The Rising Landscape: A Visual Exploration of Superstring Revolutions in Physics

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

Jasna Kuljis
Department of Information Systems and Computing, Brunel University, Uxbridge UB8 3PH, UK. E-mail: jasna.kuljis@brunel.ac.uk

Knowledge domain visualization is a visual exploratory approach to the study of the development of a knowledge domain. In this study, we focus on the practical issues concerning modeling and visualizing scientific revolutions. We study the growth patterns of specialties derived from citation and cocitation data on string theory in physics. Special attention is given to the two superstring revolutions since the 1980s. The superstring revolutions are visualized, animated, and analyzed using the general framework of Thomas Kuhn's structure of scientific revolutions. The implications of taking this approach are discussed.

Introduction

Capturing and understanding the dynamics of a scientific revolution is a daunting challenge due to the complexity, the diversity, and the scale of the contributing factors and relationships involved. Thomas Kuhn's structure of scientific revolutions represents an influential work in the field. His notion of paradigm shift has become widely known in a diverse range of scientific disciplines. However, its use is limited and time consuming because scientists, especially junior scientists, often have to sift through mountains of publications and other documents to digest and synthesize a wide variety of diverse views into a clear picture of a scientific revolution. One way to simplify this process is by constructing such pictures automatically based on some publicly accessible data.

A paradigm is a world view of underlying theories and methodologies that the members of a scientific community share (Kuhn, 1996). According to Kuhn, a scientific discipline evolves through a high-level state transition process,

namely: normal science → crisis → scientific revolutions → normal science. A normal science period is typically predominated by a stable paradigm. The core of the dominating paradigm is not seriously challenged in this period. In contrast, a crisis period is marked by fundamental challenges to the dominating paradigm due to anomalies that cannot be simply explained away within the existing paradigm. New paradigms of a greater explanation power begin to challenge the authoritative position of the dominating paradigm. A scientific revolution takes place when the current dominating paradigm gives way to a new paradigm.

In looking at paradigm shifts researchers have attempted to facilitate the process of understanding the difference between competing theories in terms of their conceptual structures. Paul Thagard (1992) suggests that explanation coherence holds the key for theory selection in science. A theory is more likely to gain acceptance if it has greater explanation coherence than its competitors. To study conceptual revolutions, Thagard designed a computational method that can compare the structural characteristics of conceptual systems before, during, and after conceptual revolutions. He conducted several case studies of conceptual revolutions using this method. Thagard's method is based on specific assumptions and major assertions about a theory. This strict dependency becomes a double-edged sword in practice. On the one hand, it tends to provide highly focused qualitative analysis of a conceptual structure. On the other hand, the overhead of conducting such studies is relatively high.

Knowledge domain visualization (KDV) provides an alternative route to the study of conceptual changes in scientific theories. KDV is concerned with the development of a scientific discipline. It emphasizes the importance of information visualization techniques and readily available data such as the citation databases compiled by the Institute for Scientific Information (ISI). Recently, an increasing number

of KDV studies began to emerge in a variety of knowledge domains, including information science (White & McCain, 1998), computer graphics (Chen, Paul, & O'Keefe, 2001), mass extinction (Chen, Cribbin, Macredie, & Morar, 2002), mad cow disease Chen, Kuljius, & Paul, 2001), and visual interfaces of digital libraries (Boyack, Wylie, & Dividson, 2002).

In this article, we focus on some of the practical issues concerning visualizing the dynamics of specialties in a scientific discipline. What are the key characteristics of scientific revolutions that should be featured in visualization models? To what extent do citation patterns track scientific revolutions? What are the growth patterns of scientific paradigms in terms of their citation impact and connectivity? In addition, we visualize the growth patterns of superstring revolutions in physics so as to illustrate the feasibility and viability of our approach.

The rest of this article is organized as follows. First, we introduce the context of our study. Second, we describe the key concepts and principles to be used. Third, we outline our methodology and analyze the growth patterns associated with the two superstring revolutions in physics. Finally, we discuss the significance and implications of our work.

Related Work

The following review of the literature focuses on studies based on structures derived from citation data. In the 1970s, Small and Griffith (1974) examined issues concerning with identifying and mapping specialties from the structure of scientific literature, especially based on cocitation patterns. Small (1977) also conducted a longitudinal study of collagen research and found that rapid changes of focus took place in collagen research. He grouped documents into clusters based on cocitation strengths between pairs of documents; these clusters represent leading specialties, or paradigms. Rapid shifts in research focus was found when a number of key documents abruptly disappeared from the leading cluster in 1 year and were replaced by a set of new documents in the following year. This is an important type of specialty change, which is an informative indicator of "revolutionary" changes. More recent studies in related areas include (Garfield, 1994, 1998; Small, 1999a, 1999b).

White and McCain (1998) mapped the intellectual structure of information science based on author cocitation patterns. Their work is also methodologically important; they demonstrate the power of author cocitation relationships as the underlying grouping principle. Chen (1999a, 1999b) incorporated Pathfinder network scaling into the methodology for author cocitation analysis and produced an author cocitation map of the hypertext community. Chen and Paul (2001) further extended the methodology to a generic approach to visualizing a knowledge domain's intellectual structure in a three-dimensional spatial model. Their approach produces visualization models of document cocitation networks as well as author cocitation networks. Colors, shapes, and animations are used in their models to display

salient structures of specialties in a given knowledge domain.

Until recently, researchers generated many valuable insights using snapshots of a domain image. Researchers began to explore the potential of information visualization and domain analysis techniques in understanding the growth patterns of specialties in a scientific discipline (Chen, 2002). For example, Chen et al. (2002) built animated visualization models of the literature of mass extinction and demonstrated the potential of a citation-based modeling and visualization approach.

Visualizing scientific revolutions has special requirements. Some of the requirements are likely to differ from what are readily available from existing approaches. Even if we narrow our options down to network visualizations, there are still a wide range of design decisions to be made. For example, static and dynamic networks could refer to vertices only, edges only, or both. Although the structure of a network changes in all these ways simultaneously in reality, one often chooses to work with a simplified model and gradually introduce more variables into the system. The following conceptual framework categorizes the major variables of interest in this study.

Conceptual Framework

A conceptual framework is developed for modeling and visualizing the growth patterns of scientific paradigms. The framework is based on the following assumptions: the structural model is based on a graph-theoretical network; the visualization is based on various representations of a network; and association relationships between each pair of data points are defined in the source data. The conceptual framework provides a reference point for a number of procedures that one can apply to the source data.

Procedures

There are two types of procedures: structural and nonstructural. Although structural procedures alter the structure of a network, nonstructural procedures leave the structure intact. The most commonly seen structural procedures are link-reduction algorithms, such as minimum-spanning-tree (MST) generation algorithms and Pathfinder network scaling (Schvaneveldt, 1990). These procedures can reduce the number of edges in a network. They are useful tools for simplifying a network representation. Pathfinder networks were used in our previous studies (Chen, 1999a, Chen & Paul, 2001).

A useful nonstructural procedure is the partitioning of a network. In the simplest case, vertices of a given network are divided into mutually exclusive groups. In more complex models, the partition may generate nonmutual exclusive groups. Partition is typically applied to the vertex domain of a network. Partition is particularly useful for identifying specialties in a scientific community. Chen and Paul (2001) partitioned Pathfinder networks into specialties
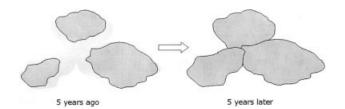
FIG. 1. The "continental drift" metaphor. A visualization of scientific paradigm highlights the movement of an evolving paradigm over time in an intellectual context.

of different colors using results of principle component analysis (PCA).

Patterns in a network are often directly determined by the structure of the network. For example, the connectivity of a network may in itself reveal interesting patterns. In timed networks, both the structure of a network and the various attributes of its vertices are variables of time. This type of network is particularly useful for modeling the growth patterns of scientific revolutions. An animation sequence of network visualizations enables users to recognize temporal patterns more easily than static snapshots. Animations have been used to show annual increases of citation counts in previous studies (Chen et al., 2002).

The focus of this study is on representing the dynamics of scientific revolution using a citation landscape of a knowledge domain. We also aim to enable a wide variety of users to use our tool. It is important for us to make it easy to use even without specialized domain knowledge. Users only need to know the top-level topic term, such as "mass extinction" or "string theory," to start the process.

A "Continental Drift" Metaphor

Our visual metaphor of scientific paradigms is inspired by Alfred Wegener's continental drift theory proposed in the early 20th century. We choose a retrospective approach to study the immediate past of science advancements. If this strategy works, then we can repeat it frequently so as to bring the big picture up to date.

Figure 1 is a sketch of our strategy. As scientific paradigms grow, clusters of articles would make them look like an intellectual continent. The course of an evolving paradigm can be tracked by a sequence of snapshots of the intellectual world over time. A very useful feature of Pathfinder networks is that the most significant and relevant work tends to be located in the center of its Pathfinder network representation, whereas other relevant but less predominant work appears at the outskirt of such networks.

In this study, we demonstrate the use of animation techniques to represent state transitions of vertices and edges in a Pathfinder network. Briefly speaking, we define a state space for vertices and a state space for edges. The vertex state space defines three states and associated transitions: prepublication → publication → citation. The edge state space also defines three states: not available for cocita-

tion → available but not cocited → cocited. State transitions are represented by decreased levels of transparency. We represent a document as a sphere and decrease the level of transparency of the sphere as the document moves up a state. The diagram in Figure 2 outlines the design.

String Theory

String theory is a vibrant topic in physics. In ScienceWatch®'s 1997 top-10 most cited physics paper list, four papers were about string theory (Mitton, 1997). Many believe that it has the great potential of unifying fundamental theories in modern physics, namely general relativity theory and quantum field theory. More interestingly, the field of string theory has gone through two conceptual revolutions since 1980s. The two revolutions provide a valuable reference point for us to verify the applicability of our visual exploratory approach, especially with reference to Kuhn's paradigm shift framework. Schwarz's historical accounts of the superstring revolutions have made it clear that string theory would be an excellent subject to test our visualization approach using Kuhn's philosophic and historic framework of scientific revolutions as high-level guidance for qualitative explanations.

Background

Quantum field theory in physics has successfully unified special relativity and quantum mechanics. However, physicists have encountered difficulties in incorporating general relativity because an ordinary quantum field theory cannot accommodate gravity. On the other hand, it has been realized that gravity comes about naturally in string theory. As a relativistic quantum theory, string theory shows great promise as a unified quantum theory of all fundamental forces.

String theory replaces the traditional notion of a point particle in quantum field theory with a string. Strings can be open or closed loops. The vibrations of strings and their interactions with the topology of the space where they exist
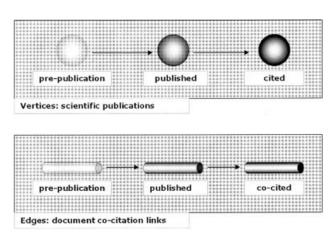

FIG. 2. Visual representations of state transitions.

dramatically simplified the representations of various particles. Scientists are very interested in string theory because it provides a promising vehicle to study various particles and forces in a consistent framework.

The Scientist named Michael B. Green of Queen Mary College in London, John H. Schwarz of the California Institute of Technology in Pasadena, and Edward Witten of Princeton University as the potential Nobel Prize winners in string theory (Martello, 1990). Green and Schwarz are known for their anomaly cancellation calculation in the 1980s (Green and Schwarz, 1984). Witten is the leading proponent of the superstring theory of unification. He is the most-cited physicist for the period 1981–1988.

Timeline

String theory was initially developed in the late 1960s to describe strong nuclear forces. The history of string theory has many "bizarre twists and turns" (Schwarz, 1996). A brief summary of important events related to the two revolutions in string theory is given here.[1]

In 1971, space-time supersymmetry was realized to be a generic feature of consistent string theories. The term superstring refers to a supersymmetric string.

In 1974, Joël Scherk and John Schwarz showed that string theory naturally accommodates gravitons. Because gravitons are necessary for a quantum gravity theory, string theory is viable to become a unified theory of all four observed forces in Nature (Scherk & Schwarz, 1974).

1984 was regarded as the beginning of the first superstring revolution. The anomaly crisis in string theory was resolved by the famous anomaly cancellation calculation. String theory became part of mainstream physics. Five distinct consistent string theories emerged as a result of the first superstring revolution.

In 1995, the second superstring revolution took place. It was discovered that five superstring theories are all equivalent because there is a unique underlying theory behind them. Edward Witten coined this underlying theory as the M-theory. The five distinct theories turned out to be perturbation expansions of M-theory about five different points. The second superstring revolution is also called the M-theory revolution, or the duality revolution because of the crucial role of duality transformation. A duality transformation can translate a difficult question into a much more accessible one.

Two Superstring Revolutions

Schwarz (1996) provides detailed accounts of the two superstring revolutions. Furthermore, his description follows Kuhn's notion of scientific revolutions. According to Kuhn, the progress of science is characterized by a sequence of phase transitions: normal science → crisis → revolution.

[1] http://superstringtheory.com/history/history3.html.

The two superstring revolutions are outlined as follows with reference to this framework.

The first superstring revolution (1984). The 1984 revolution shows an interesting resemblance to Kuhn's structure of scientific revolutions, in particular, in terms of the arrival of a crisis and revolutionary changes.

The crisis was sparkled by the anomaly problem identified by Alvarez-Gaumé and Witten ((1983). They showed that higher dimensional supersymmetric theories suffered from mathematical disasters called anomalies. It was a life-and-death crisis as far as string theory was concerned. In the summer of 1984, Green and Schwarz (1984) discovered that in superstring theory, there was a way to avoid the anomaly problem and still have a theory with sensible and realistic quantum gravity and particle interactions. This is known as the famous anomaly cancellation calculation.

The first revolution led to five distinct and consistent superstring theories, namely type I, type IIA, type IIB, E8 × E8 heterotic (HE), and SO(32) heterotic (HO), each requiring 10 dimensions (nine space and one time). Schwarz describes the significance of this revolution to string theory: "Almost overnight, the subject was transformed from an intellectual backwater to one of the most active areas of theoretical physics, which it has remained ever since."

The second superstring revolution (1995). The second superstring revolution is also known as the M-theory revolution. A revolutionary discovery is that the five distinct theories that emerged from the first revolution are fundamentally connected. The five theories are connected through duality transformations. They are five different versions of a single underlying theory—the M-theory.

A revolutionary extension of the idea of a string is the notion of membranes, or p-branes, where p is the dimensionality of a brane. Strings are 1-branes in the new notation. The second superstring revolution is directly related to a particular type of p-branes called Dirichlet branes (D-branes). Scientists studied D-branes for several years, but their significance was not clear until Polchinski's breakthrough work (Polchinski, 1995). Polchinski's paper quickly became a citation classic; it had 167 citations by August 1997 according to Science Watch®'s record—this is quite high for a purely theoretical paper. D-branes have many interesting applications, but most remarkably in the study of black holes (Strominger & Vafa, 1996). Indeed, research in black holes is a major specialty in string theory. See later sections for more detailed discussions on this topic.

The impact of the second superstring revolution is evident in several ways. According to the more recent Science-Watch® report in 1999, four of the top-10 physics papers are string theory papers (Mitton, 1999). The 6th and 9th papers are about D-branes, and the 10th paper is on M-theory, which is the underlying theory behind all five distinct string theories. Note that the above historical background is pro-

vided to set our research problem in an adequate context. This is not a prerequisite for using our approach.

Visualizing Scientific Revolutions

A paradigm is a world view of underlying theories and methodologies that the members of a scientific community share (Kuhn, 1996). Paradigms can be discovered by examining the behavior of a given community's members. In terms of the size of a community, Kuhn envisaged communities of 100 members, occasionally significantly fewer. Usually individual scientists will belong to several such groups either simultaneously or in succession (Kuhn, 1996).

Kuhn emphasized the significance of the transition from what he called the pre- to the postparadigm period in the development of a scientific field. The following is how Kuhn describes the transition:

> Before it occurs, a number of schools compete for the domination of a given field. Afterward, in the wake of some notable scientific achievement, the number of schools is greatly reduced, ordinarily to one, and a more efficient mode of scientific practice begins (Thomas Kuhn (1996, 3rd ed., p. 178).

The most striking resemblance to Kuhn's description is how M-theory unifies the five distinct consistent superstring theories in the second superstring revolution. Our basic assumption is that the intellectual structure of the string theory community, or any subject domain for that matter, will transform itself under the influence of an intellectual revolution. The differences between citation patterns before and after a scientific revolution should be a good indicator of the influence of the revolution. For example, in the anomaly cancellation calculation case, can we observe a sharp increase of citations to Schwarz's paper? Similarly, can we observe a surge of citations to Pochinski's paper on D-branes in the second superstring revolution? By modeling and visualizing how citation structures characterize conceptual revolutions over time, one can explore and understand these revolutions intuitively so that we will have a better understanding of how a discipline evolves. What was the most heavily cited work 10 years ago, 5 years ago, or 3 years ago? What are the most striking changes in citation patterns? What are the causes such changes? Our approach aims to provide a generic tool that will enable the study of such questions in a visual-spatial context.

Extracting Cocitation Patterns

To make our approach generic and easy to use, we incline to use widely available citation data and a simple procedure to select citation data for modeling and visualization. In particular, we first searched the Web of Science (WoS) for articles published between 1981 and 2002 on string theory; in particular, we searched for articles that
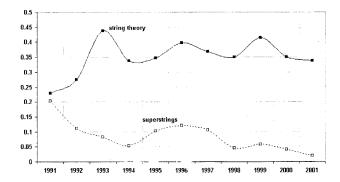

FIG. 3.   The frequency distributions of the search terms.

contain one of the two top-level descriptors of the topic, namely "string theory" and "superstring" in the topic field.

Using top-level descriptors has the following advantages over using more specific terms such as D-branes and dualities. First, it makes the method easy to use. Users only need to know the topic name to start the process. Second, this is a logical starting point. Users can use more specific search terms once they learn more domain-specific terminology and the knowledge structure of the topic. Third, using top-level topic terms enables us to work with a data set that is broad enough to lead to a coherent coverage of significant patterns. The integrity of the patterns is particularly important for a knowledge domain visualization study like this. The search extracted a total of 3,708 articles. We refer to these articles as citing articles because we are interested in their citing behavior in the string theory literature collectively. This set of citing articles has cited 137,655 references altogether, including journal articles, conference papers, books, and other types of documents.

Many records in WoS contain a field called the ID field, which includes a number of terms assigned by an indexer to specific articles. In this article, ID terms refer to such terms in the ID field. We examined the frequency distribution of ID terms in the citing data set. Because the ID field was not included in records prior to 1991, the frequency distribution analysis is limited to the period of 1991–2002. The ID term frequency distribution of the two search terms, "string theory" and "superstring," was plotted by the year of publication. The plot shows two relatively level lines across the entire period (see Fig. 3). The only difference is that string theory is a much more popular ID term than superstring. This indicates that the citation data has a consistent coverage of the topic.

In addition, we examined the frequency distribution of other popular ID terms during the same period between 1991 and 2002. The relative frequency of each ID term was computed each year to avoid the fluctuations caused by the number of ID terms assigned in a particular year. For example, the relative frequency of D-branes in year 2000 is the frequency of D-branes in the ID field of all records in 2000 divided by the total number of records with the ID field. This is still a crude indicator, but it is suffice for our purpose. Figure 4 shows the relative frequency distributions of seven most popular ID-terms in this set of citing articles.
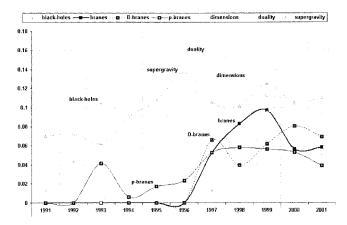
black-holes — branes — D-branes — p-branes — dimensions — duality — supergravity

FIG. 4. Relative frequency distributions of seven most frequently used ID terms.

The relative simple ID term frequency distributions have revealed some interesting patterns. Terms such as supergravity, duality, and black holes appeared more evenly than terms such as branes, D-branes, and p-branes, which increased sharply in recent years due to their role in the second superstring revolution. This simply analysis suggests that terms such as supergravity, duality, or even black holes characterize fundamental issues shared by both revolutions, whereas branes, D-branes, or p-branes represent new ideas crucial to the second revolution only.

Citation analysis traditionally focuses on the cream of crop in that the construction of a citation image is based on the most productive and influential scientists. This study only considers those articles that have at least a total of 35 citations accumulated between 1981 and 2002. A document cocitation network was formed based on 62,588 nonzero cocitation relationships between 375 such articles. The cocitation strength of a pair of documents is determined by a corresponding Pearson's correlation coefficient derived from the original 375-by-375 cocitation matrix. Pathfinder network scaling ($q = N - 1$, $r = \infty$) was applied to the correlation matrix to reduce its complexity. As a result, a Pathfinder network of 375 nodes and 381 links was produced. For technical details of Pathfinder network scaling, see Chen and Paul (2001) and Schvaneveldt (1990).

Principle component analysis (PCA) was also applied to the correlation matrix to identify leading specialties in string theory research. PCA extracted 18 components, which collectively accounted for 95% of the variances in the cocitation data. The largest three components accounted for 66% of the variances. One of the useful features of PCA is that if we use extract components to identify specialties, we can characterize documents that simultaneously belong to more than one specialty, which is very common in a scientific community. See White and McCain (1998) and Chen and Paul (2001) for technical details.

Base Map and Thematic Overlay

The Pathfinder-enhanced cocitation network was rendered for a three-dimensional display. The three-dimensional visualization model consists of a two-dimensional base map and a thematic overlay along the third dimension. The base map is essentially the Pathfinder network partitioned by specialty memberships derived from PCA.

The network partition was designed to clarify the interrelationships among specialties. Each document was mapped to a triplet of factor loading coefficients of the largest three components extracted by PCA. These triplets were in turn mapped to colors in the RGB color scheme (R for red, G for green, and B for blue). This mapping scheme characterizes a document by its profiles in three specialties rather than in its strongest profile in a single specialty. For example, if a document is representative in the largest specialty but not in the other two specialties, then it would be shown in a blend of a lot of red, little or nothing green and blue, which will appear almost as red. Similarly, a typical document in the second- or the third largest specialty would be in green or in blue accordingly. In contrast, if a document has about the same strength in all three primary specialties, then it would appear in white or gray, depending on the magnitude of the strengths. Finally, if a document has almost nothing to do with any of the three primary specialties, then it would be shown in dark colors—in such cases, the "home specialty" of these documents tends to be a minor specialty, which is represented by an extracted component that only explains a small amount of variances in PCA.

PCA-based network partition helps users identify specialties in a domain at disciplinary levels. We utilized this technique in our previous studies and found some interesting patterns (Chen & Paul, 2001). For example, leading specialties tend to occupy the center of the cocitation network, whereas minor specialties tend to appear towards the outer rim of such networks.

The thematic overlay represents annual citation counts of documents featured in the base map. A series of annual citation counts of a document is depicted as a multisectioned color bar growing upwards from the document

TABLE 1. Total variance explained.

Component	Total	% of Variance	Cumulative %
1	125.749	33.533	33.533
2	68.127	18.167	51.700
3	53.980	14.395	66.095
4	28.848	7.693	73.788
5	15.908	4.242	78.030
6	14.726	3.927	81.957
7	12.218	3.258	85.215
8	9.897	2.639	87.854
9	6.677	1.781	89.635
10	4.275	1.140	90.775
11	3.794	1.012	91.786
12	2.723	.726	92.512
13	2.355	.628	93.141
14	1.978	.527	93.668
15	1.661	.443	94.111
16	1.544	.412	94.522
17	1.081	.288	94.810
18	1.016	.271	95.081

TABLE 2. Title words frequently used by the top-10 articles in each of the first three specialties.

Factors	1: The second revolution (Red)		2: The first revolution (Green)		3: Black holes (Blue)
Freq.	Extracted Phrases	Freq.	Extracted Phrases	Freq.	Extracted Phrases
4	Dirichlet branes (D-branes)	4	superconformal field-theories	5	black-holes
2	Born-Infeld action	4	string theory	3	string theory
2	p-branes	2	moduli	2	Dilaton

sphere. The color of each section corresponds to a year in the citing window between 1981 and 2002. Lighter colors indicate citations made in more recent years, whereas darker colors indicate citations made in earlier years. This design enables users to identify potential trends in citation impact. The annual growth of citations was animated. Together with information in the base map, an animated thematic overlay provides users additional cues to understand the citation dynamics of a paradigm.

To trace various bibliographic transitions in the history of string theory, we introduced a new feature in this study and displayed various state transitions in animated visualizations. Two state spaces were defined: one for individual documents, and the other for cocitation relationships. The document state space contains three states: prepublication, publication, and citation. Given a document, before its publication, the document is in the state of prepublication; after its publication but before it has ever been cited, it is in the state of publication; and it is in the state of citation if it has been cited. Three state transitions are defined in this space: prepublication $\rightarrow$ publication, publication $\rightarrow$ citation. A document's state transitions are shown as the document sphere becomes more and more visible, from highly transparent, to moderately transparent, and finally to completely solid. For example, given a particular point in the time variable, a moderately transparent document sphere indicates a published document with no citations at that time.

Similarly, the cocitation state space also contains three states and each state specifies the strength of a cocitation link: very weak connection, weak connection, and strong connection. A cocitation link is in the very weak connection state if at least one of the two documents to be connected by the link is in the prepublication state; a cocitation link is weak if both documents are in the publication state or even in the citation state, but there is no cocitation; a cocitation link becomes a strong connection once the two documents have been actually cocited. As cocitation links join documents together, small islands of documents are merged into large continents of specialties. Visualizing such processes is an interesting implementation of the "continental drift" metaphor we proposed earlier in this article.

The Rising Landscape

Before we examine visual patterns in the intellectual landscape of string theory, it is useful to identify the primary specialties. PCA has extracted 18 components. Each component corresponds to a specialty. The first three components account for 66% of the variances (see Table 1). Because each component corresponds to a specialty, it is useful to concentrate on leading specialties.

The nature of a specialty is determined by its member documents. We extracted most frequently used title terms from the 10 documents that were strongly featured in the corresponding component. This procedure can be implemented algorithmically. This method is illustrated in Table 2. Headings of factor components are explained shortly.

TABLE 3. Factor 1: The second superstring revolution: Polchinski's article in 1995 is a core paper in this revolution.

Factor Loading	Publications
0.946	Callan, C. G. (1998). Brane dynamics from the Born-Infeld action. Nuclear Physics B, 513(1–2), 198–212.
0.941	Polchinski, J. (1996). HEPTH9611050.
0.940[a]	Witten, E. (1996). Bound states of strings and p-branes. Nuclear Physics B, 460(2), 335–350.
0.939	Strominger, A. (1996). Open p-branes. Physics Letters B, 383(1), 44–47.
0.934	Douglas, M. R. (1995). HEPTH9512077.
0.934[b]	Polchinski, J. (1995). Dirichlet branes and Ramond-Ramond charges. Physical Review Letters, 75(26), 4724–4727.
0.915	Banks, T., Fischler, W., Shenker, S. H., & Susskind, L. (1997). M theory as a matrix model: A conjecture. Physical Review D, 55(8), 5112–5128.
0.908	Douglas, M. R., Kabat, D., Pouliot, P., & Shenker, S. H. (1997). D-branes and short distances in string theory. Nuclear Physics B, 485(1–2), 85–127.
0.905	Tseytlin, A. A. (1996). Self-duality of Born-Infeld action and dirichlet 3-brane of type IIB superstring theory. Nuclear Physics B, 469(1–2), 51–67.
0.890	Green, M. B., Harvey, J. A., & Moore, G. (1997). I-brane inflow and anomalous couplings on D-branes. Classical and Quantum Gravity, 14(1), 47–52.

[a] Number 6 in the top-10 most-cited physics papers ranked by Science Watch in 1997.
[b] Number 2 in the top-10 most-cited physics papers ranked by Science Watch in 1997.

TABLE 4. Factor 2: The first superstring revolution.

Factor loading	Publications
0.857	Green, M.B., & Schwarz, J.H. (1984). Anomaly cancellations in supersymmetric D = 10 gauge theory and superstring theory. Physics Letter B, 149, 117–122.
0.817	Cecotti, S., Ferrara, S., & Girardello, L. (1989). Geometry of Type-II superstrings and the moduli of superconformal field-theories. International Journal of Modern Physics A, 4(10), 2475–2529.
0.817	Seiberg, N. (1988). Observations on the moduli space of superconformal field-theories. Nuclear Physics B, 303(2), 286–304.
0.806	Scherk, J. (1979), Nuclear Physics B, 153, 61.
0.783	Narain, K.S. (1986). New heterotic string theories in uncompactified dimensions less-than 10. Physics Letters B, 169(1), 41–46.
0.777	Witten, E. (1982). Constraints on supersymmetry breaking. Nuclear Physics B, 202(2), 253–316.
0.764	Narain, K.S. (1987). Nuclear Physics B, 279, 369.
0.749	Dine, M., Huet, P., & Seiberg, N. (1989). Large and small radius in string theory. Nuclear Physics B, 322(2), 301–316.
0.739	Candelas, P., Delaossa, X.C., Green, P.S., & Parkes, L. (1991). A pair of Calabi-yau manifolds as an exactly soluble superconformal theory. Nuclear Physics B, 359(1), 21–74.
0.738	Sagnotti, A. (1992). A note on the Green-Schwarz mechanism in open-string theories. Physics Letters B, 294(2), 196–203.

Documents in the largest specialty characterize the second superstring revolution, or the M-theory revolution. Among the 10 most representative documents in this specialty, four are on D-branes, two on p-branes, and one on M-theory (see Table 3). Polchinski's revolution-making article on D-branes also has a strong presence in this specialty.

Similarly, the second specialty is essentially represented by the first superstring revolution. Green and Schwarz's anomaly cancellation paper in 1984 has the strongest factor loading in this specialty (see Table 4). It was this anomaly cancellation calculation that was regarded as the beginning of the first superstring revolution.

The third specialty is about black holes (see Table 5). Research in black holes played an important role in the development of superstring theory.

Landscape of Revolutions

In subsequent analysis, we focus on the citation landscape of superstring revolutions as well as the formation of cocitation networks. Once the growth pattern is translated into a visual language, it is easier to trace the dynamics of interesting patterns as well as drill down to a finer granularity of details.

Figures 5, 6, and 7 are three snapshots of the citation landscape of string theory. Figure 5 shows an early snapshot of the landscape. The scene was dominated by the first revolution specialty, shown as clusters of articles in green at the far side of the landscape. Elsewhere in the scene, the majority of documents were barely visible because they were not published at that time. Most cocitation links were very weak, rendered as highly transparent, in this scene.

Figure 6 shows a snapshot of the landscape a few years later. The first revolution consolidated itself with more citations and more member articles. Citation bars became taller, more document spheres emerged, and cocitation links became stronger.

In addition, the specialty of black holes, mainly located in the eastern part of the scene, emerged and proliferated outside-in from the rim of the cocitation network. However, the central area and the landscape as a whole were still only sporadically occupied.

TABLE 5. Factor 3: Black holes.

Factor loading	Publications
0.894	Gibbons, G.W., & Maeda, K. (1988). Black-holes and membranes in higher-dimensional theories with Dilaton fields. Nuclear Physics B, 298(4), 741–775.
0.891	Garfinkle, D., Horowitz, G.T., & Strominger, A. (1991). Charged black-holes in string theory. Physical Review D, 43(10), 3140–3143.
0.887	Shapere, A., Trivedi, S., & Wilczek, F. (1991). Dual dilaton dyons. Modern Physics Letters A, 6(29), 2677–2686.
0.872	Maharana, J., & Schwarz, J.H. (1993). Noncompact symmetries in string theory. Nuclear Physics B, 390(1), 3–32.
0.869	Horowitz, G.T., & Steif, A.R. (1990). Spacetime singularities in string theory. Physical Review Letters, 64(3), 260–263.
0.866	Gibbons, G.W. (1982). Anti-gravitating black-hole solitons with scalar hair in N = 4 supergravity. Nuclear Physics B, 207(2), 337–349.
0.863	Hassan, S.F., & Sen, A. (1992). Twisting classical-solutions in heterotic string theory. Nuclear Physics B, 375(1), 103–118.
0.819	Susskind, L., & Uglum, J. (1994). Black-Hole Entropy in Canonical Quantum-Gravity and Superstring Theory. Physical Review D, 50(4), 2700–2711.
0.807	Holzhey, C.F.E., & Wilczek, F. (1992). Black-holes as elementary-particles. Nuclear Physics B, 380(3), 447–477.
0.800	Kallosh, R., Linde, A., Ortin, T., Peet, A., & Vanproeyen, A. (1992). Supersymmetry as a cosmic censor. Physical Review D, 46(12), 5278–5302.

FIG. 5. The early landscape of string theory was dominated by the first superstring revolution, shown as green clusters of documents located at the far side of the image.

The snapshot in Figure 7 shows the arrival of the second superstring revolution. A large number of new articles appeared in previously uninhabited areas. Even more remarkably, many of these new articles began to make their impact almost immediately—their citations were growing in leaps and bounce.

Less is More

The goal of this study is to find out in what way visual exploration can enhance traditional analysis of scientific revolutions. What features of scientific revolutions can be made visible? How important and how useful are they for studying once intangible ideas such as invisible colleges and paradigms? The two revolutions in string theory provide a good opportunity for us to investigate these questions from new perspectives.

Figure 8 shows a snapshot of the landscape just before the second superstring revolution. Because at that point, key papers that triggered the new revolution were not yet published, document spheres representing these papers remained semitransparent. A useful function of such information hiding is that it enables users to identify the growth patterns of closely connected articles as a group as opposed to as individual articles.

Figures 9 and Figure 10 show close-up snapshots of the landscape before and after the second revolution. The rapid growth of the second revolution specialty is clearly visible

FIG. 6. A few years later, the landscape was characterized by the continuous growth of the first revolution specialty (document spheres in red) and the emergence of the black holes specialty (document spheres in blue).
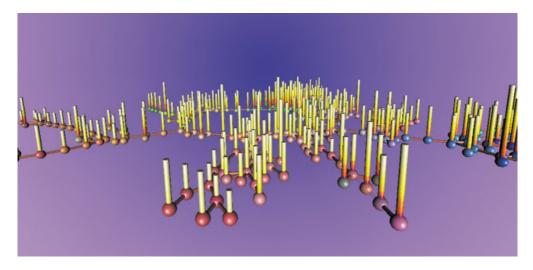
FIG. 7. The more recent landscape of string theory was overwhelmed by the arrival of the second superstring revolution (the branch in the middle and the one stretching to the left).

in the animated version. Figure 11 shows an overview of the landscape in the middle of a formation process. Green and Schwarz's anomaly cancellation paper was strongly connected by a cocitation link to Witten's 1984 paper on gravitational anomalies. Polchinski's 1995 paper on D-branes was in the prepublication state. In fact, its neighboring papers on the central ring of the network were also written by Polchinski.

The string theory study has demonstrated the potential of visual exploration approaches to the study of scientific revolutions and competing paradigms. The modeling and visualization techniques experimented in this study should be consolidated and integrated into methodologies for the study of the dynamics of a scientific paradigm.

Discussions and Conclusions

We have visualized and explored the changing landscape of string theory over the last 20 years based on the growth of citations and the strengths of cocitation links. The analysis has focused on the use of visualization techniques in the study of the two revolutions in string theory. We have made special reference to Kuhn's structure of scientific revolutions.

What have we actually visualized? Would Kuhn regard the two superstring revolutions as scientific revolutions? A Kuhn's revolution is preceded by a crisis. The first superstring revolution was triggered by the anomalies discovery. The second superstring revolution was in part caused by the puzzle over the coexistence of five distinct consistent string
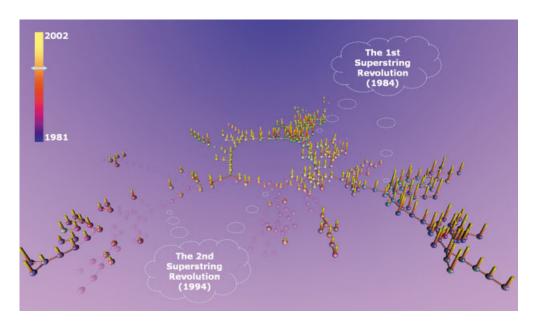

FIG. 8. Information hiding in the landscape view can help users identify the growth patterns of documents in groups.
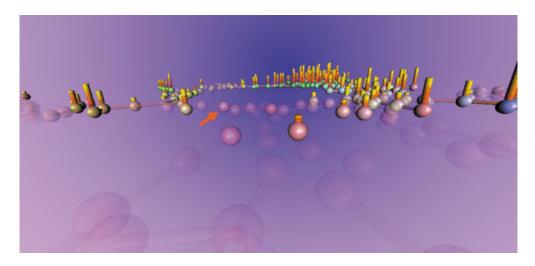
FIG. 9. A close-up snapshot of the landscape before the second revolution. The red arrow in the scene points to Polchinski's 1995 paper on D-branes.

theories. A Kuhn's revolution also includes a special turning point—an intellectual Gestalt switch. To the first revolution, it was the anomaly cancellation calculation; to the second revolution, it was the M-theory and Polchinski's work on D-branes. However, one needs to be cautious and avoid overinterpreting visual patterns. It is important to note that our approach is intended to facilitate the study of developments of a scientific discipline, rather than replace critical thinking in philosophy of science or other realms.

Will the visualizations be useful for information science, philosophy of science, or other subject domains? The usefulness of the visualizations is likely to be as enabling techniques rather than as discovery and decision-making devices. Visualizations enable us to augment our perceptual and cognitive abilities in exploring complex patterns more easily and more vividly than what can be done in traditional science studies. Schwarz's historic accounts have been very helpful for us to interpret various patterns in the visualization model. The generic applicability of our methodology means that one may apply it repeatedly to the same knowledge domain or to different domains.

Kuhn's earlier version of his theory is akin to a fittest survival model: The wining paradigm takes it all and the losing paradigm will die out. His later version modified this view with a specialization model, in which different paradigms can coexist and evolve along their own paths. This issue is not directly addressed in our current visualization. In this model, because growth of citation bars is based on the actual citations instead of citation rates, a specialty can still show a growing impact even if its growing rate is slowing down. Therefore, one should probably consider alternative impact measures so as to distinguish the growth pattern of a growing paradigm from that of a shrinking paradigm.

In conclusion, we have demonstrated some useful functions of our approach. The relative frequency distributions have shown the stability of the top-level indexing terms of string theory. The analysis of the visualized intellectual landscape has produced meaningful results. This is a viable approach to the study of developments of a scientific field. On the other hand, a lot more needs to be done. This study has raised theoretic and practical issues concerning the research in knowledge domain visualization in general. Our approach is just one of many potentially effective routes to the study of knowledge discovery in science. We used Kuhn's theory as our framework; we used Wagner's conti-
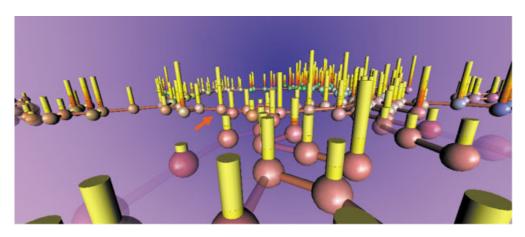

FIG. 10. A close-up snapshot of the landscape after the second revolution. The newly formed specialty was gaining citations rapidly.
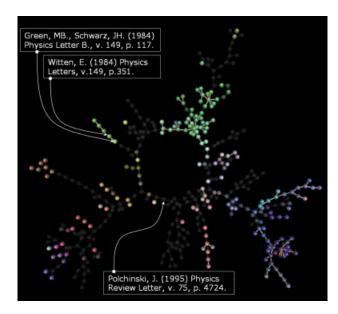
FIG. 11. An overview of the intellectual landscape in the making. Green and Schwarz's paper and Witten's paper belong to the first superstring revolution, whereas Polchinski's paper belongs to the second revolution.

nental drift as our visual metaphor; we used cocitation networks as the basis of an evolving landscape an indicator of intellectual changes. We took the advantage of Kuhn's "history-friendly" theory, the vivid continental drift visual metaphor, and the well-established cocitation methodology. There are, of course, many other ways, proven and untried, to derive a new methodology. From the philosophy of science point of view, we could use theories from Kuhn's followers or even his criticizers, such as Karl Popper, Paul Feyerbend, and Imre Lakatos. From the visual metaphor point of view, we could revisit the work of Henry Small and other pioneers in automating the study of specialty dynamics. From the visualization point of view, we could explore the use of fractals and other visualization schemes rather than networks (van Raan, 1991, 2000). The list is obviously not comprehensive, and we hope this study provides a useful exemplar in its own right and stimulates more interests and more studies in modeling and visualizing scientific paradigms.

References

Alvarez-Gaumé, L., & Witten, E. (1983). Gravitational anomalies. Nuclear Physics B, 234, 269–330.

Boyack, K.W., Wylie, B.N., & Davidson, G.S. (2002). Domain visualization using VxInsight for science and technology management. Journal of the American Society for Information Science and Technology 43(9), 764–774.

Chen. C. (1999a). Information visualisation and virtual environments. London: Springer-Verlag.

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.

Chen, C., & Paul, R.J. (2001). Visualizing a knowledge domain's intellectual structure. Computer, 34(3) 65–71.

Chen, C., Cribbin, T., Macredie, R., & Morar, S. (2002). Visualizing and tracking the growth of competing paradigms: Two case studies. Journal of the American Society for Information Science and Technology, 53(8), 678–689.

Chen, C., Kuljis, J., & Paul, R.J. (2001). Visualizing latent domain knowledge. IEEE Transactions on System, 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.

Garfield, E. (1994). Scientography: Mapping the tracks of science. Current Contents: Social & Behavioural Sciences, 7(45), 5–10.

Garfield, E. (1998, February 14). Mapping the world of science. Paper presented at the 150 anniversary meeting of the AAAS, Philadelphia, PA.

Green, M.B., & Schwarz, J.H. (1984). Anomaly cancellations in supersymmetric D = 10 gauge theory and superstring theory. Physics Letters B, 149, 117–122.

Kuhn, T.S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Kuhn, T.S. (1996). The structure of scientific revolutions (3rd ed.). Chicago: Univesity of Chicago Press.

Martello, A. (1990). Twelve prolific physicists: Likely 1990 Nobel contenders. The Scientist, 4(17), 16.

Mitton, S. (1997). Strings maintain tight wrap on physics top 10. Science Watch®, 8(6), 7.

Mitton, S. (1999). Fifty orders of magnitude stretch physics to the limits. Science Watch®, 10(2), 6.

Polchinski, J. (1995). Dirichlet branes and Ramond-Ramond charges. Physical Review Letters, 75(26), 4724–4727.

Scherk, J., & Schwarz, J. (1974). Dual models for non-hadrons. Nuclear Physics, B81, 118.

Schvaneveldt, R.W. (Ed.). (1990). Pathfinder associative networks: Studies in knowledge organizaiton. Norwood, NJ: Ablex Publishing Corporation.

Schwarz, J.H. (1996). The second superstring revolution (arXiv:hep-th/9607067).

Small, H.G. (1977). A co-citation model of a scientific specialty: A longitudinal study of collagen research. Social Studies of Science, 7, 139–166.

Small, H. (1999a). A passage through science: Crossing disciplinary boundaries. Library Trends, 48(1), 72–108.

Small, H. (1999b). Visualizing science by citation mapping. Journal of the American Society for Information Science, 50(9), 799–813.

Small, H.G., & Griffith, B.D. (1974). The structure of scientific literatures I: Identifying and graphing specialties. Science Studies, 4, 17–40.

Strominger, A., & Vafa, C. (1996). Microscopic origin of the Bekenstein-Hawking entropy. Physics Letters B, 379(1–4), 99–104.

Thagard, P. (1992). Conceptual revolutions. Princeton, NJ: Princeton University Press.

van Raan, A. (1991). Fractal geometry of information space as represented by co-citation clustering. Scientometrics, 20, 439–449.

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

White, H.D., & McCain, K.W. (1998). Visualizaing a discipline: An author co-citation analysis of information science, 1974–1995. Journal of the American Society for Information Science, 49(4), 327–356.