Visualizing Latent Domain Knowledge

Chaomei Chen, Jasna Kuljis, and Ray J. Paul

Abstract—Knowledge discovery and data mining commonly rely on finding salient patterns of association from a vast amount of data. Traditional citation analysis of scientific literature draws insights from strong citation patterns. Latent domain knowledge, in contrast to the mainstream domain knowledge, often consists of highly relevant but relatively infrequently cited scientific works. Visualizing latent domain knowledge presents a significant challenge to knowledge discovery and quantitative studies of science. In this paper, we build upon a citation-based knowledge visualization procedure and develop an approach that not only captures knowledge structures from prominent and highly cited works, but also traces latent domain knowledge through low-frequency citation chains. We apply this approach to two cases: 1) identifying cross-domain applications of Pathfinder networks (PFNETs) and 2) clarifying the current status of scientific inquiry of a possible link between Bovine spongiform encephalopathy (BSE), also known as mad cow disease, and a new variant Creutzfeldt–Jakob disease (vCJD), a type of brain disease in human.

Index Terms—Citation chains, knowledge discovery, knowledge domain visualization (KDViz), latent domain knowledge.
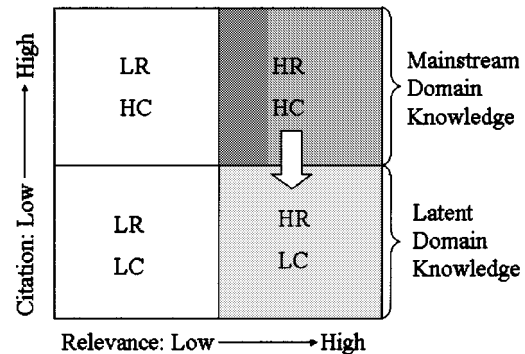
Fig. 1. Mainstream domain knowledge is typically high in both relevance and citation, whereas latent domain knowledge can be characterized as HR and LC.
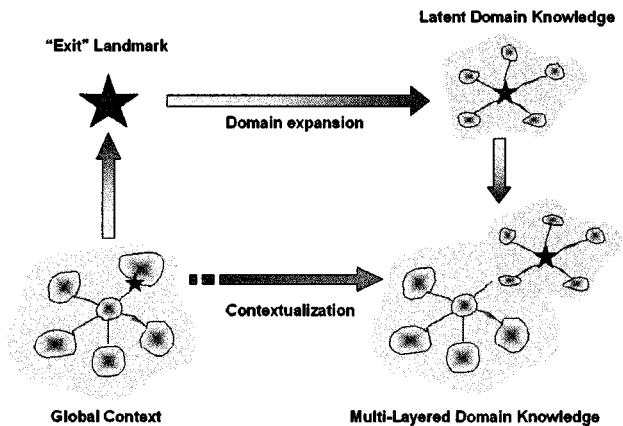

Fig. 2. Strategy of visualizing latent domain knowledge. The global context is derived from co-citation networks of highly cited works. An "exit" landmark is chosen from the global context to serve as the seeding paper in the process of domain expansion. The expanded domain consists of papers connecting to the seeding paper by citation chains of no more than two citation links. Latent domain knowledge is represented through a citation network of these papers.

I. INTRODUCTION

KNOWLEDGE tracking and technology monitoring become increasingly important for knowledge management. The rapid advances of information visualization in recent years have highlighted its great potential in knowledge discovery and data mining [1]–[3].

Knowledge discovery and data mining commonly rely on finding salient patterns of association from a vast amount of data. Traditional citation analysis of scientific literature draws new insights from strong citation patterns. Such citation patterns typically reflect the view of the so-called "cream of the crop" in a knowledge domain, such as eminent scientists and their highly cited classic works. Latent domain knowledge, in contrast to the mainstream domain knowledge, often consists of highly relevant but relatively infrequently cited scientific works. Strong citation patterns become less common for latent domain knowledge: They could be completely missing or overwhelmed by the body of the mainstream domain knowledge. Visualizing latent domain knowledge presents a significant challenge to knowledge discovery and quantitative studies of science.

There may be many reasons why a particular line of research may fall outside the body of the mainstream domain knowledge and become latent to a knowledge domain. In a cross-disciplinary research program, researchers face an entirely unfamiliar scientific discipline. Tracking the latest development into a different discipline can be rather challenging. One example of such problems is the cross-disciplinary use of Pathfinder networks (PFNETs), a structural and procedural modeling method developed by cognitive psychologists in the 1980s [4], [5]. Pathfinder is a generic tool that has been adapted by several fields of study, including some quite different adaptations from its original cognitive applications. For example, we have adapted PFNET scaling as an integral component of our generic structuring and visualization framework [1], [2], [6]. It is a challenging task to track down how applications of PFNETs have evolved over the past two decades across a number of apparently unconnected disciplines.

Another type of latent domain knowledge can be explained in terms of scientific paradigms. Kuhn [7] described the development of science as interleaved phrases of normal science and
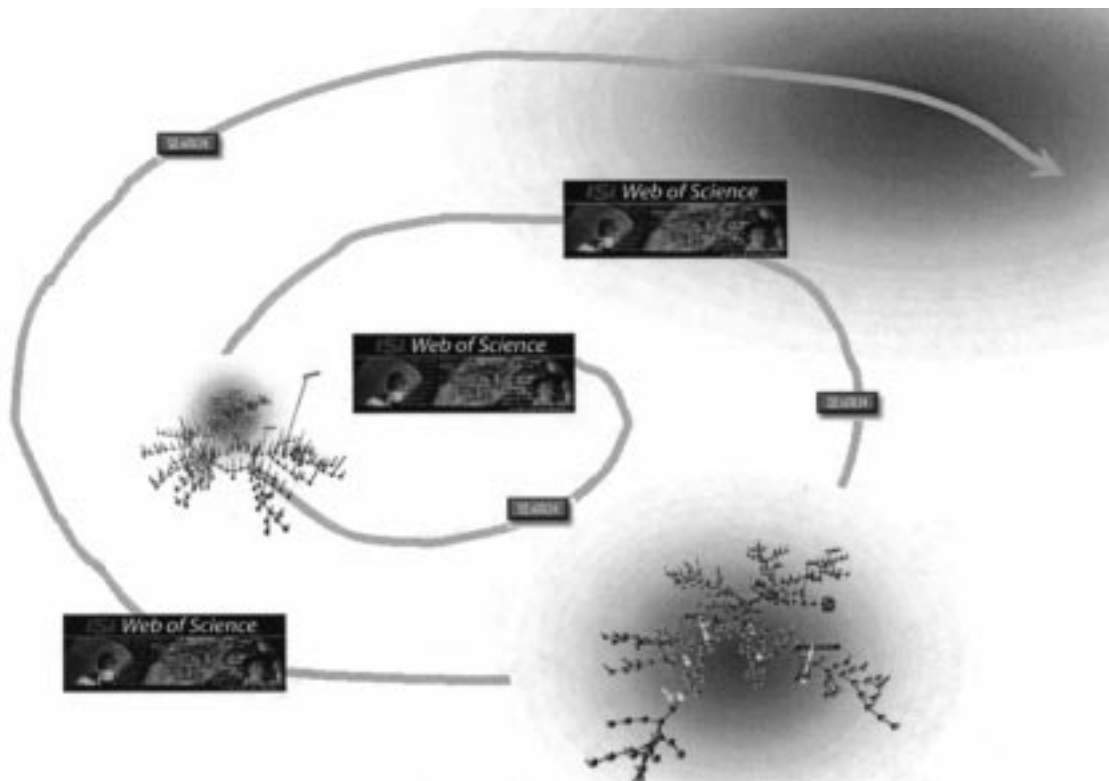
Fig. 3. Spiral methodology of domain expansion and visualization.

scientific revolutions. A period of normal science is typically marked by the dominance of an established framework. 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 are introduced. New theories replace with greater explanatory power the ones in trouble in a revolutionary manner. Science regains another period of normal science.

Kuhn suggested that a paradigm shift in science should lead to a corresponding change of citation patterns in scientific literature; therefore, the study of such patterns may provide indicators of the development of a scientific paradigm. Indeed, a number of researchers pursued this line of research since 1970s. For example, Small [8] studied the movement of highly cited publications on the topic of collagen as a means of tracking major paradigm shifts in this particular field. White and McCain [9] used INDSCAL to depict changes in author co-citation maps over consecutive periods . We have started to investigate how information visualization can help us characterize the dynamics of scientific paradigms [10]. In particular, our focus is on contemporary puzzle-solving topics in science and medicine: What caused dinosaurs' mass extinction? Are Bovine spongiform encephalopathy (BSE) and the new variant Creutzfeldt–Jakob disease (vCJD) connected? What powers active galactic centers, super-massive black holes, or something else?

In this paper, we introduce an approach to visualizing latent domain knowledge. We demonstrate how one can accommodate latent domain knowledge and the mainstream domain knowledge within the same visualization framework. This paper includes two case studies: 1) PFNET applications and 2) theo-
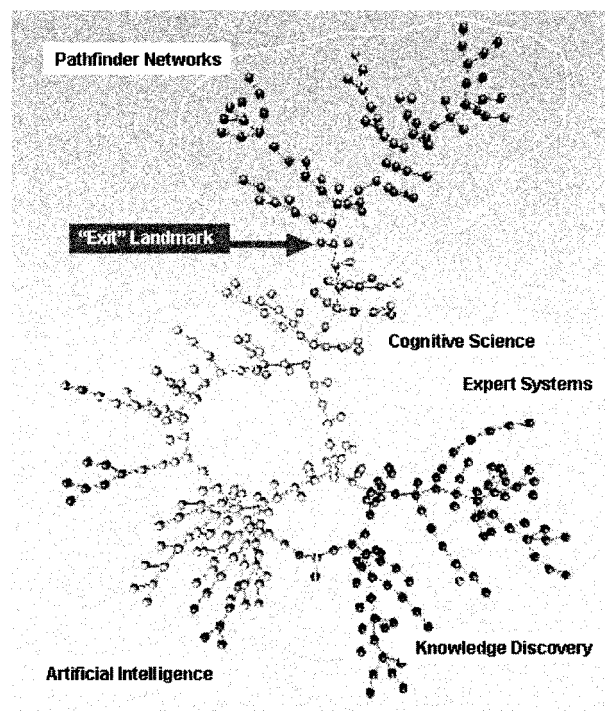

Fig. 4. Overview of the mainstream domain knowledge.

ries of BSE, commonly known as mad cow disease. The rest of the paper is organized as follows. First, we outline existing work, including citation analysis, knowledge discovery, and examples. We then extend our domain visualization approach to visualize latent domain knowledge. We apply this approach to
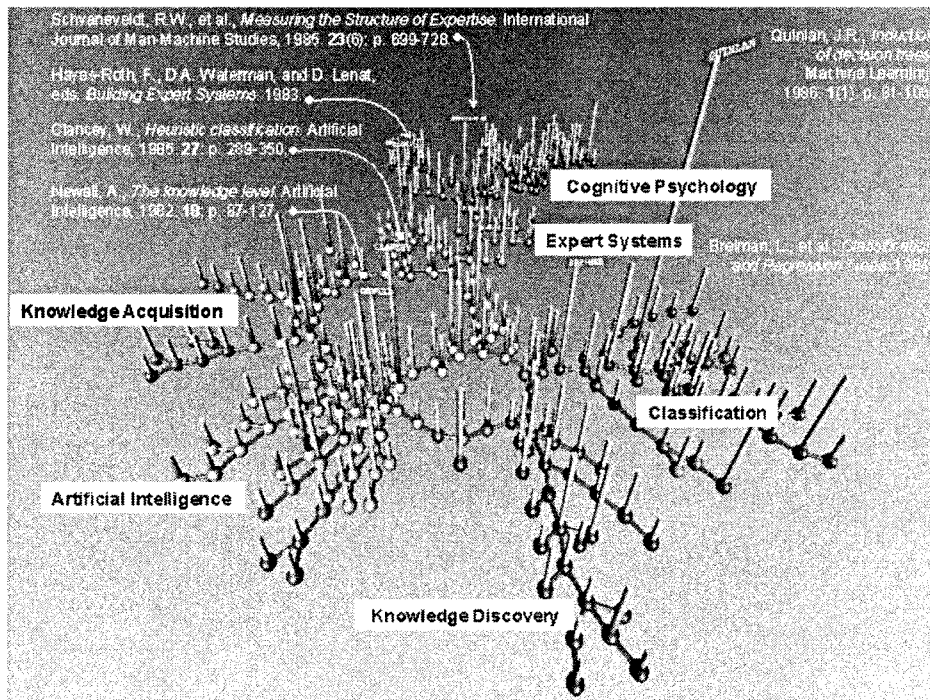
Fig. 5.   Global context of the Pathfinder case. Applications of PFNETs are found in a broader context of knowledge management technologies, such as knowledge acquisition, knowledge discovery, and artificial intelligence. A majority of PFNET users are cognitive psychologists.

two cases in which visualizing latent domain knowledge is involved: 1) tracing applications of PFNETs and 2) connecting a controversial theory of BSE, mad cow disease, to the mainstream intellectual structure of research in BSE.

II. KNOWLEDGE DISCOVERY

The advances of information visualization have revived the interest in a number of challenging issues concerning knowledge tracking. We contrast two strands of research in this paper: 1) the citation-based paradigm of knowledge discovery and 2) the undiscovered public knowledge approach. The key prerequisite for the citation-based paradigm is a target scientific literature that is rich in citations, whereas the undiscovered public knowledge deals with exactly the opposite situation when citation links are missing or are considerably rare. A synergy of the two would lead to a more powerful tool to facilitate knowledge discovery and knowledge management in general.

A. Citation-Centric Paradigm for Knowledge Visualization

Pioneering examples of using citation data to map the structure of science include the creation of the historical map of research in DNA [11], mapping scientific networks [12], and a study of Nobel Prize winners' citation profiles [13].

Since the 1970s, information scientists began to look for ways to reveal patterns and trends in science through studies of scientific literature. In 1974, Small and Griffith [14] addressed issues concerning mapping specialties based on co-citation patterns . In 1977, Small conducted a longitudinal study of collagen research and showed that some rapid changes of focus took place in the research [8]. He used data from the science citation indexing (SCI) to compute co-citation strengths between

pairs of documents and subsequently clustered documents to identify leading specialties, or paradigms. He then used multidimensional scaling (MDS) to map highly cited papers each year in clusters on a two-dimensional (2-D) plane. An abrupt disappearance of a few key documents in the leading cluster in one year and a rapidly increased number of documents in the leading cluster in the following year indicate an important type of specialty change in terms of rapid shift in research focus, which is an indicator of "revolutionary" changes. Garfield [15] discussed longitudinal mapping as a series of chronologically sequential maps to depict how knowledge advances.

White and Griffith [16] introduced author co-citation analysis (ACA) in 1981 as a literature measure of intellectual structure . In ACA, the unit of analysis is authors and their intellectual relationships as reflected through scientific literature. McCain [17] produced a comprehensive technical review of mapping authors in intellectual spaces. In 1998, White and McCain applied ACA on information science [9].

Intellectual groupings of highly cited authors or their papers represent the underlying knowledge domain. We have developed a four-step domain visualization procedure to support both ACA and document co-citation analysis (DCA) [2]. Co-citation patterns are typically represented by a co-citation matrix based on citation data drawn from citation databases such as the SCI. There are a range of valid entities for co-citation analysis, e.g., author–author, document–document, journal–journal, descriptor–descriptor, and even country–country. A co-citation matrix of more than 100 entities can be rather complex. Researchers routinely use clustering algorithms to divide such matrices into smaller ones. Matrix analysis techniques such as singular value decomposition (SVD) and eigenvalue decomposition are also available for this divide-and-conquer purpose. We

incorporate PFNET scaling into our approach. The strength of PFNET scaling is that it can simplify a complex co-citation network but also keep the salient structure intact.

In co-citation analysis, factor analysis is often used to reduce the dimensionality of co-citation data. Each identified factor represents a specialty in a given knowledge domain. Therefore, one can break a knowledge domain down to several specialties in such a way. Normally, we are interested in the most significant factors identified by factor analysis in terms of their explanation power. In contrast, co-citation analysis often uses MDS to project a high-dimensional data space to a 2-D or three-dimensional (3-D) one so that inter-entity relationships can be depicted in a 2-D or 3-D graphical representation. However, it had not been a common practice in co-citation analysis to consolidate the results of both factor analysis and MDS configuration into the same graphical representation. Our domain visualization procedure renders the distribution of specialties across a co-citation landscape. In addition to structural connectivity, the membership of a specialty is color-coded. Entities in similar colors are likely to belong to the same specialty. The advantage of this approach is that the structural connectivity and the special membership color can reinforce each other and make the visualization easy to understand. Finally, the magnitude of impact of a particular entity is depicted through the height of its citation bar, showing a stack of color-coded annual citation sections.

B. Undiscovered Public Knowledge—The Missing Link Paradigm

Swanson describes three aspects of the context and nature of knowledge fragmentation [18].

1) There is an enormous and constantly growing gap between the entire body of recorded knowledge and the limited human capacity to make sense of it.
2) Inadequate cross-specialty communication causes knowledge fragmentation. In response to the information explosion, specialties are increasingly divided into more and more narrowly focused subspecialties.
3) One specialty might not be aware of potentially valuable information in another specialty. Two specialized literatures may be isolated in terms of explicit citation links, but they may have implicit, latent connections at the text level.

Swanson has been pursuing his paradigm since 1986 when he found two sizeable biomedical literatures: one is on the circulatory effects of dietary fish oil and the other is on the peripheral circulatory disorder, Raynaud's disease. Swanson noticed that these two literatures were not bibliographically-related, i.e., no one from one camp cited works in the other [19]. On the other hand, he was pondering the question that apparently no one had asked before: Is there a connection between dietary fish oil and Raynaud's disease?

Prior to Swanson's research, no medical researcher had noticed this connection, and the indexing of these two literatures was unlikely to facilitate the discovery of any such connections. Swanson's approach can be represented in a generic form. Given two premises that A causes B ($A \rightarrow B$) and that B causes C

TABLE I
LEADING ARTICLES IN THE THREE LARGEST SPECIALTIES RANKED BY THE STRENGTH OF FACTOR LOADING. ABSOLUTE VALUES LESS THAN 0.500 ARE SUPPRESSED FROM THE TABLE. FACTORS F1, F2, AND F3 DEFINE THREE SPECIALTIES. THE "EXIT" LANDMARK BELONGS TO THE FIRST SPECIALTY

Article	F1	F2	F3
CARD SK, 1983, PSYCHOL HUMAN COMPUT	.872		
JOHNSONLAIRD PN, 1983, MENTAL MODELS	.858		
NISBETT RE, 1977, PSYCHOL REV, V84, P231	.855		
GLASER R, 1988, NATURE EXPERTISE, PR15	.850		
GAMMACK JG, 1985, RES DEV EXPERT SYSTE, P105	.841		
CHI MTH, 1981, COGNITIVE SCI, V5, P121	.841		
COOKE NM, 1986, P IEEE, V74, P1422	.836		
COOKE NM, 1987, INT J MAN MACH STUD, V26, P533	.830		
ANDERSON JR, 1982, PSYCHOL REV, V89, P369	.814		
ANDERSON JR, 1987, PSYCHOL REV, V94, P192	.813		
MCKEITHEN KB, 1981, COGNITIVE PSYCHOL, V13, P307	.811		
CHI MTH, 1989, COGNITIVE SCI, V13, P145	.810		
ANDERSON JR, 1983, ARCHITECTURE COGNITI	.807		
CORDINGLEY ES, 1989, KNOWLEDGE ELICITATIO, P89	.804		
COOKE NJ, 1994, INT J HUM-COMPUT ST, V41, P801	.798		
HOFFMAN RR, 1987, AI MAG, V8, P53	.797		.528
CHASE WG, 1973, COGNITIVE PSYCHOL, V4, P55	.794		
KLEIN GA, 1989, IEEE T SYST MAN CYB, V19, P462	.792		.508
SCHVANEVELDT RW, 1985, INT J MAN MACH STUD, V23, P699	.789	-.532	
MARCUS S, 1988, AUTOMATING KNOWLEDGE		.951	
MUSEN MA, 1987, INT J MAN MACH STUD, V26, P105		.949	
BENNETT JS, 1985, J AUTOMATED REASONIN, V1, P49		.947	
CLANCEY WJ, 1989, MACH LEARN, V4, P285		.942	
NEWELL A, 1982, ARTIF INTELL, V18, P87		.942	
MUSEN MA, 1989, KNOWL ACQUIS, V1, P73		.941	
CLANCEY WJ, 1985, ARTIF INTELL, V27, P289		.940	
FORD KM, 1993, INT J INTELL SYST, V8, P9		.933	
KAHN G, 1985, 9TH P INT JOINT C AR, P581		.933	
MUSEN MA, 1989, AUTOMATED GENERATION		.930	
NECHES R, 1991, AI MAG, V12, P36		.929	
MARCUS S, 1989, ARTIF INTELL, V39, P1		.926	
CHANDRASEKARAN B, 1986, IEEE EXPERT, V1, P23		.925	
LENAT DB, 1990, BUILDING LARGE KNOWL		.923	
CHANDRASEKARAN B, 1983, AI MAG, V4, P9		.921	
DAVIS R, 1982, KNOWLEDGE BASED SYST		.920	
DAVIS R, 1979, ARTIF INTELL, V12, P121		.918	
GRUBER TR, 1987, INT J MAN MACH STUD, V26, P143		.914	
SHADBOLT N, 1990, CURRENT TRENDS KNOWL, P313		.912	
DEKLEER J, 1984, ARTIF INTELL, V24, P7		.910	
HOLLAND JH, 1986, INDUCTION PROCESSES			.771
OLEARY DE, 1987, DECISION SCI, V18, P468			.713
WATERMAN DA, 1986, GUIDE EXPERT SYSTEMS	.526		.712
MICHALSKI RS, 1980, INT J MAN MACH STUD, V12, P63		.593	.674
OLSON JR, 1987, EXPERT SYST, V4, P152	.668		.672
MILLER GA, 1956, PSYCHOL REV, V63, P81			.671
HART A, 1986, KNOWLEDGE ACQUISITIO	.640		.664
PRERAU DS, 1990, DEV MANAGING EXPERT			.657
MESSIER WF, 1988, MANAGE SCI, V34, P1403	-.611		.635
QUINLAN JR, 1979, EXPERT SYSTEMS MICRO	-.644		.631
JACKSON P, 1990, INTRO EXPERT SYSTEMS	.530		.627
JOHNSON PE, 1983, J MED PHILOS, V8, P77	.510		.612
BOOSE JH, 1986, EXPERTISE TRANSFER E	.578		.601
RUMELHART DE, 1986, PARALLEL DISTRIBUTED	-.575		.599
HARMON P, 1985, EXPERT SYSTEMS		.546	.597
KIM J, 1988, DECISION SUPPORT SYS, V4, P269	.654		.591
SHAW MLG, 1987, KNOWL ACQUIS, P109	.580		.585
QUINLAN JR, 1979, EXPERT SYSTEMS MICRO, P168			.585
SAATY TL, 1980, ANAL HIERARCHY PROCE		.508	.580
MICHALSKI R, 1980, INT J POL ANAL INF S, V4, P125	-.664		.571

($B \rightarrow C$), the question to ask is whether A causes C ($A \rightarrow C$). If the answer is positive, the causal relation has the transitive property. In the biological world, such transitive properties may not always be there. Therefore, scientists must explicitly es-
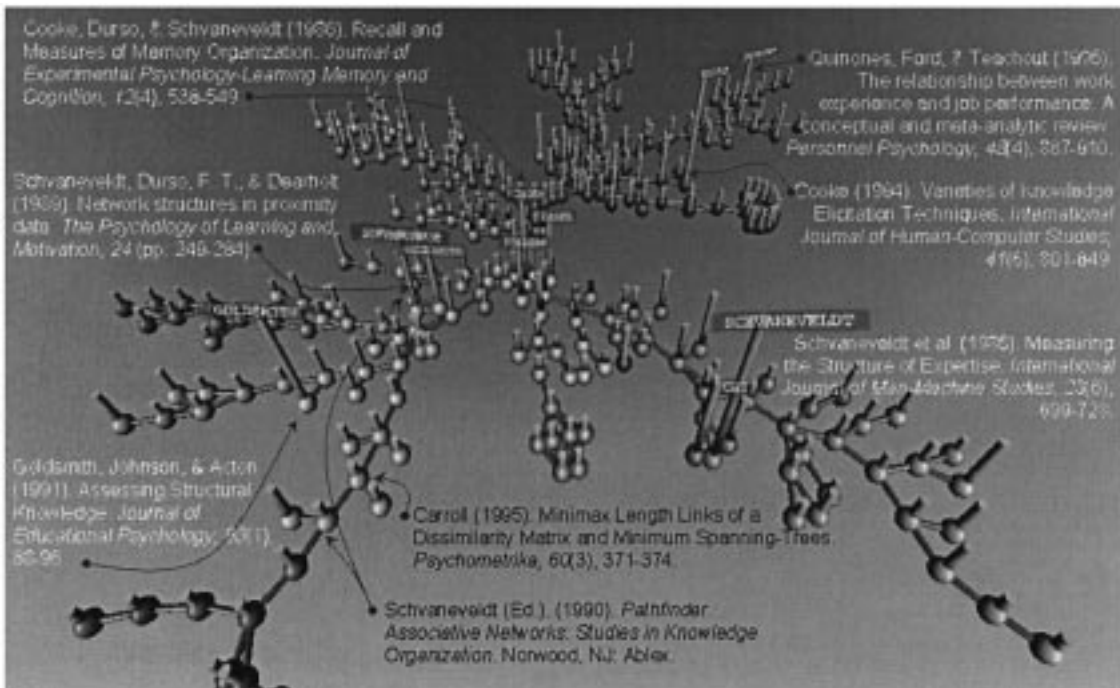
Fig. 6. Citation map showing that the most prolific themes of PFNET applications include measuring the structure of expertise, eliciting knowledge, measuring the organization of memory, and comparing mental models. No threshold is imposed.

tablish such transitivity relationships. Swanson [18] suggests that once information scientists identify such possibilities, they should recommend domain experts to validate them.

Swanson's approach focuses on the discovery of such hypotheses from the vast amount of implicit, or latent, connections. Swanson and Smalheiser [20] defined the concept of *noninteractive* literature. If two literatures have never been cited together at a notable level, they are noninteractive, i.e., scientists have not considered both of them together. In the past 15 years, Swanson identified several missing links of the same pattern, notably migraine and magnesium [21] and arginine and somatomedin C [22]. Since 1994, the collaboration between neurologist Smalheiser and Swanson led to a few more such cases [23], [24]. They also made their software *Arrowsmith* available on the Internet [25].

Swanson's approach relies on the identification of the two premises: $A \rightarrow B$ and $B \rightarrow C$. In a large knowledge domain, it is crucial for analysts to have sufficient domain knowledge. Otherwise, to find two such premises is like searching for needles in a haystack. Knowledge domain visualization (KDViz) can narrow down the search space and increase the chance of finding a fruitful line of scientific inquiry.

C. Visualizing Latent Domain Knowledge

In this paper, we distinguish mainstream domain knowledge and latent domain knowledge along two dimensions: relevance and citation. Scientific documents in the literature can be classified into four categories according to their relevance to the subject domain and their citations received from the scientific literature: mainstream domain knowledge, which typically consists of documents of HR and high citations (HC); latent domain knowledge, which are typically made of documents of

HR but low citations (LC); and two categories of documents of low relevance. The traditional knowledge discovery such as citation analysis and domain visualization focuses on the mainstream domain knowledge (HR+HC). The focus of latent domain knowledge discovery and visualization is on the category of HR and LC. In this paper, we introduce an approach that can extend the coverage of KDViz from mainstream to latent domain knowledge (see Fig. 1).

In our earlier work [3], we developed a four-step procedure for visualizing mainstream domain knowledge. In particular, the procedure consists of the following four steps.

Step 1) Select highly relevant and highly cited documents from a citation database.
Step 2) Derive citation networks based on the selected population of documents and simplify citation networks using PFNET scaling.
Step 3) Partition the resultant PFNET according to specialties identified through principal component analysis.
Step 4) Superimpose the citation history of a document or author over the citation network.

Our solution to visualizing latent domain knowledge is built upon this four-step procedure. Instead of simply applying the procedure on highly relevant and highly cited documents, we incorporate this procedure into a recursive process particularly suitable for detecting patterns in highly relevant but sparsely cited documents. Fig. 2 illustrates the overall strategy of our approach. This approach has three subprocesses. The purpose of the first process is to establish a global context by subsequent analysis and visualization. Indeed, in this process, we apply our four-step procedure to the mainstream domain knowledge and generate a citation landscape. The second process is domain expansion, which means that we expand our field of view from

mainstream domain knowledge to latent domain knowledge. A key component in this domain expansion process is the selection of a so-called "exit" landmark from the citation landscape. This "exit" landmark will play a pivot role in tracking latent knowledge by "pulling" highly relevant but relatively rarely cited documents into the scene. The "exit" landmark is selected based on both structural and topical characteristics. Structurally important documents in the citation landscape include branching points, from which one can reach more documents along citation paths preserved by the network. Topically important documents are the ones that are closely related to the subject in question. Ideally, a good "exit" landmark should be a classic work in a field of study and it can link to a cluster of closely related documents by citation. We will explain in more detail through case studies how we choose "exit" landmarks. Once an "exit" landmark is chosen from the citation landscape, the four-step procedure can be applied again to all the documents within a citation chain of up to two citation links. The resultant citation network represents the latent domain knowledge. Finally, we embed this local structure back into the global context by providing a reference from the "exit" landmark in the global context to the latent knowledge structure.

In this paper, we describe how we applied this approach to two case studies. One is about cross-domain applications of PFNET scaling techniques. The other is about clarifying the perceived connection between BSE and vCJD in contemporary literature. We use the Web of Science, a Web-based interface to citation databases compiled by the Institute for Scientific Information (ISI). We start with a search in the Web of Science using some broad search terms in order to generate a global context for subsequent visualization. For example, in the Pathfinder case, we chose to use search terms such as *knowledge discovery*, *knowledge acquisition*, *knowledge modeling*, and *Pathfinder*. Once the global context is visualized, it is straightforward to identify an "exit" landmark. In the Pathfinder case, a classic citation of PFNETs is chosen as an "exit" landmark. This "exit" landmark paper serves as the seed in a citation search within the Web of Science. The citing space of the seeding paper s contains papers that either cite the seeding paper directly or cite an paper that in turn cites the paper

$$C_{\text{One-Step}}(s) = \{c | c \rightarrow s\}$$

$$C_{\text{Two-Step}}(s) = \{c | \exists c' \Rightarrow c \rightarrow c' \wedge c' \rightarrow s\}$$

$$\text{Citing Space}_{\text{Theme}}(s) = C_{\text{One-Step}}(s) \cup C_{\text{Two-Step}}(s).$$

Such citing spaces may contain papers beyond the boundary of the mainstream domain knowledge. The spiral shape of the diagram shown in Fig. 3 implies that one can repeatedly apply this method by identifying another "exit" landmark. Papers connected to the landmark by two-step citation chains are gathered to represent latent domain knowledge. By using different ways to select citing papers, we can visualize latent knowledge structures with reference to highly established and frequently cited knowledge structures. In the following two case studies, we apply the same spiral methodology to illustrate our approach.

TABLE II
LEADING ARTICLES IN THE THREE MOST PROMINENT SPECIALTIES RANKED BY THE STRENGTH OF FACTOR LOADING. ABSOLUTE VALUES LESS THAN 0.500 ARE SUPPRESSED FROM THE TABLE. AT LEAST ABOVE-THRESHOLD FACTOR LOADING IS REQUIRED TO BE INCLUDED IN THE LISTING. THE FIRST MEMBER OF THE FIRST SPECIALTY IS THE "EXIT" LANDMARK CHOSEN FOR DOMAIN EXPANSION

Publication	F1 Pathfinder, Cognitive Psychology	F2 Educational Psychology	F3 Knowledge Acquisition
SCHVANEVELDT RW, 1985, INT J MAN MACH STUD, V23, P699	0.916		
ANDERSON JR, 1983, ARCHITECTURE COGNITI	0.906		
REITMAN JS, 1980, COGNITIVE PSYCHOL, V12, P554	0.874		
FRIENDLY ML, 1977, COGNITIVE PSYCHOL, V9, P188	0.861		
MCKEITHEN KB, 1981, COGNITIVE PSYCHOL, V13, P307	0.848		
ERICSSON KA, 1984, PROTOCOL ANAL	0.845		
COOKE NM, 1987, INT J MAN MACH STUD, V26, P533	0.837		
CHI MTH, 1981, COGNITIVE SCI, V5, P121	0.825		
KRUSKAL JB, 1977, STATISTICAL METHODS	0.822		
COOKE NM, 1986, P IEEE, V74, P1422	0.822		
HAYESROTH F, 1983, BUILDING EXPERT SYST	0.807		
MURPHY GL, 1984, J EXP PSYCHOL LEARN, V10, P144	0.806		
ROSKEHOESTRAND RJ, 1986, ERGONOMICS, V29, P1301	0.803		
ANDERSON JR, 1982, PSYCHOL REV, V89, P369	0.801		
COOKE NJ, 1988, INT J MAN MACH STUD, V29, P407	0.800	0.514	
TVERSKY A, 1977, PSYCHOL REV, V84, P327	0.798		
KELLY GA, 1955, PSYCHOL PERSONAL CON	0.790		
BUTLER KA, 1986, ARTIFICIAL INTELLIGE	0.789		
COLLINS AM, 1969, J VERB LEARN VERB BE, V8, P240	0.784		
SCHVANEVELDT RW, 1985, MCCS859 NEW MEX STAT	0.777		
GOLDSMITH TE, 1991, J EDUC PSYCHOL, V83, P88		0.840	
GONZALVO P, 1994, J EDUC PSYCHOL, V86, P601		0.789	
ACTON WH, 1994, J EDUC PSYCHOL, V86, P303		0.777	
GOMEZ RL, 1996, J EDUC PSYCHOL, V88, P572		0.754	
JOHNSON PJ, 1994, J EDUC PSYCHOL, V86, P617		0.747	
NOVAK JD, 1990, J RES SCI TEACH, V27, P937		0.747	
NOVAK JD, 1984, LEARNING LEARN		0.744	
SCHVANEVELDT RW, 1989, PSYCHOL LEARN MOTIV, P249		0.744	
FENKER RM, 1975, INSTR SCI, V4, P33		0.737	
SCHVANEVELDT RW, 1988, COMPUT MATH APPL, V15, P337		0.734	
SCHVANEVELDT RW, 1990, PATHFINDER ASS NETWO	0.601	0.726	
WILSON JM, 1994, J RES SCI TEACH, V31, P1133		0.734	
ARABIE P, 1993, CONTEMP PSYCHOL, V38, P66		0.720	
PREECE PFW, 1976, J EDUC PSYCHOL, V68, P1		0.716	
ROSCH E, 1975, J EXPT PSYCHOL GENER, V104, P192		0.711	
GOMEZ RL, 1996, J HLTH PSYCHOL, V1, P107		0.710	
GOMEZ RL, 1994, J EXP PSYCHOL LEARN, V20, P396		0.710	
CRAIK KJW, 1943, NATURE EXPLANATION		0.706	
CANAS JJ, 1994, INT J HUM-COMPUT ST, V40, P795		0.698	
SCHVANEVELDT RW, 1989, PSYCHOL LEARN MOTIV, V24, P249		0.696	0.501
SHAW MLG, 1989, KNOWL ACQUIS, V1, P341			0.623
KITTO CM, 1989, INT J MAN MACH STUD, V31, P149			0.618
KITTO CM, 1987, P WESTEX 87 W C EXP, P96			0.571
SANDERSON PM, 1994, HUMAN COMPUTER INTER, V9, P251			0.566
COOKE NJ, 1996, HUM-COMPUT INTERACT, V11, P29			0.560
COOKE NJ, 1992, INT J MAN MACH STUD, V37, P721		-0.551	0.517
WALSH JP, 1988, ORGAN BEHAV HUM DEC, V42, P194			0.511
ROWE AL, 1996, J EXP PSYCHOL-APPL, V2, P31			0.503
WIELINGA BJ, 1992, KNOWL ACQUIS, V4, P5			0.503

III. CASE STUDY I: PATHFINDER NETWORKS

In the Pathfinder case, we focus on cross-domain applications of PFNETs. It is particularly important to ensure that nonmainstream applications of PFNETs are not overshadowed by mainstream citation peaks.
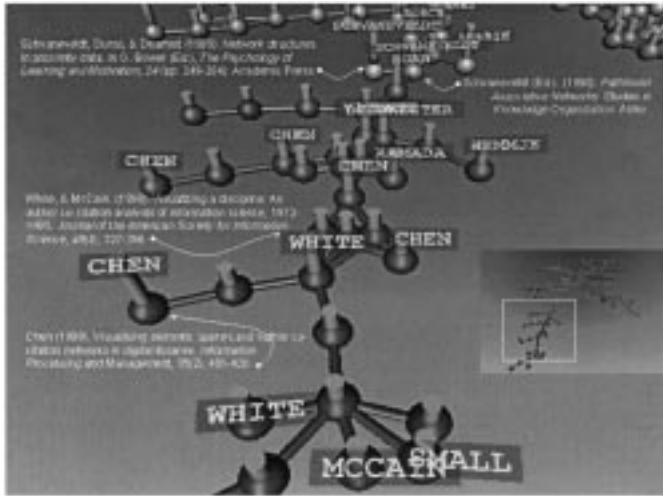
Fig. 7. Branch representing a new paradigm of incorporating PFNETs into GSA, a generic framework for structuring and visualization, and its applications especially in strengthening traditional citation analysis.

A. Pathfinder Network (PFNETs) Scaling

PFNET scaling was originally developed by cognitive psychologists for structuring modeling [5]. It relies on a triangle inequality condition to select the most salient relations from proximity data. PFNETs have the same set of vertices as the original graph. The number of edges in a PFNET, on the other hand, can be largely reduced.

The topology of a PFNET is determined by two parameters (q and r) and the corresponding network is denoted as PFNET (r, q). The q-parameter controls the scope that the triangular inequality condition should be imposed. The r-parameter refers to the Minkowski metric used for computing the distance of a path. The weight of a path P with k links, $W(P)$, is determined by weights $w_1, w_2, \ldots, w_k$ of each individual link as follows:

$$W(P) = \left(\sum_{i=1}^{k} w_i^r \right)^{1/r}$$

The Minowski distance (geodetic) depends on the value of the r-metric. For $r = 1$, the path weight is the sum of the link weights along the path; for $r = 2$, the path weight is computed as Euclidean distance; and for $r = \infty$, the path weight is the same as the maximum weight associated with any link along the path

$$W(P) = \left(\sum_{i=1}^{k} w_i^r \right)^{1/r} = \begin{cases} \sum_{i=1}^{k} w_i, & r = 1 \\ \left(\sum_{i=1}^{k} w_i^2 \right)^{1/2}, & r = 2 \\ \max_i w_i, & r = \infty. \end{cases}$$

The q-parameter specifies that triangle inequalities must be satisfied for paths with $k \leq q$ links

$$w_{n_1 n_k} = \left(\sum_{i=1}^{k-1} w_{n_i n_{i+1}}^r \right)^{1/r} \qquad \forall k \leq q.$$

TABLE III
STRONG NEGATIVE FACTOR LOADING IN FACTOR ONE SUGGESTING A UNIQUE SPECIALTY. THESE ARTICLES PATHFINDER NETWORKS ARE USED, BUT NOT IN ANY WAY SIMILAR TO A TYPICAL PUBLICATION IN THE PATHFINDER SPECIALTY

Publication	F1
MCCAIN KW, 1995, J AM SOC INFORM SCI, V46, P306	-0.619
BUSH V, 1945, ATLANTIC MONTHLY, V176, P101	-0.631
KAMADA T, 1989, INFORM PROCESS LETT, V31, P7	-0.651
CHEN CM, 1996, HUM-COMPUT INTERACT, V11, P125	-0.652
CONKLIN J, 1987, IEEE COMPUT, V20, P17	-0.657
BRAAM RR, 1991, J AM SOC INFORM SCI, V42, P233	-0.661
MARSHALL C, 1994, P ECHT 94 ED SEPT, P13	-0.661
DILLON A, 1996, INT J HUM-COMPUT ST, V45, P619	-0.664
GREEN SJ, 1998, P 7 INT WORLD WID WE	-0.664
BENYON D, 1997, P HUM COMP INT INTER, P39	-0.664
CAMPAGNONI FR, 1989, ACM T INFORM SYST, V7, P271	-0.666
MCCAIN KW, 1990, J AM SOC INFORM SCI, V41, P433	-0.667
WHITE HD, 1981, J AM SOC INFORM SCI, V32, P163	-0.668
HEMMJE M, 1994, P 17 ANN INT ACM SIG, P249	-0.670
WHITE HD, 1997, ANNU REV INFORM SCI, V32, P99	-0.672
SMALL H, 1973, J AM SOC INFORM SCI, V24, P265	-0.673
CHEN C, 1997, NEW REV HYPERMEDIA M, V3, P67	-0.675
VICENTE KJ, 1988, INT J MAN MACH STUD, V29, P647	-0.680
DEERWESTER S, 1990, J AM SOC INFORM SCI, V41, P391	-0.680
SMALL H, 1999, J AM SOC INFORM SCI, V50, P799	-0.682
CHEN C, 1998, P 9 ACM C HYP HYP HY, P77	-0.684
CHALMERS M, 1992, P 15 ANN INT ACM SIG, P330	-0.688
CHEN CM, 1998, J VISUAL LANG COMPUT, V9, P267	-0.693
CHEN CM, 1998, INTERACT COMPUT, V10, P107	-0.695
SALTON G, 1983, INTRO MODERN INFORMA	-0.697
WHITE HD, 1998, J AM SOC INFORM SCI, V49, P327	-0.724
SMALL H, 1997, SCIENTOMETRICS, V38, P275	-0.724
HETZLER B, 1998, P 5 INT ISKO C STRUC	-0.724
SMALL H, 1994, SCIENTOMETRICS, V30, P229	-0.724
CHEN HC, 1998, J AM SOC INFORM SCI, V49, P582	-0.727
FOX KL, 1999, J AM SOC INFORM SCI, V50, P616	-0.736
CHEN CM, 1999, INFORM PROCESS MANAG, V35, P401	-0.743

When a PFNET satisfies the following three conditions, the distance of a path is the same as the weight of the path:

1) The distance from a document to itself is zero.
2) The proximity matrix for the documents is symmetric; thus, the distance is independent of direction.
3) The triangle inequality is satisfied for all paths with up to q links.

If q is set to the total number of nodes less one, then the triangle inequality is universally satisfied over the entire network. The number of links in a network can be reduced by increasing the value of parameter r or q. The geodesic distance between two nodes in a network is the length of the minimum-cost path connecting the nodes. A minimum-cost network (MCN), PFNET $(r = \infty, q = n - 1)$, has the least number of links. See [1], [2], [5], and [6] for further details.

Typical applications of PFNETs include modeling a network of concepts based on similarity ratings given by human experts, constructing procedural and protocol analysis models of complex activities such as air-traffic control, and comparing learners' PFNETs at various stages of their learning [4].
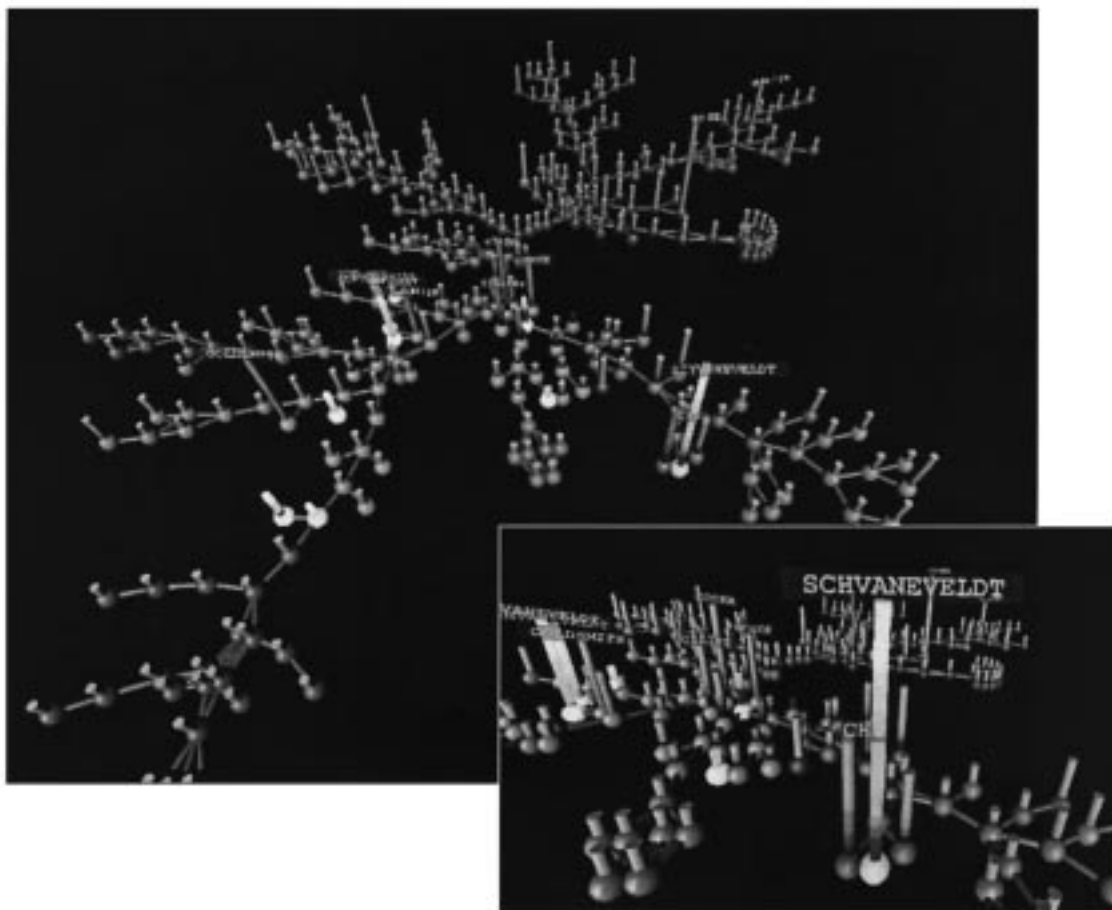
Fig. 8. Schvaneveldtl's "exit" landmark in the landscape of the thematic visualization.

In our earlier research, we incorporated PFNETs into our generalized similarity analysis (GSA) framework [1], [26]. Traditionally a typical application of PFNETs rely on proximity data judged manually. The number of nodes in a typical PFNET ranges from 30 to 50, although PFNETs of 2000 nodes were reported in one occasion in the 1980s [5]. We introduced a variety of computer-generated proximity measures along with GSA including document-document similarity computed based on information retrieval models, state transition probabilities derived from a Web navigation, and co-citations of authors as well as documents [6]. These proximity data have extended the scope of PFNETs to a much wider variety of phenomenon beyond the amount of proximity data one can measure by hand. This extension has not only enriched the topological properties of PFNETs but also led to valuable insights into the meaning of PFNETs. The Pathfinder case study is motivated by the question: How does this extension fit into the general picture of PFNET applications with reference to traditional Pathfinder applications?

B. Mainstream Domain Knowledge

The global context of the Pathfinder case shown in Fig. 5 contains clusters of papers on knowledge discovery, knowledge acquisition, classification and machine learning, artificial intelligence, expert systems, and domain knowledge modeling. Pathfinder-related papers are located in the far side of the landscape view, near the area labels of cognitive psychology and expert systems (see Fig. 4). This indicates that applications of PFNETs are closely to these two broad categories. In order to pursue latent knowledge structures associated with PFNETs. Schvaneveldt's 1985 paper was chosen as the first "exit" landmark because it is located at a point connecting the Pathfinder "peninsula" to other areas in the landscape.

Table I lists further details concerning the structure of the global context as derived from factor analysis. Up to 20 leading papers in each of the three largest factors, or specialties, are listed. In essence, Factor 1 corresponds to research in PFNETs. Factor 2 corresponds to classic artificial intelligence. Factor 3 corresponds to expert systems and decision support systems. The higher a factor loading, the more typical an paper is as a representative member of the specialty. On the other hand, if an paper has a wide impact, then its loadings on individual factors may not be exceedingly high.

C. Latent Domain Knowledge

Fig. 6 shows the latent knowledge structure derived from the citing space of the "exit" landmark paper. This structure is not overshadowed by HC of classic artificial intelligence papers, but it maintains a connecting point with the global context through the "exit" landmark, which is the highest citation bar half way down in the branch pointing to the lower right corner. This detailed local structure shows more papers related to the use of Pathfinder.

Similarly, Table II shows leading papers in this latent knowledge structure. The classification is more detailed than the one in the global context.

Fig. 7 shows an extended branch from the main PFNET. This branch represents a new area of applying PFNETs. In fact, this is the area in which PFNETs have been adapted for citation-based visualizations.

Table III reveals the fact that papers in this branch all have negative loadings on Factor 1 and are virtually absent from the remaining factors. This is interesting because, on the one hand, the first specialty provides a dimension that can account for both the traditional applications of Pathfinders and the new branch of applications. On the other hand, since documents in the new branch are so consistently classified by factor loading, they can be treated as a subspecialty.

Fig. 8 shows a simple research function which lights up all the papers by Schvaneveldt, a central figure in the development of PFNET scaling. The position of each lit paper and the direction of the hosting branch provide insightful information into the nature of the paper and the branch.

IV. CASE STUDY II: POSSIBLE LINKS BETWEEN BSE AND vCJD

Prusiner, a Nobel prize winner for his discovery of prions—a type of bad protein, suggested that an abnormal form of a protein is responsible for diseases such as scrapie in sheep, BSE in cattle (mad cow disease), and Creutzfeldt–Jakob disease (CJD) in humans. These diseases are known as transmissible spongiform encephalopathy (TSE).

A. Mainstream Domain Knowledge

BSE was first found in 1986 in the U.K.. A sponge-like malformation was found in the brain tissue from affected cattle. It was identified as a new prion disease, a new TSE disease. The BSE epidemic in the U.K. reached its peak in 1992 and has since steadily declined. CJD was first discovered in the 1920s by two German neurologists. It is the principal form of a number of human TSE diseases. In humans, the prion-based disease is related to CJD, Kuru (transmitted by cannibalism), Gerstmann–Sträussler–Scheinker disease (GSS), and fatal familial insomnia (FFI).

The vCJD is an unrecognized variant of CJD discovered by the National CJD Surveillance Unit in Edinburgh, U.K. The vCJD is characterized clinically by a progressive neuropsychiatric disorder. Neuropathology shows marked spongiform change throughout the brain. The media reported a growing concern in the general public that BSE may have passed from cattle to humans. The British government assured the public that the beef is safe, but in 1996 it announced there is possibly a link between BSE and vCJD. The central question in this case study is what scientific literature tells us about the possible link between BSE and vCJD.

First, we generated a mainstream-driven thematic landscape of the topic of BSE and CJD by searching the Web of Science with the term "BSE or CJD" (see Fig. 9). The strongest specialty, Prion protein, is colored in red; the BSE specialty is in green; and the CJD specialty is in blue. In particular, the very light color
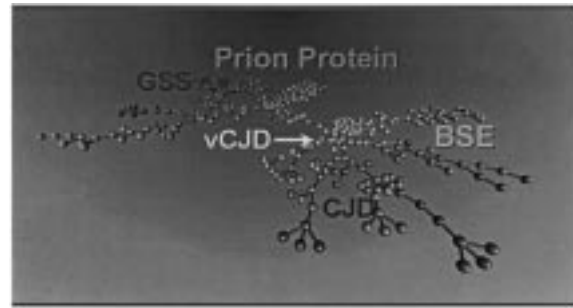

Fig. 9. Global view of mainstream research in BSE and vCJD.

TABLE IV
THE CITATION PROFILE OF PURDEY'S ARTICLES. THE ID NUMBERS ARE CROSS-REFERENCES TO LABELS IN FIG. 12. THE HIGHEST CITATION COUNT IS 9, WHICH IS CONSIDERABLY LESS THAN CITATIONS OF MORE THAN 900 TIMES RECEIVED BY KEY ARTICLES ON PRION THEORY

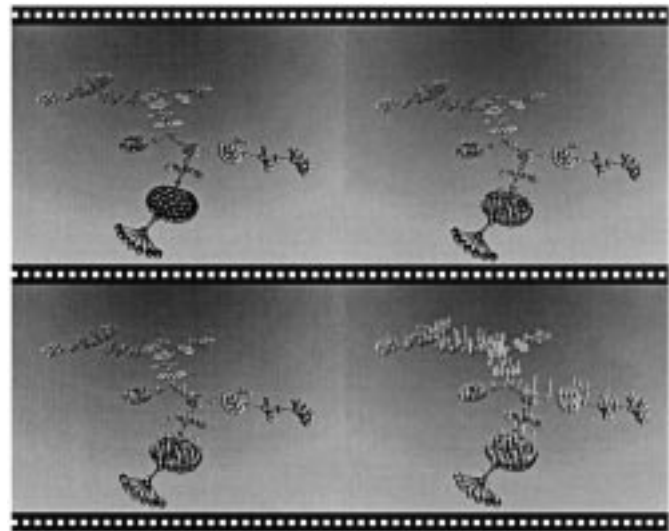| ID	Cites	Author	Year	Source
1	9	Purdey, M	1996	Medical Hypotheses, 46(5), 445-454
2	5	Purdey M	1994	Journal of Nutritional Medicine, 4, 43-82
3	5	Purdey M	1996	Medical Hypotheses, 46(5), 429-443
4	5	Purdey, M	1998	Medical Hypotheses, 50(2), 91-111

Fig. 10. Year-by-year animation highlights the dynamics of citation trends. Four snapshots are extracted from the animation sequence of the BSE data (1981–2001), containing papers on two-step citation chains to Brown's 1997 paper.

of the vCJD specialty indicates that this is an area where other specialties overlap.

B. Manganese–Copper Hypothesis

The mainstream view on BSE has focused on the food chain: cows got BSE by eating feed made from sheep infected with scrapie, and, similarly, humans get vCJD by eating BSE infected beef. However, Purdey, a British organic dairy farmer, believed that the unbalanced manganese and copper in the brain is the real cause of BSE and vCJD [27]. He studied the environment
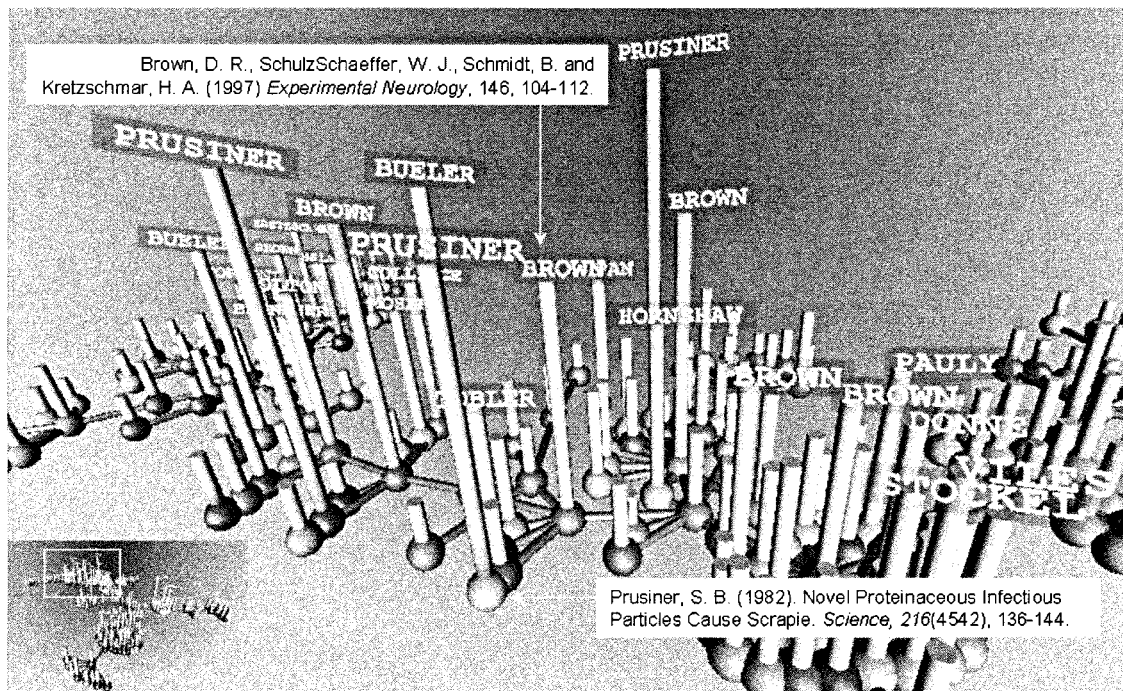
Fig. 11. Expanded knowledge domain, featuring Brown's paper, which is the "exit" landmark, and Prusiner's original paper on prion theory. The inset shows the relative location of this local view in the global context.
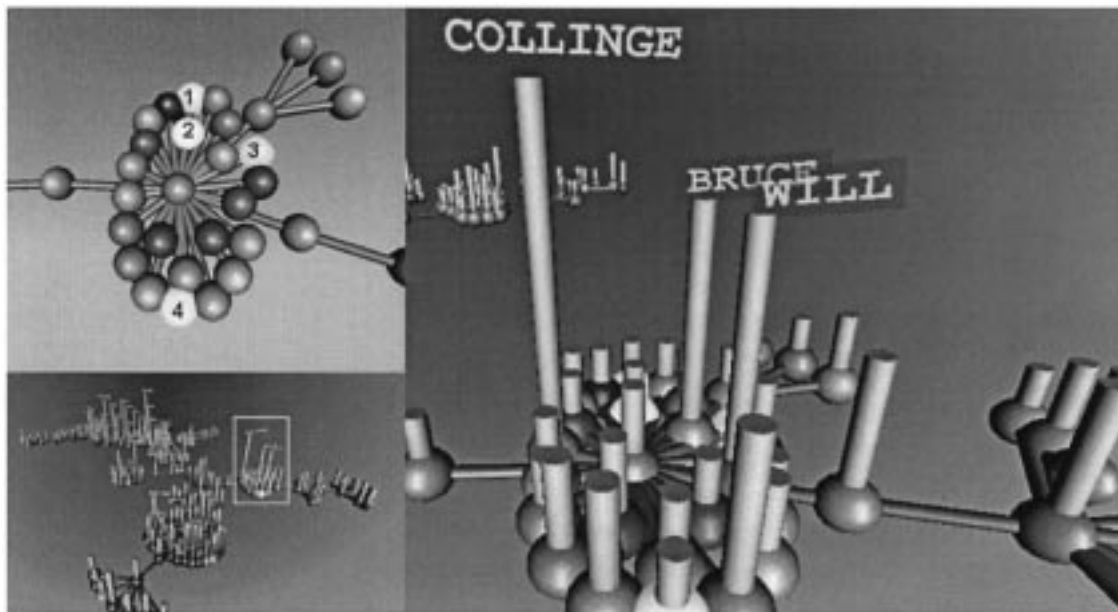

Fig. 12. All Purdey's papers are located in the same cluster as three highly cited papers on possible connections between BSE and CJD. The upper-left frame shows the locations of Purdey's papers in this cluster. The paper numbers are cross references to the bibliographic details in Table IV. The lower-left frame shows the position of this cluster in the global context.

in areas known to have found spongiform diseases, such as Colorado in the U.S., Iceland, Italy, and Slovakia. He found a high level of manganese and low levels of copper in all of them.

Purdey's research on the manganese–copper hypothesis shows the sign of latent domain knowledge. He has published in scientific journals, but they are not highly cited by other researchers (see Table IV). We need to find a gateway from which we can expand the global landscape of mainstream research in BSE and vCJD and place Purdey's research into

the big picture of this issue. Recall that we need to find an "exit" landmark in the global landscape to conduct the domain expansion, but none of Purdey's publications was featured in the scene. To solve this problem, we need to find someone who is active in the manganese–copper paradigm and also included in the mainstream visualization view.

Brown, a biochemist at Cambridge University, Cambridge, U.K., is among scientists who did cite Purdey's publications. Brown provides a good candidate for an "exit" landmark.

On the one hand, Brown is interested in the role of the manganese–copper balance in prion diseases [28] and he cited Purdey's papers. On the other hand, he is interested in Prusiner's prion theory and published about 50 papers on prion diseases. Indeed two of his papers are featured in the mainstream view visualization of the case study. We chose his 1997 paper published in *Experimental Neurology* as the "exit" landmark (see Fig. 11). The citing space to his work is shown in Figs. 10 and 11.

Fig. 11 shows the location of the "exit" landmark paper by Brown. Prusiner's famous 1982 paper on prion theory is also located nearby. Brown's research interest in prion theory is apparent from its closeness to Prusiner's paper. Fig. 12 shows more details of the manganese–copper paradigm, in particular the locations of Purdey's papers. Four of his papers are located within the same cluster as three scientists well cited for their works in vCJD, namely, Collinge, Bruce, and Will.

Because of the relatively LC rates of Purdey's papers, conventional citation analysis is unlikely to take them into account. Predominant papers in this cluster all address the possible link between BSE and vCJD. This observation suggests how Purdey's papers might fit into the mainstream domain knowledge.

We have demonstrated it that our approach can be successfully applied to find connections that would be otherwise obscured. The BSE case study has shown that Purdey's theory is feeding in the mainstream research on BSE and CJD through Brown and his group.

V. DISCUSSIONS AND CONCLUSIONS

In order to track the development of scientific paradigms, it is important to take into account latent domain knowledge as well as mainstream domain knowledge. By incorporating an information visualization procedure originally developed for visualizing mainstream domain knowledge into a recursive process, we have been able to demonstrate the potential of our approach to visualizing not only highly relevant and highly cited documents, but also highly relevant but infrequently cited documents.

Typical citation-based domain visualization approaches have focused on citation frequencies of high-profiled research in a knowledge domain. Consequently, resultant visualizations are strongly biased toward highly cited works. Although highly cited works constitute the core knowledge of a domain, their presence inevitably outshines the presence of latent domain knowledge if they are measured by the same yardstick. The use of two-step citation chains allows us to glean latent domain knowledge and maintain the global picture of where such latent domain knowledge fits.

A natural extension of the research is to explore ways that can combine approaches based on citation patterns and those based on word-occurrence patterns to pinpoint a significant mismatch between the citation strength and word co-occurrence patterns. There are other potentially useful ways to uncover latent domain knowledge. Many techniques developed in scientometrics for quantitative studies of science can be used to generate structural representations of domain knowledge. By comparing and contrasting differences across a variety of structural representa-

tions one can expect to spot missing links and potentially noteworthy connections. For example, if a co-word analysis reveals a strong link between intellectually-related works. In contrast, if such links are absent or weak in citation networks, then it could be important for scientists to know whether they might have overlooked something potentially significant.

On the one hand, visualizing domain knowledge in general is a revival of a long established quest for quantitative studies of scientific discoveries and scientific paradigms, especially due to the advances in enabling techniques such as digital libraries and information visualization. On the other hand, visualizing domain knowledge should set its own research agenda in the new era of science and technology so as to provide valuable devices for scientists, philosophers of science, sociologists of knowledge, librarians, government agencies, and others to grasp crucial developments in science and technology.

In this paper, we have examined the role of citation chains in visualizing latent domain knowledge. The new visualization approach can not only capture the intellectual structure of highly cited works but also make it possible to uncover connections between latent domain knowledge and the body of the mainstream domain knowledge. The two case studies have shown that this approach has the potential to be a new way of supporting knowledge tracking and knowledge management.

REFERENCES

[1] C. Chen, *Information Visualization and Virtual Environments*. London, U.K.: Springer-Verlag London, 1999.
[2] C. Chen and R. J. Paul, "Visualizing a knowledge domain's intellectual structure," *IEEE Computer*, vol. 34, pp. 65–71, 2001.
[3] C. Chen, "Visualization of knowledge structures," in *Handbook of Software Engineering and Knowledge Engineering*, S. K. Chang, Ed, Singapore: World Scientific, 2002, vol. 2, p. 700.
[4] R. W. Schvaneveldt, *Pathfinder Associative Networks: Studies in Knowledge Organization*. ser. Ablex Series in Computational Sciences, D. Partridge, Ed.   Norwood, NJ: Ablex, 1990.
[5] R. W. Schvaneveldt, F. T. Durso, and D. W. Dearholt, "Network structures in proximity data," in *The Psychology of Learning and Motivation, 24*, G. Bower, Ed.   New York: Academic, 1989, pp. 249–284.
[6] C. Chen, "Visualising semantic spaces and author co-citation networks in digital libraries," *Inform. Process. Manage.*, vol. 35, pp. 401–420, 1999.
[7] T. S. Kuhn, *The Structure of Scientific Revolutions*.   Chicago, IL: Univ. Chicago Press, 1962.
[8] H. G. Small, "A co-citation model of a scientific specialty: A longitudinal study of collagen research," *Social Stud. Sci.*, vol. 7, pp. 139–166, 1977.
[9] H. D. White and K. W. McCain, "Visualizing a discipline: An author co-citation analysis of information science, 1972–1995," *J. Amer. Soc. Inform. Sci.*, vol. 49, pp. 327–356, 1998.
[10] C. Chen, T. Cribbin, R. Macredie, and S. Morar, "Visualizing and tracking the growth of competing paradigms: Two case studies," *J. Amer. Soc. Inform. Sci.*, 2001, to be published.
[11] E. Garfield, S. I. H. , and R. J. Torpie, *The Use of Citation Data in Writing the History of Science*.   Philadelphia, PA: Inst. Sci. Inform., 1964.
[12] D. D. Price, "Networks of scientific papers," *Sci.*, vol. 149, pp. 510–515, 1965.
[13] I. Sher and E. Garfield, "New tools for improving and evaluating the effectiveness of research," in *Proc. Res. Program Effectiveness*, Washington, DC, 1966.

[14] H. G. Small and B. C. Griffith, "The structure of scientific literatures—Part I: Identifying and graphing specialties," *Sci. Stud.*, vol. 4, pp. 17–40, 1974.

[15] E. Garfield, "Scientography: Mapping the tracks of science," *Current Contents: Social Behavioral Sci.*, vol. 7, pp. 5–10, 1994.

[16] H. D. White and B. C. Griffith, "Author co-citation: A literature measure of intellectual structure," *J. Amer. Soc. Inform. Sci.*, vol. 32, pp. 163–172, 1981.

[17] D. C. He and L. Wang, "Texture unit, texture spectrum, and texture analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, 1990.

[18] D. R. Swanson, "On the fragmentation of knowledge, the connection explosion, and assembling other people's ideas," *Bull. Amer. Soc. Inform. Sci. Technol.*, vol. 27, pp. 12–14, 2001.

[19] ——, "Fish oil, Raynauds syndrome, and undiscovered public knowledge," *Pers. Biol. Med.*, vol. 30, pp. 7–18, 1986.

[20] D. R. Swanson and N. R. Smalheiser, "An interactive system for finding complementary literatures: A stimulus to scientific discovery," *Artif. Intell.*, vol. 91, pp. 183–203, 1997.

[21] D. R. Swanson, "Migraine and magnesium—Eleven neglected connections," *Pers. Biol. Med.*, vol. 31, pp. 526–557, 1988.

[22] ——, "Somatomedin-C and arginine—Implicit connections between mutually isolated literatures," *Pers. Biol. Med.*, vol. 33, pp. 157–186, 1990.

[23] N. R. Smalheiser and D. R. Swanson, "Assessing a gap in the biomedical literature—Magnesium-deficiency and neurologic disease," *Neurosci. Res. Commun.*, vol. 15, pp. 1–9, 1994.

[24] ——, "Indomethacin and Alzheimer's disease," *Neurol.*, vol. 46, p. 583, 1996.

[25] D. R. Swanson, "Computer-assisted search for novel implicit connections in text databases," *Abstr. Papers Amer. Chem. Soc.*, vol. 217, p. 010-CINF, 1999.

[26] C. Chen, "Generalized similarity analysis and pathfinder network scaling," *Interacting Comput.*, vol. 10, pp. 107–128, 1998.

[27] E. Stourton, *Mad Cows and an Englishman*. London, U.K.: BBC2, 2001.

[28] D. R. Brown, F. Hafiz, L. L. Glasssmith, B. S. Wong, I. M. Jones, C. Clive, and S. J. Haswell, "Consequences of manganese replacement of copper for prion protein function and proteinase resistance," *EMBO J.*, vol. 19, pp. 1180–1186, 2000.

Chaomei Chen received the Ph.D. degree in computer science from the University of Liverpool, Liverpool, U.K., in 1995.

Currently, he is an Associate Professor at the College of Information Science and Technology, Drexel University, Philadelphia, PA. His research interests include information visualization, knowledge discovery, domain visualization, human–computer interaction, and hypertext. .

Dr. Chen is a Member of the ACM, the IEEE Computer Society, and the American Society for Information Science and Technology.

Jasna Kuljis received the Ph.D. degree in information systems from the London School of Economics, University of London, London, U.K., in 1995.

Currently, she is a Senior Lecturer at the Department of Information Systems and Computing, Brunel University, Uxbridge, U.K. Her current research interests include human–computer interfaces, graphical user interface design, and the usability of interactive computer systems.

Ray J. Paul received the Ph.D. degree in operational research from Hull University, Hull, U.K., in 1974.

Currently, he is a Professor of simulation modeling in the Department of Information Systems and Computing, Brunel University, Uxbridge, U.K. His research interests are simulation modeling processes, software environments for simulation modeling, and information systems development.

Dr. Paul is a Member of the ACM, the IEEE Computer Society, the Operational Research Society, the British Computer Society, and the Society for Computer Simulations. He is the European Chapter Chairman of ACM SIGSIM.