organized maps into co-word analysis to depict patterns and trends derived from text (Noyons & van Raan, 1998).

Cocitation analysis is another vibrantly active camp in scientometrics. The first wave of landmark cocitation anal-

Received October 1, 2001; revised December 17, 2001; accepted December 17, 2001

ysis studies appeared in the 1970s, notably (Small, 1973, 1977, 1997, 1999). Instead of deriving a salient semantic structure from the distribution of word frequencies, cocitation analysis emphasizes the unique value of higher order interrelationships between documents or between authors. A good example of combining cocitation and co-word analysis can be found in Braam, Moed, and van Raan (1991). Author cocitation analysis was first introduced in 1981, and it is now a routinely used method in citation analysis (White & Griffith, 1981; White & McCain, 1998). The more recent wave of cocitation analysis studies include (Chen, 1999b, Chen, 2002; Chen, Kuljis, & Paul, 2001; Chen & Paul, 2001). Mapping scientific frontiers have begun to attract the attention of researchers and practitioners. Both co-word analysis and cocitation analysis are playing an increasingly important role (Chen, 2002).

The advances of information visualization as a field have particularly revived the interest in knowledge discovery and knowledge tracking. In this article, we focus on the use of a particular approach to visualizing and tracking the growth of scientific paradigms. We illustrate the potential of this approach with two case studies. The first case study investigates the role of information visualization in tracking the growth of the study of mass extinctions. The second case study tracks down the line of research concerning whether there is a connection between mad cow disease and new variant Creutzfeldt-Jakob disease (vCJD). The rest of the article is organized as follows: we first provide a brief introduction to the key concepts and principles. Then we explain how our approach works and what types of structural and visual properties we should look for in the case studies. We describe two case studies in detail. We finally reflect on our experience with these case studies in the broader context of knowledge tracking and technology monitoring.

Scientific Revolutions

Thomas Kuhn's *The Structure of Scientific Revolutions* (Kuhn, 1962) is one of the most influential works in the 20th century. According to Kuhn's theory, most of the time scientists are engaged in what is known as normal science. A period of normal science is typically marked by the dominance of an established framework. The majority of scientists would work on specific hypotheses within such frameworks. The foundations of such frameworks largely remain unchallenged until new discoveries begin to cast doubts over fundamental issues—science falls into a period of crises. To resolve such crises, radically new theories with greater explanatory power are introduced. New theories replace the ones in trouble in a revolutionary manner. Science regains another period of normal science. According to Kuhn, scientific revolutions are an integral part of science and science progresses through such revolutionary changes.

Kuhn characterized the structure of scientific revolutions in terms of the dynamics of scientific paradigms. The revolutionary transform of science from one paradigm to another is now widely known as a paradigm shift. "Paradigm shifts" have become one of the most widely used concepts in everyday language, as well as in scientific language. Classic examples of paradigm shifts include the Copernican revolution, which established the Sun as the center of our solar system rather than the earth, and Einstein's theory of general relativity, which replaced the predominant place of Newtonian mechanics.

Kuhn's theory has been well received on the one hand by, for example, sociologists of science. On the other hand, philosophers have launched various criticisms, especially on Kuhn's earlier views on incommeasurability, which refers to the communicative barrier between different paradigms. Paul Thagard proposed an approach called conceptual revolutions (Thagard, 1992). Thagard argued that the explanatory coherence of a scientific theory should be used as a key criterion for the selection of theories. For example, if a theory with fewer assumptions can explain more phenomena than an alternative theory, then the simpler one is better, and should replace the more complex one. Thagard explained the idea of conceptual revolutions with examples such as the conceptual development of plate tectonics in the latest geological revolution and Darwin's natural selection theory.

Information scientists are concerned with patterns of scientific communications and intellectual structures of scientific disciplines. Since the 1970s, information scientists began to look for signs of competing paradigms in scientific literature, for example, a rapid change of research focus within a short period of time. In 1974, Small and Griffith were among the first to address issues concerning identifying and mapping specialities from the structure of scientific literature, especially based on cocitation patterns (Small & Griffith, 1974). In 1977, Small conducted a longitudinal study of collagen research and showed how and when some rapid changes of focus had taken place (Small, 1977). He used data from the Science Citation Index (SCI) to compute cocitation strengths between pairs of documents and subsequently clustered documents to identify leading specialities, or paradigms. He then used multidimensional scaling to map highly cited articles each year in clusters on a two-dimensional plane. An abrupt disappearance of a few key documents in the leading cluster in one year and the rapidly increased number of documents in the leading cluster in the following year indicate an important type of specialty change—rapid shift in research focus—that is an indicator of "revolutionary" changes.

We have also drawn some useful insights from studies of thematic maps of geographic information. For example, if people study a geographic map first and read relevant text later, they can remember more information from the text (Rittschof, Stock, Kulhavy, Verdi, & Doran, 1994). Traditionally, a geographic map shows two important types of information: structural and feature information. Structure information helps us to locate individual landmarks on the map and determine spatial relations among them. Feature information refers to detail, shape, size, color, and other visual properties used to depict particular items on a map. When people study a map, they first construct a mental image of the map's general spatial framework and add the landmarks into the image subsequently (Rittschof et al., 1994). The mental image integrates information about individual landmarks in a single relatively intact piece, which allows rapid and easy access to the landmarks embedded. The greater the integration of structural and feature information in the image, the more intact the image is. The more intact the image, the more easily landmark information can be located and help retrieval of further details. If we visualize a paradigm as a cluster of highly cited landmark articles and combine citation and cocitation into the same visualization model, then this will allow users to construct an intact image of a network of top-sliced articles from the chosen subject domain.

Recently, Chen and Paul described an integrated approach to visualizing a knowledge domain's intellectual structure based on author cocitation analysis by utilizing techniques such as Pathfinder network scaling. See (Chen & Paul, 2001) for technical details. An in-depth domain analysis using this technique is reported in (Chen, Paul, & O'Keefe, 2001). This approach integrates techniques such as Pathfinder network scaling and factor analysis to visualize an intellectual structure as a network of salient cocitation relationships. It is this network that provides a context in which visualizing the growth of citation impact reflects the consequence of a paradigm shift.

In this article, we extend the existing approach and draw upon Kuhn's notion of the structure of scientific revolution (Kuhn, 1962), in particular, the notion of paradigm, so as to focus specifically on identifying and tracking the growth of competing paradigms. Kuhn's paradigm theory is among a broad range of potentially feasible theories that can drive a new wave of studies in general public understanding of science as well as scientometrics. The details of the method are described in the next section.

Methods

Our intention is to make the process simple and easy to use, especially for people who may not have specialized knowledge. The best place to find competing paradigms is a topic that is known or is likely to be controversial and hotly debated. We include two case studies in this article to illustrate our approach. One is about the mass extinction debates, and the other is about the possible connection between mad cow disease and the new variant of CJD. Our users may have heard about these topics elsewhere, but all they need to get started is the top-level terms: mass extinction for the first case study, and BSE and CJD for the second case study. BSE is the abbreviation of the formal name of mad cow disease—Bovine Spongiform Encephalopathy.

Once our users decide which topics they want to explore, the next step is to gather citation data. Currently, the science and social science citation databases compiled by the Institute for Scientific Information (ISI) are among the best sources of citation data. It is also possible to use other citation databases. The Web of Science is a Web-based user interface to ISI's citation databases. The coverage of the version used for this study ranged from 1981 to the present date. Our users can search the Web of Science with the topic terms identified, such as "mass extinction" and "BSE or CJD." The result of search may contain hundreds or even thousands of bibliographic records to be analyzed subsequently.

The bibliographic records from the Web of Science contain two important types of information: (1) citation—how frequently a publication has been cited, and (2) cocitation—how often a pair of publications has been cited side by side in later publications. Then we follow the procedure described in (Chen & Paul, 2001). In both case studies, only those publications that have been cited at least 10 times since 1981 can enter the next stage. This is to ensure that the subsequent analysis is based on a set of well-recognized publications. Admittedly, this may exclude some potentially significant ones before they reach the threshold, but this problem can be resolved by running such studies periodically. Our users can adjust the level of the threshold accordingly.

We use classic factor analysis, namely Principle Component Analysis (PCA), to divide selected publications, the cream of the crop, into specialties. Instead of applying PCA directly on the raw cocitation data, we follow the procedure described in White and McCain's widely cited study (White & McCain, 1998) and apply it to the correlation matrix of cocitations. Each factor represents a specialty. Pathfinder network scaling then derives a structural model—a cocitation network of publications. The structural model and the speciality memberships are superimposed on each other to achieve the greatest degree of clarity. The citation impact of a publication is depicted as a color-mapped stacked bar, and the height of each segment of the bar corresponds to the total citations of a particular year. Another threshold is available for users to control the minimum citations needed for a semitransparent label displayed on top of a citation bar. The color of each year's citation bar indicates the recentness of citations. Identifying a landmark in such knowledge landscape becomes a simple task: a tall citation bar with a large amount of segments in bright color is likely to be a landmark article in the given knowledge domain. In our approach, the membership of speciality, sometimes also known as a subdomain, or a theme, is colored according to the factor loadings of the three largest factors identified. This enables users to identify landmark publications easily.

Tracking the growth of competing paradigms in our approach relies on the visualization and animation of the dynamics of clusters of landmark publications over time. There are three options to depict the growth of the virtual landscape in terms of the stability of the base map: a fully dynamic base map mode, a partially dynamic base map mode, or a static base map mode. The key question is which articles and cocitation relationships are used to determine the layout of a base map in each time interval. In a fully dynamic base map mode, only articles that are cited in a specific time interval are used, and their cocitation relationships are used to determine the layout of this particular base map. Base maps across different time intervals tend to have different topology. As found in the literature of bibliometrics as well as our own previous studies, such a design tends to give the viewer a relatively high cognitive load because one has to compare different shaped base maps across different time intervals. In a partially dynamic base map mode, all articles are used to determine the layout of a universal base map, but articles are only shown in time intervals when they are cited. This is potentially a good design option. We may investigate this option in our future studies. In a static base map mode, all articles are used to determine the layout of a universal base map, and all articles are shown all the time.

In this study, we choose to depict the cocitation network as a static base map so that the viewer can concentrate on the growth of citation rates and trace the changes of clusters in a stable reference framework. This is consistent with the design guidelines used for designing thematic maps of geographic information. The more intact a mental image, the more easily users can find structural relationships between landmarks. In our visualization, while the cocitation network remains static, the citation bars grow year by year in corresponding to their actual citations. By comparing the growth rates of clusters of landmark publications in the landscape users have an intuitive means of forming a mental image—the big picture of competing paradigms. One can spot a rapidly growing area as well as a long established area. A once highly cited cluster of landmark publications can be replaced by rapidly growing new clusters within a relatively short period of time, indicating a rapid change of research focus. We hope scientists will benefit from the provision of an alternative view of their field on a regular basis. In the following two case studies, we highlight structural and feature information associated with debates between competing paradigms.

Case Study I: The Mass Extinction Debates

Five mass extinctions have occurred in the past 500 million years on earth, including the greatest ever Permian-Triassic extinction 248 million years ago, and the Cretaceous-Tertiary extinction 65 million years ago, which wiped out the dinosaurs, among many other species. The Cretaceous-Tertiary extinction, also known as the KT extinction, has been the focus of intensive debates over the last 20 years, involving over 80 theories of what may have caused the mass extinction of dinosaurs. Paleontologists, geologists, physicists, astronomers, nuclear chemists, and many others are all involved. And the debate is not completely over.

Competing Paradigms

Throughout the 1980s the KT debate was largely between the impact paradigm and the volcanism paradigm. According to the impact paradigm, the KT extinction was the result of an asteroid or meteorite, but the volcanism paradigm insists that it was due to a massive volcanic eruption (Alvarez, 1997).

Impact Theory

In 1980, Luis Alvarez, a Nobel-prize winning physicist, and his colleagues, a geologist and two nuclear chemists, came up with an impact theory to establish a connection between a layer of iridium found in rocks and the KT extinction. A layer of iridium sediment in rocks was found in many places around the world, and the iridium layer was dated the same time of the KT extinction—65 million years ago. Iridium can be found in the earth's mantle and in extraterrestrial objects such as meteors and comets. Scientists could not find other layers of iridium like this above or below the KT boundary. The impact theory proposes that an asteroid or a comet smashed into the earth and caused the mass extinction.

The impact theory is the prime representative of the so-called catastrophism. Gradualism is a view held by many paleontologists that the mass extinction took place over a span of probably more than one million years. In 1982, Phil Signor and Jere Lipps demonstrated that for a truly abrupt extinction might appear to be a gradual one if the fossil record was rare. This is now known as the Signor-Lipps effect, which is a landmark for the debate. The impact paradigm also led to an extensive search for an impact crater. A breakthrough came in 1991 when a growing body of evidence suggested that the Chicxulub crater on the Yucatan Peninsula in Mexico might well be the impact site.

Volcanism

The volcanism paradigm has a different explanation of where the iridium layer in the KT boundary came from. According to the volcanism paradigm, this iridium layer was due to a massive volcanic eruption instead of an asteroid. The volcanism paradigm was able to point to not only the Deccan Traps in India dated 65 million years ago, which coincided with the KT extinction, but also the Siberia Traps, dated 248 million years ago, which coincided with another mass extinction—the Permian-Triassic extinction. The huge amount of lava produced by such volcanic eruptions would cause intense climatic and oceanic change worldwide. Gradualists hypothesized that mass extinctions occurred gradually instead of catastrophically, and the volcanism theory would have led to such gradual extinction.

The Periodicity Paradigm

The periodicity paradigm is not in direct competition with the impact and volcanism paradigms. Instead, it poses additional challenges to the two paradigms. The periodicity hypothesis suggests that mass extinctions happen every 26 million years. If a paradigm can explain not only the KT extinction, but also other mass extinctions, then it would be regarded as a more powerful paradigm and is more likely to prevail. This is the explanatory-coherence criterion explained by Paul Thagard (Thagard, 1992). The pursuit of this type of extended coverage has led to some significant publications.

Visualizations of competing paradigms. Figure 1 shows an overview of a knowledge landscape, consisting of 273 publications above the 10-citation threshold based on a 20-year citation-sampling window ranging between 1981 and 2001. Citation profiles indicate four areas of interest. The original publication that first established the impact theory has the highest citation bar.

Figure 2 shows slightly enlarged images of the four areas so that some labels of landmark publications become more readable. Each landmark article must have at least 50 citations. Labels show the names of leading authors. Labels in the KT impact cluster include Hildebrand, Keller, Signor, and Smit. The label of Alvarez is well beyond the ceiling of this box. Labels in the volcanism cluster include Officer and Hup. Labels in the periodicity cluster include Raup, Alvarez, and Davis. Finally, labels in the Permian-Triassic extinction cluster include Erwin, Knoll, and Wignall. We examined major landmark publications in each cluster to understand their role in scientific debates between different paradigms. The details of this analysis should be read along with references listed in Table 1.

In his recent book (Alvarez, 1997), Walter Alvarez, a central figure in the debate, shows high regard to Smit's contribution to the impact theory: Alvarez found the iridium abnormally in Italy, whereas Smit confirmed the iridium abnormally in Spain. Smit's 1980 article in *Nature* is at the top of the list because it has the strongest factor loading. In the cocitation network, Smit's article is located immediately next to the 1980 *Science* paper by Alvarez et al. Both articles are strongly connected via a Pathfinder network link. The KT impact cluster also includes Glen's 1994 book "Mass Extinction Debates," which could be an additional

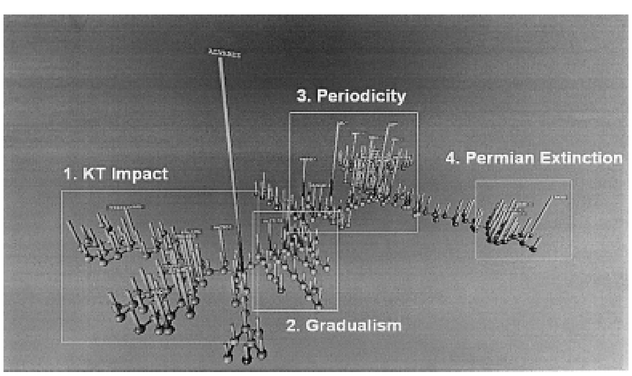

FIG. 1. A paradigmatic visualization of the mass extinction debates (1981–2001). The visualization highlights four thematic areas. Each corresponds to a loosely defined paradigm. The mainstream theme is colored in red, the secondary one is in green, and the third one is in blue.

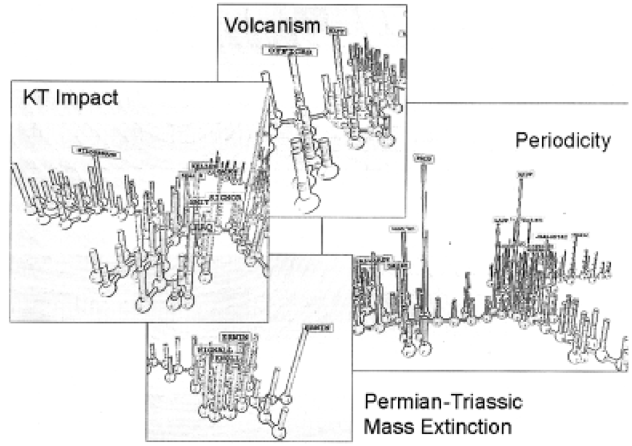

FIG. 2. Landmark articles are those that have 50 or more citations in four areas. They are labeled with the names of leading authors.

source of reference for verifying the analysis if Walter Alvarez's more recent book were not available.

The KT Impact Paradigm

Figure 2 shows articles in the KT Impact cluster. This cluster also corresponds to the most predominant specialty of the mass extinction research, and this is the home of the most server mass extinction debates in the 1980s. The red color shared by the majority of articles in this cluster indicates that they represent the mainstream of the research.

The article with the highest citations in the entire thematic map was the one by Luis Alvarez, Walter Alvarez, Frank Asaro, and Helen Michel, published in *Science* in 1980 (Table 1: ME-A4). It was this article that formally established the possible connection between the iridium anomaly found at the KT boundary and the KT mass extinction.

Scientists found a layer of iridium in deep-sea limestones exposed in Italy, Denmark, and New Zealand. The amount of iridium was found to be about 30, 160, and 20 times, respectively, above the background level at precisely the time of the Cretaceous-Tertiary extinctions, 65 million years ago.

If the impact theory is correct, then the impact should have left a crater on the earth. Alvarez et al. estimated the size of the crater must be between 150–200 km in diameter, but in 1980, only three craters were known to be 100 km or more in diameter. Among the three candidates, Sudbury and Vrdefort were dated to the Precambrian age, which was too old for the KT impact, and Popigay Crater in Siberia, dated only at 28.8 million years old, would be too young for the KT impact. In their initial impact theory article in 1980 they suggested that there was a two-thirds probability that the impacting object fell in the ocean. However, because a substantial portion of the Pre-tertiary Ocean no longer exists, we may never be able to find the crater.

Searching for the impact crater had become a Holy Grail for the supporters of the impact paradigm. A breakthrough

TABLE 1. Landmark articles in mass extinction research (ME): the KT impact cluster (ME-A), the Periodicity cluster (ME-B), and the Permian extinction cluster (ME-C).

ME	Factor loading	Source	Landmark (>50)
		KT impact	
ME-A1	0.964	Smit, J., & Hertogen, J. (1980). An extraterrestrial event at the Cretaceous-Tertiary boundary. Nature, 285, 198–200.	Highest loading
ME-A2	0.918	Hildebrand, A.R., Penfield, G.T., Kring, D.A., Pilkington, M., Carmargo, Z.A., Jacobsen, S.B., & Boynton, W.V. (1991). Chicxulub crater: A possible Cretaceous-Tertiary boundary impact crater on the Yucatan Peninsula, Mexico. Geology, 19(9), 867–871.	*
ME-A3	0.917	Keller, G. (1993). Is there evidence for Cretaceous-Tertiary boundary age deep-water deposits in the Caribbean and Gulf of Mexico. Geology, 21(9), 776–780.	*
ME-A4	0.877	Alvarez, L.W., Alvarez, W., Asaro, F., & Michel, H.V. (1980). Extraterrestrial cause for the Cretaceous-Tertiary extinction. Science, 208(4448), 1095–1098.	*
ME-A5		Signor, P.W., & Lipps, J.H. (1982). Sampling bias, gradual extinction patterns, and catastrophes in the fossil record. Geological Society of America Special Paper, 190, 291–296.	
		Periodicity	
ME-B1	0.898	Patterson, C., & Smith, A.B. (1987). Is the periodicity of extinctions a taxonomic artifact? Nature, 330(6145), 248–251.	Highest loading
ME-B2	0.873	Raup, D.M., & Sepkoski, J.J. (1986). Periodic extinction of families and genera. Science, 231(4740), 833–836.	*
ME-B3	0.859	Raup, D.M., & Sepkoksi, J.J. (1984). Periodicity of extinctions in the geologic past. Proceedings of the National Academy of Sciences of the United States of America—Biological Sciences, 81(3), 801–805.	*
ME-B4	0.720	Jablonski, D. (1986). Causes and consequences of mass extinctions: A comparative approach. In: D.K. Elliott (Ed.), Dynamics of extinction (pp. 183–230). New York: Wiley Interscience.	*
ME-B5	0.679	Benton, M.J. (1985). Mass extinction among non-marine tetrapods. Nature, 316(6031), 811–814.	*
ME-B6	0.629	Davis, M., Hut, P., & Muller, R.A. (1984). Extinction of species by periodic comet showers. Nature 308, 715–717.	*
ME-B7	0.608	Jablonski, D. (1986). Background and mass extinctions: The alternation of macroevolutionary regimes. Science, 231(4734), 129–133.	*
		Permian extinction	
ME-C1	0.812	Magaritz, M. (1989). ^{13}C minima follow extinction events: A clue to faunal radiation. Geology, 17, 337–340.	Highest loading
ME-C2	0.444	Renne, P., Zhang, Z., Richards, M.A., Black, M.T., & Basu, A. (1995). Synchrony and causal relations between Permian-Triassic boundary crises and Siberian flood volcanism. Science, 269, 1413–1416.	*
ME-C3	0.436	Stanley, S.M., & Yang, X. (1994). A double mass extinction at the end of the paleozoic era. Science, 266, 1340–1344.	*
ME-C4	0.426	Erwin, D.H. (1994). The Permo-Triassic extinction. Nature, 367, 231–236.	*
ME-C5	0.426	Wignall, P.B. (1996). Oceanic anoxia and the end Permian mass extinction. Science, 272, 1155.	*
ME-C6		Knoll, A.H., Bambach, R.K., Canfield, D.E., & Grotzinger, J.P. (1996). Comparative earth history and Late Permian mass extinction. Science, 273, 452–457.	*

came in 1991 when Alan Hildebrand made the link between the Chicxulub crater and the KT impact. The Chicxulub crater is 180 km in diameter and completely buried under the Yucatan Peninsula in Mexico. As early as 1950s, a Mexican petroleum company searched for oil fields and the gravity abnormality suggested that the Chicxulub crater could be rich in oil. They did not find oil, and the Chicxulub crater did not make it into the scientific literature. Hildebrand's 1991 paper (Table 1: ME-A2) became one of the most highly cited landmark articles of the KT impact paradigm.

According to Walter Alvarez (Alvarez, 1997), Keller was regarded as the number one opponent of the impact theory. A number of Keller's papers appear in the KT impact cluster, including the 1993 paper, which challenged the available evidence of impact-generated tsunami deposits (Table 1: ME-A3). This pattern seems to suggest that we should pay particular attention to articles from both competing paradigms. In this case, Keller's 1993 paper that challenged the impact paradigm as well as Hildebrand's 1991 paper on the Chicxulub crater, which strongly supported the impact paradigm, are found in the same cluster.

In Figure 2 the KT impact cluster shows an article labeled as Signor (Table 1: ME-A5). This is an article by Signor and Lipps on what became known as the Signor-Lipps effect. In essence, the Signor-Lipps effect says if few fossils are preserved, an abrupt distinction can appear gradual. Whether the KT mass extinction was gradual or catastrophic in nature is central to the KT impact debate. The high profile of Signor and Lipps' article indicates its key role in this debate.

The Volcanism Paradigm

The cluster located between the KT impact cluster and the periodicity cluster represents the home of gradualism; in

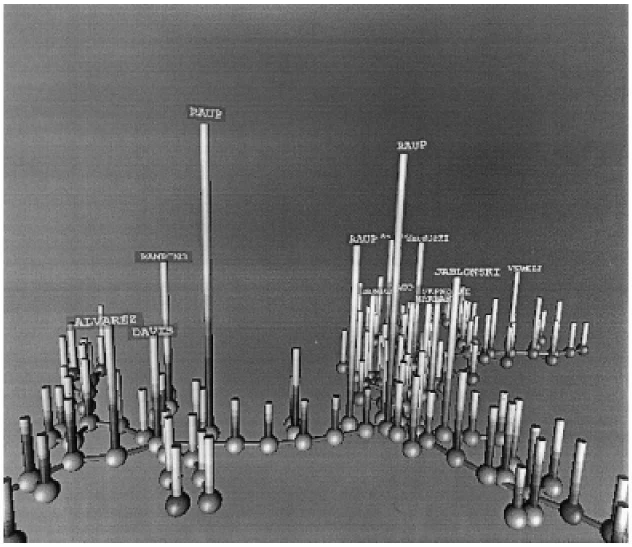

FIG. 3. The periodicity cluster.

particular, it shows volcanism as a competing paradigm to the impact paradigm. Landmark articles in this cluster include ones from Chunk Officer, a key opponent of the impact theory. McLean's article is also in this cluster, but below the landmark threshold of 50 citations. McLean was the first to propose that prolonged volcanic eruptions, known as the Deccan Traps in India, were the cause of the KT mass extinction.

Hut's 1987 *Nature* article, with coauthors such as Alvarez and Keller, proposed that comet showers are a cause of mass extinctions. This theme marks a transition point between the KT impact paradigm and the periodicity hypothesis, underlining the role of this article in seeking an explanation of the periodicity of mass extinctions within the impact paradigm.

The Periodicity Paradigm

The second largest area in the thematic landscape highlights the periodicity paradigm. The periodicity frame in Figure 3 shows two predominant landmarks, both from David Raup and John Sepkoski. The one on the left is their 1984 article published in the *Proceedings of the National Academy of Sciences of the United States of America—Biological Sciences,* entitled *Periodicity of extinctions in the geologic past.* They showed a graph of incidences of extinction of marine families through time, in which peaks tend to coincide with the time of most major extinction events, and suggested that evidence shows that mass extinctions tend to occur every 26 million years. The one on the right is their 1982 article in *Science,* entitled *Mass extinctions in the marine fossil record.*

Patterson and Smith's 1987 article in *Nature* (Table 1: ME-B1) questioned whether the periodicity really existed, and this article has the highest factor loading (0.898), indicating its unique position in this paradigm. In response to the periodicity paradigm, the impact paradigm came up with a hypothesis that an invisible death star is causing the periodical mass extinction, but the hypothesis is still essentially theoretical. Landmark articles labeled as Alvarez and Davis in the scene are notable examples of research in this direction.

The Permian-Triassic Mass Extinction Paradigm

The Permian-Triassic (PT) mass extinction was much more severe than the KT extinction. Because it happened 248 million years ago, it is extremely hard to find evidence, especially the kind of evidence required for an impact theory. Landmark articles in this cluster are from Erwin, Wignall, and Knoll. Erwin is the leading scientist on the Permian mass extinction. Erwin's 1994 article in *Nature* (Table 1: ME-C4) is among the most highly cited articles in the Permian cluster. He listed causes such as intense climatic, tectonic, and environmental change.

The 1996 article in *Science* by Knoll et al. (Table 1: ME-C6) suggested that the overturn of anoxic deep oceans during the Late Permian introduced high concentrations of carbon dioxide into surficial environments. Wignall's 1996 *Science* article (Table 1: ME-C5) is on a similar topic, suggesting anoxic oceans may have caused the Permian extinction.

Just below the 50-citation threshold, the PT cluster includes the 1995 *Science* article by Paul Renne and his colleagues (Table 1: ME-C2), in which he suggested that the Siberian plume changed the environment and climate, which in turn, led to the Permian mass extinction.

The dynamics of paradigms. Figure 4 shows a few frames from a year-by-year routinely generated animation of the growing impact of articles associated with different paradigms. The skyline of citations articles indicates that the volcanism paradigm is one of the pioneering ones in the study of mass extinctions, and that the KT impact became the most prevalent paradigm much more recently.

Discussion

In this case study, our citation-based visualization approach has demonstrated its potential for tracking the development of competing paradigms. The animated growth of citations of landmark articles helps us to focus on significant movements of landmark articles in a broad context of the mass extinction debate.

To verify the major paradigms identified in this case study, we consulted a book written by Walter Alvarez, one of the leading figures in the mass extinction research, especially in the impact paradigm (Alvarez, 1997). We were able to find a substantial level of consistency between the book and the visualization, especially for the KT impact paradigm.

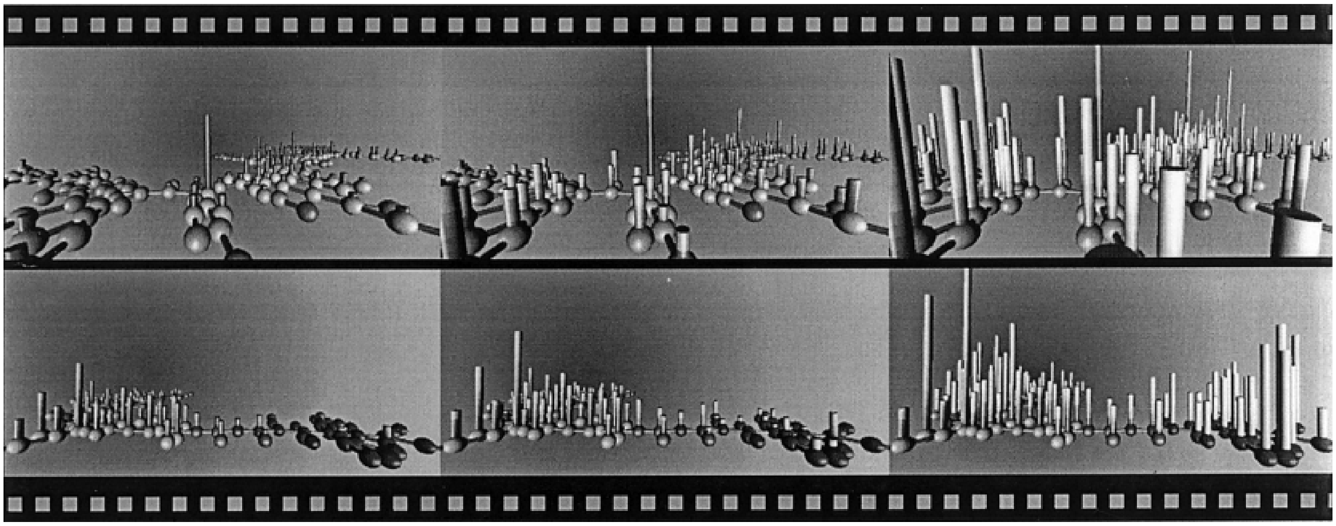

FIG. 4. A year-by-year animation shows the growing impact of articles in the context of relevant paradigms. The top-row snapshots show the citations gained by the KT impact articles (center), whereas the bottom-row snapshots highlight the periodicity cluster (left) and the Permian extinction cluster (right).

Henry Small, in his longitudinal study of collagen research, included a questionnaire-based validation process (Small, 1977). He sent questionnaires to researchers in the field and asked them to describe major rapid changes of focus in the subject domain. In this article, because we have chosen the mass extinction case, we can benefit from the availability of authoritative information such as Alvarez's book and verify the case directly. This also helps us to better understand what our visualization approach entails and what insights it can possibly offer to scientists in the field as well as new researchers to the field.

Case Study II: Mad Cow Disease (BSE) and vCJD

The 1997 Nobel Prize in physiology or medicine was awarded to Stanley Prusiner, professor of neurology, virology, and biochemistry, for his discovery of prions—a type of rogue protein. Prusiner introduced the prion theory in his 1982 article published in *Science.* He suggested that an abnormal form of a protein is responsible for diseases such as scrapie in sheep, Bovine Spongiform Encephalopathy (BSE) in cattle—the mad cow disease, and Creutzfeldt-Jakob disease (CJD) in humans. The prion theory is widely accepted now, but some crucial questions still remain unanswered, for example, the lack of definitive evidence that prions can actually cause disease by themselves.

Mad Cow Disease

The mad cow disease was first identified in Britain in 1986. A sponge-like malformation can be found within the brains of infected cattle. The BSE epidemic in Britain reached its peak in 1992, and has since steadily declined. During the outbreak, much of the "national herd" had to be slaughtered, and financial claims were massive. The British Ministry of Agriculture, Fisheries, and Food (MAFF) provides an authoritative source of information on BSE.[1] The incidence of BSE in other countries is available on the Office International des Epizooties (OIE) Web site.[2]

vCJD

Two German neurologists were the first to discover Creutzfeldt-Jakob disease (CJD) in the 1920s. It is the principal form of a number of human transmissible encephalopathy (TSE) diseases. BSE in cattle, and scrapie in sheep, are examples of animal TSE diseases. Several diseases are related to the prion protein in different species; in humans, the prion-based disease is related to CJD, Kuru (transmitted by cannibalism), Gerstmann-Sträussler-Scheinker Disease (GSS), and Fatal Familial Insomnia (FFI).

Variant CJD (vCJD) is an unrecognized variant of CJD discovered by the National CJD Surveillance Unit in Edinburgh. This discovery was first reported in *The Lancet* on 6 April 1996. vCJD is characterized clinically by a progressive neuropsychiatric disorder. Neuropathology shows marked spongiform change throughout the brain. There is growing concern in the general public that BSE may have passed from cattle to humans. The British government assured the public that beef was safe until, in 1996, it announced a possible link between BSE and CJD.

Visualization of the Possible Link

Figure 5 shows an overview of a thematic landscape of scientific literature on the topic "BSE or CJD." Five thematic areas are labeled in the overview: the prion paradigm, BSE, CJD, vCJD, and GSS. As usual, the colors correspond

[1] http://www.maff.gov.uk/animalh/bse/index.html

[2] http://www.oie.int/eng/info/en_esbmonde.htm

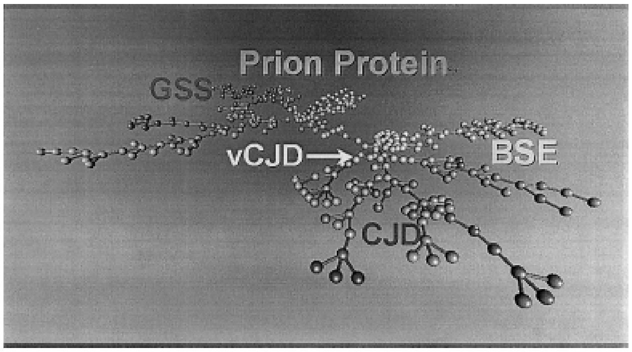

FIG. 5. The network of 379 articles identifies five major thematic areas.

to the leading factors identified by factor analysis in red, green, and blue, respectively (Table 2).

The Prion Paradigm

The prion paradigm includes Prusiner's 1982 *Science* article and Gajdusek's 1966 article. Kuru is a brain disease that affected highlanders in New Guinea who practiced ritualized cannibalism. D. Carleton Gajdusek's 1966 article demonstrated that kuru was infectious. His work eventually won him the 1976 Nobel Prize. Figure 7 shows Prusiner's article as the fourth highly cited articles during the period of 1995–2001. Gajdusek's 1966 *Nature* article is located immediately next to Prusiner's article, but its citations were below the labeling threshold.

This paradigm also includes Alper's 1967 article in *Nature.* Alper studied scrapie in sheep and found that brain tissue remained infectious even after she subjected it to radiation that would destroy any DNA or RNA. Griffith suggested in a separate article that perhaps a protein could somehow misfold and then catalyze other proteins to do so. Such an idea seemed to threaten the very foundations of molecular biology, which held that nucleic acids were the only way to transmit information from one generation to the next. Prusiner's theory was built on an idea proposed in the 1960s. He coined the term prion in 1982 to describe the "proteinaceous infectious particles" he blamed for causing scrapie in sheep and hamsters. He suggested that a variety of

TABLE 2. Landmark articles on BSE and CJD (BC): prion protein (BC-A), BSE (BC-B), and vCJD (BC-X) (Landmark thresholds of 600 and 155, respectively).

BC	Factor loading	Source	Landmark
		Prion	>600
BC-A1	0.981	Prusiner, S.B. (1982). Novel proteinaceous infectious particles cause scrapie. Science, 216(4542), 136–144.	*
BC-A2	0.935	Oesch, B., Westaway, D., Walchli, M., McKinley, M.P., Kent, S.B.H., Aebersold, R., Barry, R.A., Tempst, P., Teplow, D.B., Hood, L.E., Prusiner, S.B., & Weissmann, C. (1985). A cellular gene encodes scrapie prp 27–30 protein. Cell, 40(4), 735–746.	*
BC-A3	0.914	Prusiner, S.B. (1991). Molecular biology of prion diseases. Science, 252(5012), 1515–1522.	*
BC-A4	0.924	Alper, T., Cramp, W.A., Haig, D.A., & Clarke, M.C. (1967). Does the agent of scrapie replicate without nucleic acid? Nature, 214, 764–766.	
BC-A5	0.914	Gajdusek, D.C., Gibbs, C.J., & Alpers, M. (1966). Experimental transmission of a kuru-like syndrome to chimpanzees. Nature, 209, 794–796.	
		BSE	>155
BC-B1	0.825	Wells, G.A.H., Scott, A.C., Johnson, C.T., Gunning, R.F., Hancock, R.D., Jeffrey, M., Dawson, M., & Bradley, R. (1987). A novel progressive spongiform encephalopathy in cattle. Veterinary Record, 121(18), 419–420.	*
BC-B2	0.838	Anderson, R.M., Donnelly, C.A., Ferguson, N.M., Woolhouse, M.E.J., Watt, C.J., Udy, H.J., MaWhinney, S., Dunstan, S.P., Southwood, T.R.E., Wilesmith, J.W., Ryan, J.B.M., Hoinville, L.J., Hillerton, J.E., Austin, A.R., & Wells, G.A.H. (1996). Transmission dynamics and epidemiology of BSE in British cattle. Nature, 382(6594), 779–788.	*
BC-B3	0.855	Donnelly, C.A., Ghani, A.C., Ferguson, N.M., Wilesmith, J.W., & Anderson, R.M. (1997). Analysis of the bovine spongiform encephalopathy maternal cohort study: Evidence for direct maternal transmission. Applied Statistics-Journal of the Royal Statistical Society Series C, 46(3), 321–344.	*
BC-B4	0.823	Hadlow, W.J., Kennedy, R.C., & Race, R.E. (1982). Natural infection of Suffolk sheep with scrapie virus. Journal of Infectious Diseases, 146(5), 657–664.	*
BC-B5	0.887	Kimberlin, R.H., & Walker, C.A. (1988). Pathogenesis of experimental scrapie. Ciba Foundation Symposia, 135, 37–62.	*
		vCJD	
BC-X1		Will, R.G., Ironside, J.W., Zeidler, M., Cousens, S.N., Estibeiro, K., Alperovitch, A., Poser, S., Pocchiari, M., Hofman, A., & Smith, P.G. (1996). A new variant of Creutzfeldt-Jakob disease in the UK. Lancet, 347, 921–925.	
BC-X2		Collinge, J., Sidle, K., Meads, J., Ironside, J., & Hill, A. (1996). Molecular analysis of prion strain variation and the aetiology of "new variant" CJD. Nature, 383, 685–691.	
BC-X3		Bruce, M.E., Will, R.G., Ironside, J.W., McConnell, I., Drummond, D., Suttie, A., McCardle, L., Chree, A., Hope, J., Birkett, C., Cousens, S., Fraser, H., & Bostock, C.J. (1997). Transmissions to mice indicate that "new variant" CJD is caused by the BSE agent. Nature, 389(6650), 498–501.	

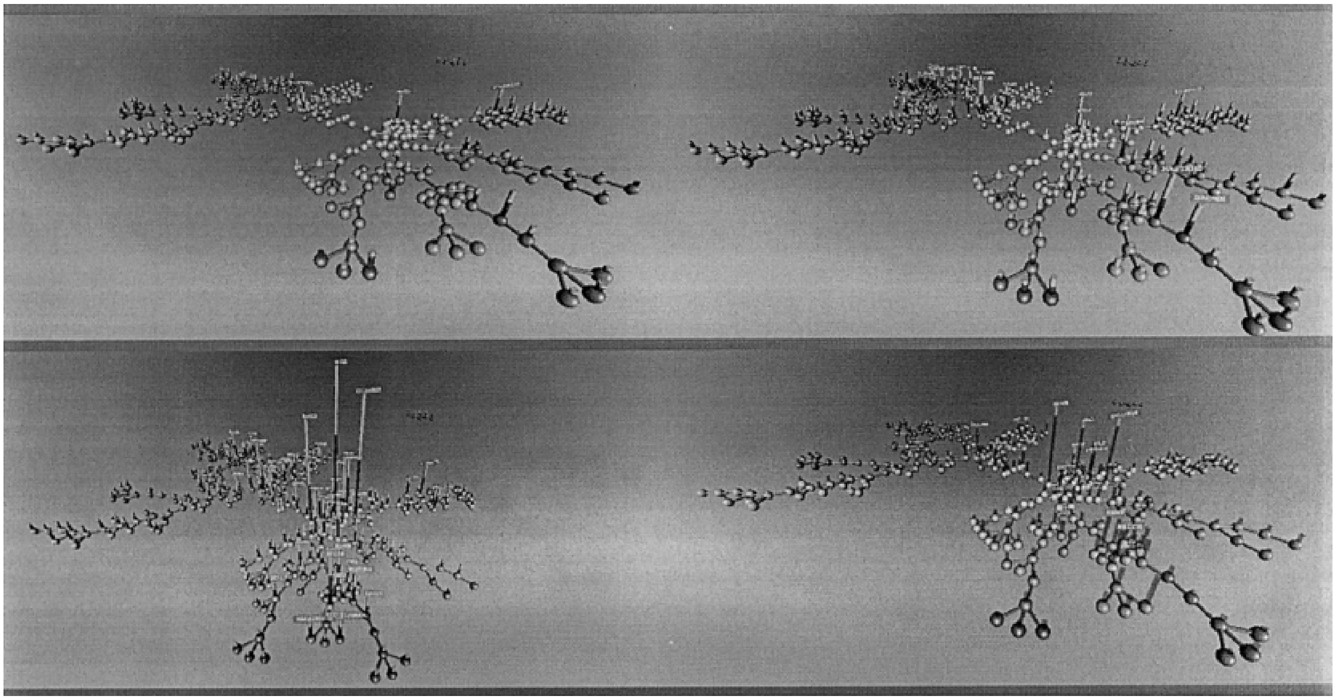

FIG. 6. A year-by-year animation shows the growing impact of research in the connections between BSE and CJD. Top left: 1991–1993; Top right: 1994–1996; Bottom left: 1997–1999; Bottom right: 2000–2001.

brain diseases were all due to a common process: a misfolded protein that propagates and kills brain cells.

Prusiner and his colleagues reported in *Science* in 1982 that they had found an unusual protein in the brains of scrapie-infected hamsters that did not seem to be present in healthy animals. Their article, entitled "Novel proteinaceous infectious particles cause scrapie," had been cited 941 times by March 2001. Taken together, the results have convinced many scientists that the protein is indeed the agent behind CJD, scrapie, mad cow disease, and others.

Snapshots in Figure 6 show the evolution of research in BSE and CJD. The fastest growth of citation impact is in the central area where studies that focused on the possible link between BSE and CJD are located. In addition, the visualization presents a lot of contextual information about our immediate topic. For example, the prion paradigm is associated with areas that are colored in red, suggesting that the paradigm represents the mainstream of the BSE and CJD research. Our domain visualization has highlighted the fundamental role of the prion paradigm in explaining brain diseases in animals and humans, including BSE and CJD.

Figure 7 highlights four highly cited publications during the period of 1995–2001. The first three articles are all about new variant CJD, which is the type of CJD that may turn out to be a link between BSE and humans.

The BSE and CJD case study has shown that a citation-based approach enables users to see not only the results that are immediately related to the initial search, but also the wider context in which the current topic has been evolving. Cocitations are probably the primary force that introduces such contextual information into the picture. Users can explore the structure of the relevant knowledge landscape and decide to which area they should direct their attention for future development.

Discussion and Conclusion

We have included two case studies that demonstrate the potential of information visualization as a means of tracking the development of scientific debates and the movement of paradigms. Our approach affords tasks such as the identification and tracking of landmark articles in a knowledge domain and comparing the growth rates between landmark articles.

In conclusion, through the two case studies, we have identified the following issues to be addressed by approaches to visualizing the growth of competing paradigms. Addressing these issues will help us to clarify further what tasks are required for scientists to make use of such facilities.

Domain-Specific versus Domain-Independent

This issue is concerned with how much domain knowledge will be required to carry out an analysis. Our approach provides some flexibility in this area. As long as users can issue a meaningful query to the system, it will visualize the relevant topic in a wider context. In case study I, we simply used the query "mass extinction," and in case study II, we used the query "BSE and CJD" to gather bibliographic data from the Web of Science, ISI's Web-based citation database.

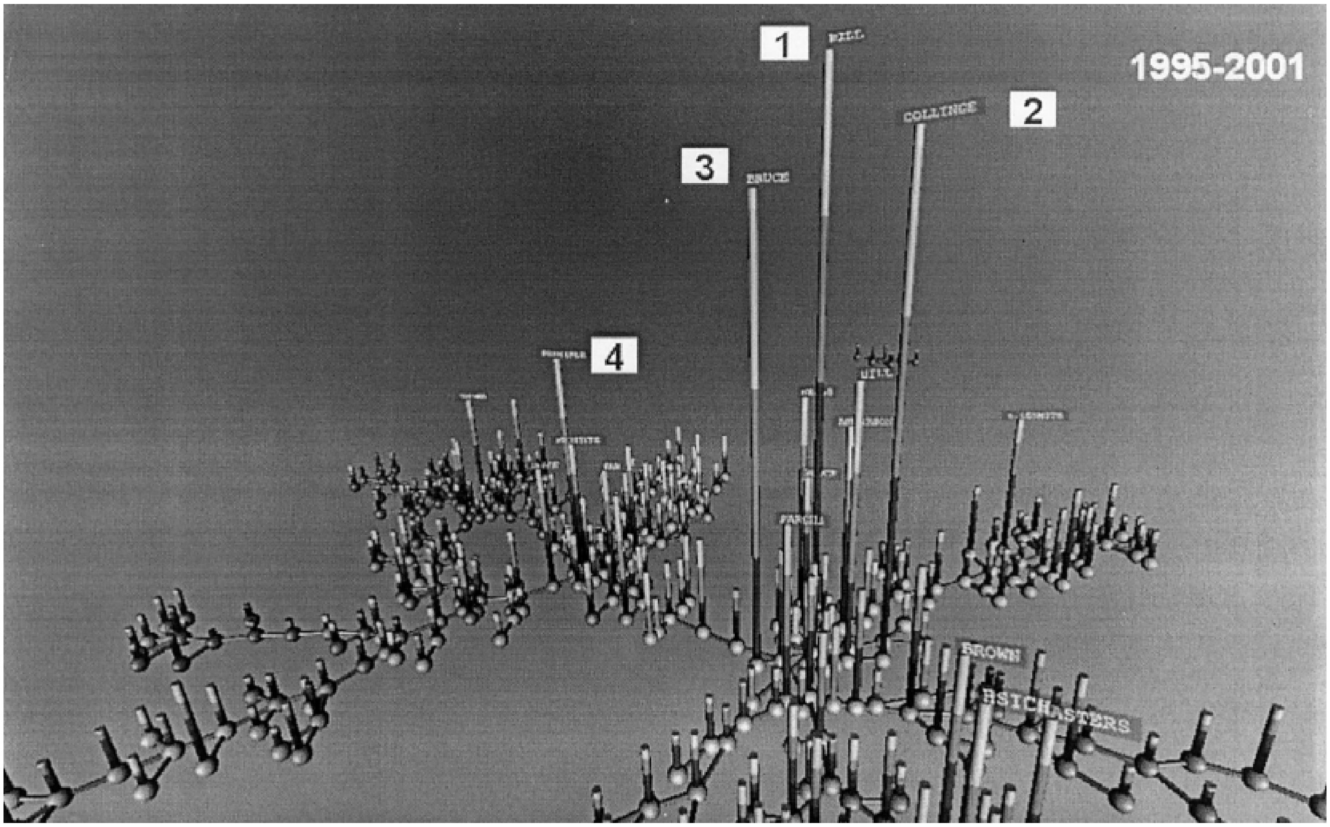

FIG. 7. Landmark articles in the period of 1995–2001. Articles labeled 1–3 directly address the BSE-CJD connection. Article 4 is Prusiner's original article on prion, which has broad implications on brain diseases in sheep, cattle, and humans.

Quality versus Timeliness

The quality comes from the collective views expressed by domain experts in their scholarly publications. The timeliness issue arises from the reality that by the time an article appears in print, it is more likely that science has moved on. Nevertheless, the history of scientific debates can provide valuable insights. If the analysis can be done frequently, such visualizations can provide useful milestones for scientists to project the trajectory of a paradigm. This issue also relates to the source of input, ranging from traditional scientific literature, gray literature such as technical reports and preprints, to communications between scientists of an invisible college.

Interdisciplinary Nature

To understand and interpret what is happening in science, we must consider the practice of closely related disciplines, particularly history of science, philosophy of science, and sociology of science as well as the scientific domain itself. This challenging issue requires an interdisciplinary approach to ensure that we are aware of the latest developments in these disciplines and integrate theoretical and methodological components properly.

Validation

This is an integral part of the process. It is crucial to develop a comprehensive understanding of the strengths and weaknesses in applying this type of approach to a wider range of case studies. By maintaining a focused visualization and exploration process, it may also be informative for the development of the information visualization field as well as for the particular scientific debates studied.

In conclusion, visualizing and tracking the growth of scientific knowledge presents a unique opportunity and a challenge to the information visualization community. We must draw challenging tasks from a diverse range of disciplines to pump up the momentum of information visualization as an independent intellectual field.

References

Alvarez, W. (1997). T. Rex and the crater of doom. New York: Vintage Books.

Benford, S., Snowdown, D., Greenhalgh, C., Ingram, R., Knox, I., & Brown, C. (1995). VR-VIBE: A virtual environment for co-operative information retrieval. Computer Graphics Forum, 14(3), C.349–C.360.

Braam, R.R., Moed, H.F., & van Raan, A.F.J. (1991). Mapping of science by combined co-citation and word analysis, I: Structural aspects and II: Dynamical aspects. Journal of the American Society for Information Science, 42, 233–266.

Callon, M., Courtial, J.P., Turner, W.A., & Bauin, S. (1983). From translation to network: The co-word analysis. Scientometrics, 5(1), 78–78.

Callon, M., Law, J., & Rip, A. (Eds.). (1986). Mapping the dynamics of science and technology: Sociology of science in the real world. London: Macmillan Press.

Card, S., Mackinlay, J., & Shneiderman, B. (Eds.). (1999). Readings in information visualization: Using vision to think. San Mateo, CA: Morgan Kaufmann.

Chalmers, M. (1992). BEAD: Explorations in information visualisation. Proceedings of SIGIR '92, June 1992, Copenhagen, Denmark (pp. 330–337).
Chen, C. (1999a). Information visualisation and virtual environments. London: Springer-Verlag London.
Chen, C. (1999b). Visualising semantic spaces and author co-citation networks in digital libraries. Information Processing and Management, 35(2), 401–420.
Chen, C. (2002). Mapping scientific frontiers: The quest for knowledge visualization. London: Springer-Verlag London.
Chen, C., & Paul, R.J. (2001). Visualizing a knowledge domain's intellectual structure. Computer, 34(3), 65–71.
Chen, C., Kuljis, J., & Paul, R.J. (2001). Visualizing latent domain knowledge. IEEE Transactions on Systems, Man, and Cybernetics: Part C. Applications and Reviews, 31(4), 518–529.
Chen, C., Paul, R.J., & O'Keefe, B. (2001). Fitting the jigsaw of citation: Information visualization in domain analysis. Journal of the American Society for Information Science, 52(4), 315–330.
Kuhn, T.S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.
Noyons, E.C.M., & van Raan, A.F.J. (1998). Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. Journal of the American Society for Information Science, 49(1), 68–81.
Olsen, K.A., Korfhage, R.R., & Sochats, K.M. (1993). Visualization of a document collection: The VIBE System. Information Processing & Management, 29(1), 69–81.
Rittschof, K.A., Stock, W.A., Kulhavy, R.W., Verdi, M.P., & Doran, J.M. (1994). Thematic maps improve memory for facts and inferences: A test of the stimulus order hypothesis. Contemporary Educational Psychology, 19, 129–142.
Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science, 24, 265–269.
Small, H. (1977). A co-citation model of a scientific speciality: A longitudinal study of collagen research. Social Studies of Science, 7, 139–166.
Small, H. (1997). Update on science mapping: Creating large document spaces. Scientometrics, 38(2), 275–293.
Small, H. (1999). Visualizing science by citation mapping. Journal of the American Society for Information Science, 50(9), 799–813.
Small, H., & Griffith, B. (1974). The structure of scientific literatures I: Identifying and graphing specialties. Science Studies, 4, 17–40.
Spence, B. (2001). Information visualization. Reading, MA: Addison-Wesley.
Thagard, P. (1992). Conceptual revolutions. Princeton, NJ: Princeton University Press.
White, H.D., & Griffith, B.C. (1981). Author co-citation: A literature measure of intellectual structure. Journal of the American Society for Information Science, 32, 163–172.
White, H.D., & McCain, K.W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. Journal of the American Society for Information Science, 49(4), 327–356.
Wise, J.A., Thomas, J.J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., & Crow, V. (1995). Visualizing the non-visual: Spatial analysis and interaction with information from text documents. Proceedings of IEEE Symposium on Information Visualization '95, October 30–31, 1995, Atlanta, GA.