

Mapping Scientific Frontiers

Chaomei Chen

Mapping Scientific Frontiers

The Quest for Knowledge Visualization

Second Edition



Springer

Chaomei Chen
College of Information Science and Technology
Drexel University
Philadelphia, Pennsylvania
USA

ISBN 978-1-4471-5127-2 ISBN 978-1-4471-5128-9 (eBook)
DOI 10.1007/978-1-4471-5128-9
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2013944066

© Springer-Verlag London 2003, 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Mapping science at first glance appears to be an oxymoron: how can we map something as abstract as science? Scientific knowledge seems to occupy an intellectual realm which can only be glimpsed by the mind of the highly trained specialist. Yet this book demonstrates that the discipline of science mapping has been going on for a number of years and is indeed beginning to flourish with new results and insights. In this endeavor we may only be in the stage of the early explorers who drew out the first crude geographic maps of the then known physical world. We might argue that science mapping is simply the logical progression of map making from the physical to the intellectual world.

It is not necessary to make the case for the importance of science in the modern world and the object of our map making efforts, even though our society has at times minimized the role of scientific research, neglected to provide adequate funding and support, and attempted to retard its educational programs. What more important artifact of human intelligence could we focus on than the current scientific landscape? Of course science can be used for good or ill, and science confers incredible power on those who master its principles. On occasion individual scientists will abuse the trust we place in them in their quest for recognition, as Chaomei Chen documents here in his study of “retractions” of scientific articles. But despite these aberrations, science is the gateway to understanding of our place in the universe, and the foundation of our social and economic well-being.

Despite the fact that the language we use to describe science is replete with spatial metaphors such as “field” and “area” of research, when we actually go about trying to create a map of science, we soon realize that the procedures used to make geographic maps no longer apply. We must deal with the abstract relations and associations among entities such as scientific ideas, specialties, fields or disciplines whose very existence may be open to debate. Thomas Kuhn described these shifting historical disciplines in a revealing interview: “Look . . . you must not use later titles for [earlier] fields. And it’s not only the ideas that change, it’s the structure of the disciplines working on them.” (Kuhn 2000, p. 290) Here we realize that science in any historical period, and indeed the current period, is a *terra incognita*. Are we

justified in seeking a spatial representation of such abstract and perhaps hypothetical entities? Are our brains hardwired to take what is relational and project it in real space?

Perhaps science mapping is difficult to grasp because three conceptual steps are required before it makes sense, two of which involve some degree of mathematical manipulation. First a unit of analysis must be chosen which should comprise the elementary particles of our science universe. Secondly a measure of association between the units must be defined. Thirdly, a means must be found for depicting the units and their relations in a low dimensional space (usually two dimensions) that can be perceived by humans. Once these intellectual leaps are made science mapping seems natural and even inevitable.

The established scholarly disciplines of the history of science, sociology of science and philosophy of science have long been regarded as providing the essential tools for understanding the origin and evolution of scientific thought, its social institutions, and its philosophical underpinnings. With the exception of some early work in the sociology of science, the methods used by these disciplines have been largely qualitative in nature. History of science has dealt largely with constructing narratives using the methods of general history, and philosophy of science with logical foundation and epistemology. Sociology of science, as exemplified by one of its founders Robert Merton, was both strongly theoretical and open to the use of quantitative evidence. This approach was also taken up by early social network researchers who studied so-called invisible colleges. In the 1970s, however, sociology of science turned away from quantitative methods to embrace more radical social theories of science, such as the social construction of scientific knowledge, and largely abandoned the quantitative approach of the earlier sociologists.

More recently, primarily as the result of the availability of large databases of scientific publications and citations and partly a reaction against constructivist sociology, a discipline emerged which has opened up a new way to study the evolution of science. This field has been called variously scientometrics, informetrics and bibliometrics. These terms reflect not only the focus on quantitative methods upon which it is built, but also its origins in what was once called library science. It cannot be claimed that the upstart discipline has achieved acceptance in the academic world, particularly on the part of the more established disciplines, and institutionalization in the form of university programs and academic positions has only just begun. Critics of scientometrics have claimed that a focus on the scientific literature as its primary source material too severely limits the data on which studies of science can be based. On the other hand, the increasing availability of the full text of scientific papers in computer readable formats, opens up many new types of data for analyses which when used in tandem with the standard on-line databases goes far beyond what has been possible using the standard indexes alone. Combined with software packages such as Chaomei Chen has pioneered, a powerful new tool for the analysis of science has come into being. There is every indication that the new field is here to stay and is exerting more and more influence on policy, even though a rapprochement and integration with the traditional fields of history, sociology and philosophy is probably a long way off.

Chaomei Chen's book is important because it builds on many of the concepts and findings of history, sociology and philosophy of science, but at the same time adds a new dimension. As an example of the power of the new methods the skeptical reader should consult chapter eight presenting a case study on recent work on induced pluripotent stem-cells which shows how mapping can inform historical studies as well as assist medical researchers to get an overview of a research area. Here we see the strength of the new methods for exploring and tracking the internal structure of revolutionary developments in contemporary science.

His book also draws on an even broader disciplinary framework from computer to information science and particularly information visualization. In the first edition Chaomei Chen commented on the disciplines that contribute ideas to science mapping. "Different approaches to mapping scientific frontiers over recent years are like streams running from several different sources A lot of work needs to be done to cultivate knowledge visualization as a unifying subject matter that can join several disciplines." (Chen 2003, p. vii) This remains true even today when scientometrics, computer science, and network science continue to evolve in a strangely independent manner yet often dealing with the same underlying data and issues. This may be an inevitable side effect of the barriers between disciplines, but hopefully this book will help bridge these various streams.

As an example of the relevance of history of science, Chaomei Chen comments that the work of Thomas Kuhn was an important backdrop to mapping because one could think of the unfolding of a revolution in science as a series of cross-sectional maps that at some point undergoes a radical structural transformation. Cross sectional thinking is also very much encouraged in the history of science because historians are exhorted to understand the ideas of a historical period by entering its mind-set, "to think as they did" (Kuhn 1977, p. 110), and not interpret older science in terms of our "current" understanding. This is a difficult requirement because once we know that a new discovery or finding has occurred it is extremely difficult for us not to be influenced by it, and our first impulse is to find precursors and antecedents. As analysts we need to take care not to allow the present to distort the past.

As an example of how various cross-currents converge in science mapping we could point out the tension between psychological factors, as exemplified by Kuhn's gestalt switching as a way of looking at conceptual change, and social forces such as collegial networks and invisible colleges. Do social relations determine cognitive relations, or vice versa? In Stanley Milgram's early work (1967) on social networks, human subjects were required to think about what acquaintances their acquaintances had several steps removed. In Don Swanson's work (1987) on undiscovered public knowledge, discoveries are made by seeking concepts that are indirectly related through other concepts that are currently unconnected. Thus the same type of thinking is involved in both the social and intellectual tasks. If we are dealing with words or references as our mapping units, then psychology clearly enters the picture because an author's memory and recall are involved in the associative process. But that memory and recall are also influenced by what authors have seen other authors or colleagues say. If we map individual scientists in their co-author relations,

then social factors must come into play but psychological factors also contribute to the selection of co-authors. Thus social and psychological factors are inexorably intertwined in both the social and intellectual structure of science.

The competition in science mapping between the various choices for unit of analysis such as words, references, authors, journals, etc. and the means of associating them such as co-word, co-citation, co-authorship, direct citation, etc. seems to boil down to the types of structures and level of relations we want to observe. To better understand the role of discovery in specialty development we might turn to co-citations because many discoveries are associated with specific papers and authors. On the other hand, if we want to include broader societal, non-scholarly factors then we might turn to co-words which can more readily capture public or political sentiments external to science. Journals, a yet broader unit of analysis, might best represent whole fields or disciplines. Choice of a unit of analysis also depends on the historical period under investigation. Document co-citation is probably not feasible prior to 1900 due to the absence of standard referencing practice. However, name co-mention within the texts of scientific papers and books is still very feasible for earlier periods. It is instructive to try to imagine how we would carry out a co-mention or other kind of mapping for some earlier era, say for scientific literature in the eighteenth century and whether we would be able to identify the schools of thought and rival paradigms active during the period.

Another important issue is the interpretation of maps. We know that the network of associations that underlies maps is hyper dimensional, and that projection in two dimensions is inevitably an approximation and can place weakly related units close together. This argues for the need to pay close attention to the links themselves which give rise to the two dimensional solution in the first place, which we can think of as the neurons of the World Brain (Garfield 1968) we are trying to visualize. Only by knowing what the links signify can we gain a better understanding of what the maps represent. This will involve looking more deeply at the context in which the linking takes place, and seeking new ways of representing and categorizing those relationships, for example, by function or type such as logical, causal, social, hypothetical, metaphorical, etc. One positive development in this direction, as described in the final chapter, is the advent of systems for “visual analytics” that allow us to more deeply probe the underpinnings of maps with the ultimate goal of supporting decision making.

Part of what is exciting about science mapping is that the landscape is continually changing: every year there is a new crop of papers and the structure changes as new areas emerge and existing areas evolve or die off. Some will find such a picture unsettling and would prefer to see science as a stable and predictable enterprise, but as Merton has argued (2004), serendipity is endemic to science, and thus also to science maps. We do not yet know if discovery is in any way predictable, if there are recognizable antecedents or conditions, or whether discovery or creativity can be engineered to happen at a quicker pace. But because discoveries are readily apparent in maps after they occur, we also have the possibility of studying maps for previous time periods to look for structural antecedents.

This is a wide ranging book on information visualization, with a specific focus on science mapping. Science mapping is still in its infancy and many intellectual challenges remain to be investigated and many of which are outlined in the final chapter. In this new edition Chaomei Chen has provided an essential text, useful both as a primer for new entrants and as a comprehensive overview of recent developments for the seasoned practitioner.

SciTech Strategies, Inc.

Henry Small

References

- Chen C (2003) Mapping scientific frontiers. Springer, London
- Garfield E (1968) "World Brain" or Memex? Mechanical and intellectual requirements for universal bibliographic control. In: Montgomery EB (ed) The foundations of access to knowledge. Syracuse University Press, Syracuse, pp 169–196, from <http://garfield.library.upenn.edu/essays/v6p540y1983.pdf>
- Kuhn TS (1977) The essential tension. University of Chicago Press, Chicago
- Kuhn TS (2000) The road since structure. University of Chicago Press, Chicago
- Merton RK, Barber E (2004) The travels and adventures of serendipity. Princeton University Press, Princeton
- Milgram S (1967) The small world problem. Psychol Today 2:60–7
- Swanson DR (1987) Two medical literatures that are logically but not bibliographically connected. J Am Soc Info Sci 38:228–233

Preface for the 2nd Edition

The first edition of *Mapping Scientific Frontiers* (MSF) was published over 10 years ago in 2002. Since then, a lot has changed. Social media has flourished to the extent that we have never seen before. News, debates, hoaxes, and scholarly blogs all fight for attention on Facebook (launched in 2004), YouTube (2005), and Twitter (2006), which are made ubiquitously accessible by popular mobile devices such as iPhone (2007) and iPad (2010).

Over the past 10 years, remarkable scientific breakthroughs have been made, for example, Grigori Perelman's proof of the century-old Poincaré Conjecture in 2002, the Nobel Prize winning research on induced pluripotent stem cells (iPSCs) by Shinya Yamanaka and his colleagues since 2006, and the recent discovery of the Higgs Boson in 2012 at CERN.

The big sciences continue to get bigger. Large-scale data collection efforts for scientific research such as the Sloan Digital Sky Survey (SDSS) (2000–2014) in astronomy represent one of many sources of big data. As old scientific fields transform themselves, new ones emerge. Visual analytics entered our horizon in 2005 as a new field and has played a critical role ever since in advancing the science and technology for solving practical issues, especially when we deal situations that are full of complex, uncertain, incomplete, and potentially conflicting data. A representative case is concerned with maintaining the integrity of scientific literature itself. The increasing number of publications has overshadowed the increase of retractions. What can be done to maintain a trustworthy body of scientific knowledge?

What is the role that *Mapping Scientific Frontiers* has played? According to Google Scholar, it has been cited by 235 scientific publications on the web. These publications are in turn cited by an even broader range of articles. These articles allow us to take a glimpse on the context in which research in science mapping has been evolving. Interestingly, the citation profile appears to show two stages. The first one ranges from 2002 to 2008 and the second one from 2009 to the present (Fig. 1). Citations in the first stage peaked in 2007, whereas citations in the second stage were evenly distributed for the first 3 years. A study of citation data in the Web of Science revealed a similar pattern.



Fig. 1 The citation profile of *Mapping Scientific Frontiers* (Source: Google Scholar)

What is the citation pattern telling us? The nature of the set of articles that cited *Mapping Scientific Frontiers* as a whole can be analyzed in terms of how they are in turn cited by subsequently published articles. In particular, we turn to articles that have strong citation bursts, or abruptly increased citation rates, during the time span of 2002–2013. Figure 2 shows 25 articles of this type. Articles in the first stage shared a unique focus on information visualization and citation analysis. The original motivation of *Mapping Scientific Frontiers* was indeed to bridge together the two fields across the boundaries of different disciplines.

The second stage is predominated by a series of publications dedicated to global science maps at disciplinary levels as opposed to the document level in the first stage. The most influential work in the second stage in terms of citation burst is a 2009 *Scientometrics* article by Alan L. Porter and Ismael Rafolsson on the interdisciplinarity of science. The second highest citation burst is attributed to a 2010 article published in the *Journal of American Society for Information Science and Technology* by Ismael Rafols, Alan L. Porter, and Loet Leydesdorff on science overlay maps. We are still in the second stage. In terms of the scale and the unit of analysis, the study of interdisciplinary interactions is a profound and potentially fruitful way to better understand the dynamics of scientific frontiers.

In addition to the conceptual and theoretical development, researchers today have a much wider range of choice than before in terms of computational tools for analyzing, visualizing, and exploring patterns and trends in scientific literature. Notable examples include CiteSpace, HistCite, VOSViewer, and Sci² for scientometric studies and science mapping; GeoTime, Jigsaw, and Tableau for visual analytics; and Gephi, Alluvial Maps, D3, and WebGL for more generic information visualization. Today, a critical mass is taking its shape and gathering its strengths as visual analytic tools, data sources, and exemplars of in-depth and longitudinal studies become increasingly accessible and inter-operable. Mapping Scientific Frontiers has reached a new level with a broad range of unprecedented opportunities to impact scientific activity across so many disciplines.

The second edition of *Mapping Scientific Frontiers* brings you some of the most profound discoveries and advances in the study of scientific knowledge and the dynamics of its evolution. Some of the new additions are highlighted as follows:

Top 25 References with Strongest Citation Bursts

References	Year	Strength	Begin	End	2002 - 2013
CHEN C, 2003, MAPPING SCI FRONTIER, V, P	2003	7.2508	2002	2006	
CARD SK, 1999, READINGS INFORMATION, V, P	1999	4.5742	2002	2006	
CHEN CM, 2002, J AM SOC INF SCI TEC, V53, P678, DOI	2002	4.1649	2002	2008	
CHEN CM, 2001, J AM SOC INF SCI TEC, V52, P315, DOI	2001	4.0433	2002	2004	
CHEN CM, 2001, IEEE T SYST MAN CY C, V31, P518, DOI	2001	4.0072	2002	2004	
CHEN CM, 2001, COMPUTER, V34, P65	2001	3.5648	2002	2005	
WHITE HD, 1997, ANNU REV INFORM SCI, V32, P99	1997	3.988	2003	2008	
GARFIELD E., 1964, USE CITATION DATA WR, V, P	1964	3.5679	2003	2008	
BATAGELJ V, 1998, CONNECTIONS, V21, P47	1998	3.6708	2004	2006	
CHEN CM, 2003, J AM SOC INF SCI TEC, V54, P435, DOI	2003	3.5478	2004	2006	
SUI DZ, 2004, PROF GEOGR, V56, P62	2004	3.7103	2005	2007	
KLAVANS R, 2006, J AM SOC INF SCI TEC, V57, P251, DOI	2006	5.0919	2009	2009	
KLAVANS R, 2006, SCIENTOMETRICS, V68, P475, DOI	2006	4.213	2009	2010	
SHIFFRIN RM, 2004, P NATL ACAD SCI USA, V101, P5183, DOI	2004	3.4738	2009	2009	
KLAVANS R, 2009, J AM SOC INF SCI TEC, V60, P455, DOI	2009	5.0572	2010	2013	
PORTER AL, 2009, SCIENTOMETRICS, V81, P719, DOI	2009	12.7909	2011	2013	
LEYDESORFF L, 2009, J AM SOC INF SCI TEC, V60, P348, DOI	2009	7.3836	2011	2013	
VAN ECK NJ, 2010, SCIENTOMETRICS, V84, P523, DOI	2010	6.4516	2011	2013	
LEYDESORFF L, 2010, J AM SOC INF SCI TEC, V61, P1622, DOI	2010	6.4334	2011	2013	
RAFOLS I, 2010, SCIENTOMETRICS, V82, P263, DOI	2010	4.6479	2011	2013	
PORTER AL, 2009, J NANOPART RES, V11, P1023, DOI	2009	4.1749	2011	2013	
CHEN CM, 2010, J AM SOC INF SCI TEC, V61, P1386, DOI	2010	3.9242	2011	2013	
VAN RAAN A. F. J., 2005, MEASUREMENT, V3, P1	2005	3.5074	2011	2011	
RAFOLS I, 2010, J AM SOC INF SCI TEC, V61, P1871, DOI	2010	8.1778	2012	2013	
WAGNER CS, 2011, J INFORMETR, V5, P14, DOI	2011	4.0553	2012	2013	

Fig. 2 A citation analysis of *Mapping Scientific Frontiers* reveals two stages of relevant research. Red bars indicate intervals of citation burst

- The Sloan Digital Sky Survey (SDSS) is featured in Chap. 2 in the context of how a map of the Universe may reveal.
- In Chap. 3, a series of new examples of visualizing a thematic evolution over time are illustrated, including the widely known ThemeRiver, the elegant TextFlow, and the versatile Alluvial Maps.
- Chapter 8 is a new chapter. It introduces the framework of a predictive analysis and demonstrates how it can apply to a fast-advancing field such as regenerative medicine, which highlights the work that was awarded the 2012 Nobel Prize in medicine on induced pluripotent stem cells (iPSCs). Chapter 8 also addresses practical implications of the retraction of a scientific publication. The second half of Chap. 8 is devoted to the design, construction, and analysis of global science maps, including our own new design of dual-map overlays.

- Chapter 9 is also a new chapter. It outlines some of the most representative visual analytic tools such as GeoTime and Jigsaw. It also describes major analytic features of CiteSpace.

The first edition concludes with ten challenges ahead. It is valuable to revisit these challenges identified over 10 years ago and see what have changed and what have newly emerged.

The second edition finishes with a set of new challenges and milestones ahead for mapping scientific frontiers.

15 April 2013
Villanova
Pennsylvania, USA

Chaomei Chen

Acknowledgements

The second edition in part reflects the result of a continuous research effort that I have been engaged in since the publication of the first edition. I'd like to acknowledge the support and contributions of my colleagues, students, and collaborators in various joint projects and publications, in particular, including my current and former students Timothy Schultz (Drexel University, USA), Jian Zhang (IBM Shanghai, China), and Donald A. Pellegrino (The Dow Chemical Company, USA), collaborators such as Pak Chung Wong (PNNL, USA), Michael S. Vogeley (Drexel University, USA), Alan MacEachren (Penn State University, USA), Jared Milbank (Pfizer, USA), Loet Leydesdorff (The Netherlands), Richard Klavans, Kevin Boyack, and Henry Small (SciTech Strategies, USA), and Hong Tseng (NIH, USA).

As a Chang Jiang Scholar, I have the opportunity to work with the WISELab at Dalian University of Technology, China, since 2008. I'd like to acknowledge the collaboration with Zeyuan Liu, Yue Chen, Zhigang Hu, and Shengbo Liu. Yue Chen is currently leading an ambitious effort to translate the second edition to Chinese.

I particularly appreciate the opportunity to work with Rod Miller, Chief Strategy Officer at the iSchool of Drexel, and Paul Dougherty, Licensing Manager at the Office of Technology Commercialization of Drexel University, over numerous fruitful discussions of various research topics.

I'd like to acknowledge the support of sponsored research and grants from the National Science Foundation (IIS-0612129, NSFDACS-10P1303, IIP 1160960), Department of Homeland Security, Pfizer, and IMS Health. I'd also like to take the opportunity to express my gratitude and appreciation to the hosts of my talks and keynote speeches, including Michael Dietrich, History of Biology, Woods Hole, MA; Stephanie Shipp, International Defense Agency (IDA), Washington D.C.; Paula Fearon, NIH, Bethesda, MD; David Chavalarias, Mining the Digital Traces of Science (MDTS), Paris, France; and Josiane Mothe, Institut de Recherche en Informatique de Toulouse, France.

Special thanks to Beverley Ford, Editorial Director Computer Science, Springer London Ltd, for her initiative and encouragement for getting the second edition on my agenda, and to Ben Bishop, Senior Editorial Assistant, Springer, for his clear and professional guidance that ensures a smooth and enjoyable process of preparation.

As always, my family, Baohuan, Calvin, and Steven, gives me all the strengths, the courage, and the inspirations.

Contents

- 1 The Dynamics of Scientific Knowledge** 1
 - 1.1 Scientific Frontiers 2
 - 1.1.1 Competing Paradigms 6
 - 1.1.2 Invisible Colleges 10
 - 1.1.3 Conceptual Revolutions 11
 - 1.1.4 TRACES 16
 - 1.2 Visual Thinking 20
 - 1.2.1 Gestalt 20
 - 1.2.2 Famous Maps 22
 - 1.2.3 The Tower of Babel 23
 - 1.2.4 Messages to the Deep Space 25
 - 1.2.5 “Ceci n’est pas une Pipe” 30
 - 1.2.6 Gestalt Psychology 34
 - 1.2.7 Information Visualization and Visual Analytics..... 35
 - 1.3 Mapping Scientific Frontiers 39
 - 1.3.1 Science Mapping..... 40
 - 1.3.2 Cases of Competing Paradigms 41
 - 1.4 The Organization of the Book 43
 - References 45
- 2 Mapping the Universe** 47
 - 2.1 Cartography 47
 - 2.1.1 Thematic Maps..... 52
 - 2.1.2 Relief Maps and Photographic Cartography..... 53
 - 2.2 Terrestrial Maps 54
 - 2.3 Celestial Maps..... 56
 - 2.3.1 The Celestial Sphere Model 58
 - 2.3.2 Constellations 63
 - 2.3.3 Mapping the Universe 66
 - 2.4 Biological Maps 77
 - 2.4.1 DNA Double Helix 77

2.4.2	Acupuncture Maps	79
2.4.3	Genomic Maps	81
2.4.4	A Map of Influenza Virus Protein Sequences	82
	References	84
3	Mapping Associations	85
3.1	The Role of Association	85
3.1.1	As We May Think	86
3.1.2	The Origin of Cognitive Maps	87
3.1.3	Information Visualization	91
3.2	Identifying Structures	91
3.2.1	Topic Models.....	91
3.2.2	Pathfinder Network Scaling	93
3.2.3	Measuring the Similarity Between Images	95
3.2.4	Visualizing Abstract Structures	101
3.2.5	Visualizing Trends and Patterns of Evolution	107
3.3	Dimensionality Reduction	111
3.3.1	Geometry of Similarity	113
3.3.2	Multidimensional Scaling	114
3.3.3	INDSCAL Analysis	119
3.3.4	Linear Approximation – Isomap	121
3.3.5	Locally Linear Embedding	124
3.4	Concept Mapping	127
3.4.1	Card Sorting.....	127
3.4.2	Clustering	128
3.5	Network Models	131
3.5.1	Small-World Networks	131
3.5.2	The Erdős-Renyi Theory	133
3.5.3	Erdős Numbers.....	134
3.5.4	Semantic Networks	135
3.5.5	Network Visualization	136
3.6	Summary	138
	References	139
4	Trajectories of Search	143
4.1	Footprints in Information Space	143
4.1.1	Traveling Salesman Problem	144
4.1.2	Searching in Virtual Worlds	146
4.1.3	Information Foraging	148
4.1.4	Modeling a Foraging Process	149
4.1.5	Trajectories of Users	154
4.2	Summary	160
	References	161
5	The Structure and Dynamics of Scientific Knowledge	163
5.1	Matthew Effect	164

5.2	Maps of Words	167
5.2.1	Co-Word Maps	167
5.2.2	Inclusion Index and Inclusion Maps	168
5.2.3	The Ontogeny of RISC	170
5.3	Co-Citation Analysis	172
5.3.1	Document Co-Citation Analysis	172
5.3.2	Author Co-Citation Analysis	180
5.4	HistCite	190
5.5	Patent Co-Citations	193
5.6	Summary	195
	References	197
6	Tracing Competing Paradigms	201
6.1	Domain Analysis in Information Science	201
6.2	A Longitudinal Study of Collagen Research	203
6.3	The Mass Extinction Debates	206
6.3.1	The KT Boundary Event	206
6.3.2	Mass Extinctions	209
6.4	Supermassive Black Holes	218
6.4.1	The Active Galactic Nuclei Paradigm	218
6.4.2	The Development of the AGN Paradigm	219
6.5	Summary	224
	References	225
7	Tracking Latent Domain Knowledge	227
7.1	Mainstream and Latent Streams	228
7.2	Knowledge Discovery	229
7.2.1	Undiscovered Public Knowledge	230
7.2.2	Visualizing Latent Domain Knowledge	234
7.3	Swanson's Impact	239
7.4	Pathfinder Networks' Impact	240
7.4.1	Mainstream Domain Knowledge	241
7.4.2	Latent Domain Knowledge	242
7.5	BSE and vCJD	248
7.5.1	Mainstream Domain Knowledge	248
7.5.2	The Manganese-Copper Hypothesis	254
7.6	Summary	255
	References	256
8	Mapping Science	259
8.1	System Perturbation and Structural Variation	259
8.1.1	Early Signs	260
8.1.2	A Structural Variation Model	262
8.1.3	Structural Variation Metrics	265
8.1.4	Statistical Models	269
8.1.5	Complex Network Analysis (1996–2004)	271

8.2	Regenerative Medicine	274
8.2.1	A Scientometric Review	275
8.2.2	The Structure and Dynamics	277
8.2.3	System-Level Indicators	281
8.2.4	Emerging Trends	286
8.2.5	Lessons Learned	288
8.3	Retraction	290
8.3.1	Studies of Retraction	293
8.3.2	Time to Retraction	296
8.3.3	Retracted Articles in Context	297
8.3.4	Autism and Vaccine	301
8.3.5	Summary	304
8.4	Global Science Maps and Overlays	304
8.4.1	Mapping Scientific Disciplines	305
8.4.2	Interdisciplinarity and Interactive Overlays	308
8.4.3	Dual-Map Overlays	312
	References	316
9	Visual Analytics	321
9.1	CiteSpace	321
9.2	Jigsaw	325
9.3	Carrot	329
9.4	Power Grid Analysis	329
9.5	Action Science Explorer (iOpener)	331
9.6	Revisit the Ten Challenges Identified in 2002	332
9.7	The Future	338
	References	339
	Index	341

Abbreviations

2dF	Two-degree field spectrograph
AAG	The Association of American Geographers
ACA	Author co-citation analysis
AGN	Active galactic nuclei
ANT	Actor Network Theory
ASE	Action Science Explorer
ASIS&T	American Society for Information Science and Technology
BSE	Bovine Spongiform Encephalopathy
CAS	Complex adaptive system
CBIR	Content-based image retrieval
CfA	Harvard-Smithsonian Center for Astrophysics
CJD	Creutzfeldt-Jakob Disease
C _{KL}	Centrality divergence
CL	Cluster linkage
DCA	Document co-citation analysis
DMO	Dual-map overlay
DNA	Deoxyribonucleic acid
DoD	Department of Defense
EOBT	Expert Opinion on Biological Therapy
GCS	Global Citation Score
GSA	Generalized Similarity Analysis
HDF	Hubble Deep Field
HMM	Hidden Markov Model
HUDF	Hubble Ultra Deep Field
iPSC	Induced pluripotent stem cell
KT Boundary	Cretaceous-Tertiary Boundary
LCS	Local Citation Score
LGL	Large Graph Layout
LLE	Locally linear embedding
LSI	Latent Semantic Indexing
MCR	Modularity change rate

MDS	Multidimensional scaling
MST	Minimum spanning tree
NB	Negative binomial
NVAC	National Visualization and Analytics Center
PCA	Principle component analysis
PFNET	Pathfinder network
PNNL	Pacific Northwest National Laboratory
PTSD	Post-traumatic stress disorder
SCI	Science Citation Index
SDSS	Sloan Digital Sky Survey
SOM	Self-organized maps
SSCI	Social Science Citation Index
SVD	Singular value decomposition
TRACES	Technology in Retrospect and Critical Events in Science
TREC	The Text Retrieval Conference
TSP	Traveling salesman problem
USPTO	The United States Patent and Trademark Office
WECC	The Western Power Grid
WoS	The Web of Science
XDF	eXtreme Deep Field
ZINB	Zero-inflated negative binomial

List of Figures

Fig. 1	The citation profile of <i>Mapping Scientific Frontiers</i> (Source: Google Scholar)	xii
Fig. 2	A citation analysis of <i>Mapping Scientific Frontiers</i> reveals two stages of relevant research. <i>Red bars</i> indicate intervals of citation burst	xiii
Fig. 1.1	Conceptual change: a new conceptual system #2 is replacing an old one #1	12
Fig. 1.2	Computer-generated “best fit” of the continents. There are several versions of this type of fit maps credited to the British geophysicists E.C. Bullard, J.E. Everett, and A.G. Smith	14
Fig. 1.3	Wegener’s conceptual system (<i>top</i>) and the contemporary one (<i>bottom</i>)	15
Fig. 1.4	The conceptual structure of Wegener’s continental drift theory ...	16
Fig. 1.5	The conceptual structure of Wegener’s opponents	17
Fig. 1.6	Pathways to the invention of the video tape recorder (© Illinois Institute of Technology)	19
Fig. 1.7	Alexander Fleming’s penicillin mould, 1935 (© Science Museum, London)	21
Fig. 1.8	Minard’s map (Courtesy of http://www.napoleonic-literature.com)	22
Fig. 1.9	Map of Cholera deaths and locations of water pumps (Courtesy of National Geographic).....	24
Fig. 1.10	The Tower of Babel (1563) by Pieter Bruegel. Kunsthistorisches Museum Wien, Vienna. (Copyright free, image is in the public domain)	25
Fig. 1.11	The Tower of Babel by Maurits Escher (1928)	26
Fig. 1.12	The gold-plated aluminum plaque on Pioneer spacecraft, showing the figures of a man and a woman to scale next to a line silhouette of the spacecraft	27

Fig. 1.13	Voyagers' message	29
Fig. 1.14	Instructions on Voyager's plaque	30
Fig. 1.15	René Magritte's famous statement	31
Fig. 1.16	The first X-ray photograph, produced by Röntgen in 1895, showing his wife's hand with a wedding ring	32
Fig. 1.17	A Gestalt switch between figure and ground. Does the figure show a vase or two faces?	33
Fig. 1.18	Is this a young lady or an old woman?	35
Fig. 2.1	Scenes in the film Powers of Ten (Reprinted from http://www.powersof10.com/film © 2010 Eames Office)	48
Fig. 2.2	The procedure of creating a thematic map	49
Fig. 2.3	The visual hierarchy. Objects on the map that are most important intellectually are rendered with the greatest contrast to their surroundings. Less important elements are placed lower in the hierarchy by reducing their edge contrasts. The <i>side view</i> in this drawing further illustrates this hierarchical concept.....	52
Fig. 2.4	Four types of relief map: (a) contours, (b) contours with hill shading, (c) layer tints, and (d) digits (Reprinted from http://www.nottingham.ac.uk/education/maps/relief.html#r5)	53
Fig. 2.5	A Landsat photograph of Britain (<i>left</i>). Central London (<i>right</i>) is shown as the <i>blue</i> area near to the lower right corner. The Landsat satellite took the photo on May 23rd, 2001 (Reprinted from http://GloVis.usgs.gov/ImgViewer.jsp?path=201&row=24&pixelSize=1000)	54
Fig. 2.6	Ptolemy's world map, re-constructed based on his work Geography c. 150 (© The British Library http://www.bl.uk/)	55
Fig. 2.7	A road map and an aerial photograph of the Westminster Bridge in London	55
Fig. 2.8	London Underground map conforms to the geographical configuration	56
Fig. 2.9	London underground map does not conform to the geographical configuration	57
Fig. 2.10	Atlas with the celestial sphere on his shoulders. This is the earliest surviving representation of the classical constellations (Courtesy of www.cosmopolis.com)	59
Fig. 2.11	Most of the 48 classical constellation figures are shown, but not the stars comprising each constellation. The Farnese Atlas, 200 BC from the National Maritime Museum, London.....	59
Fig. 2.12	Constellations in the northern Hemisphere in 1795s. The Constellations of Eratosthenes.....	60

Fig. 2.13	Constellations in the southern hemisphere in 1795s. The Constellations of Eratosthenes.....	61
Fig. 2.14	The painting of constellations by an unknown artist in 1575 on the ceiling of the Sala del Mappamondo of the Palazzo Farnese in Caprarola, Italy. Orion the Hunter and Andromeda are both located to the right of the painting (Reprinted from Sesti 1991)	62
Fig. 2.15	<i>Left:</i> M-31 (NGC-224) – the Andromeda Galaxy; <i>Right:</i> The mythic figure Andromeda	64
Fig. 2.16	Perseus and Andromeda constellations in John Flamsteed’s Atlas Coelestis (1729) (Courtesy of http:// mahler.brera.mi.astro.it/)	65
Fig. 2.17	Taurus and Orion in John Flamsteed’s Atlas Coelestis (1729) (Courtesy of http://mahler.brera.mi.astro.it/)	65
Fig. 2.18	Orion the Hunter (Courtesy of http://www.cwrl.utexas. edu/~syverson/)	66
Fig. 2.19	Large-scale structures in the Universe (Reprinted from Scientific American, June 1999).....	67
Fig. 2.20	The CfA Great Wall – the structure is 500 million light-years across. The Harvard-Smithsonian Center for Astrophysics redshift survey of galaxies in the northern celestial hemisphere of the universe has revealed filaments, bubbles, and, arching across the middle of the sample	68
Fig. 2.21	Slice through the Universe (Reprinted from Scientific American, June 1999).....	69
Fig. 2.22	Flying through in the 3D universe map (Courtesy of http://msowwww.anu.edu.au/)	69
Fig. 2.23	Part of the rectangular logarithmic map of the universe depicting major astronomical objects beyond 100 mpc from the Earth (The full map is available at http://www.astro.princeton.edu/universe/all100.gif . Reprinted from Gott et al. 2005)	71
Fig. 2.24	A map of the universe based on the SDSS survey data and relevant literature data from the web of science. The map depicts 618,223 astronomic objects, mostly identified by the SDSS survey, including 4 space probes (A high resolution version of the map can be found at http://cluster.cis.drexel.edu/~cchen/projects/ sdss/images/2007/poster.jpg)	72
Fig. 2.25	The design of the circular map of the universe	72
Fig. 2.26	The types of objects shown in the circular map of the universe ...	73
Fig. 2.27	The center of the circular map of the universe	73

Fig. 2.28	Major discoveries in the west region of the map. The 2003 Sloan Great Wall is much further away from us than the 1989 CfA2 Great Wall.....	74
Fig. 2.29	The Hubble Ultra Deep Field (<i>HUDF</i>) is featured on the map of the universe	75
Fig. 2.30	SDSS quasars associated with citation bursts.....	76
Fig. 2.31	A network of co-cited publications based on the SDSS survey. The <i>arrow points</i> to an article published in 2003 on a survey of high redshift quasars in SDSS II. A citation burst was detected for the article.....	76
Fig. 2.32	The original structure of DNA's double helix (Reprinted from Watson 1968)	79
Fig. 2.33	Ear acupuncture point map. What is the best organizing metaphor? (Courtesy of http://www.auriculotherapy-intl.com/)	80
Fig. 2.34	Musculoskeletal points (©1996 Terry Oleson, UCLA School of Medicine. http://www.americanwholehealth.com/images/earms.gif)	81
Fig. 2.35	<i>Caenorhabditis elegans</i> gene expression terrain map created by VxInsight, showing three-dimensional representation of 44 gene mountains derived from 553 microarray hybridizations and consisting of 17,661 genes (representing 98.6 % of the genes present on the DNA microarrays) (Reprinted from Kim et al. 2001)	82
Fig. 2.36	114,996 influenza virus protein sequences (Reprinted from Pellegrino and Chen 2011)	83
Fig. 3.1	Liberation by Escher. Rigid triangles are transforming into more lively figures (© Worldofescher.com).....	86
Fig. 3.2	The scope of the Knowledge of London, within which London taxi drivers are supposed to know the most direct route by heart, that is, without resorting to the A–Z street map	90
Fig. 3.3	Nodes <i>a</i> and <i>c</i> are connected by two paths. If $r = \infty$, <i>Path 2</i> is longer than <i>Path 1</i> , violating the triangle inequality; so it needs to be removed.....	95
Fig. 3.4	A Pathfinder network of the 20-city proximity data	96
Fig. 3.5	A Pathfinder network of a group of related concepts	96
Fig. 3.6	Visualization of 279 images by color histogram.....	98
Fig. 3.7	Visualization of 279 images by layout	99
Fig. 3.8	Visualizations of 279 images by texture	100
Fig. 3.9	Valleys and peaks in ThemeView (© PNNL)	102
Fig. 3.10	A virtual landscape in VxInsight	102

Fig. 3.11	A virtual landscape of patent class 360 for a period between 1980 and 1984 in VxInsight. Companies' names are color-coded: Seagate- <i>red</i> , Hitachi- <i>green</i> , Olympus- <i>blue</i> , Sony- <i>yellow</i> , IBM- <i>cyan</i> , and Philips- <i>magenta</i> (Courtesy of Kevin Boyack)	103
Fig. 3.12	A SOM-derived base map of the literature of geography (Reprinted from Skupin 2009)	104
Fig. 3.13	The process of visualizing citation impact in the context of co-citation networks (© 2001 IEEE)	105
Fig. 3.14	The design of ParadigmView (© 2001 IEEE)	107
Fig. 3.15	Examples of virtual landscape views (© 2001 IEEE)	108
Fig. 3.16	Streams of topics in Fidel Castro's speeches and other documents (Reprinted from Havre et al. 2000)	109
Fig. 3.17	The evolution of topics is visualized in TextFlow (Reprinted from Cui et al. 2011)	110
Fig. 3.18	Alluvial map of scientific change (Reprinted from Rosvall and Bergstrom 2010)	111
Fig. 3.19	Load a network in .net format to the alluvial map generator	112
Fig. 3.20	An alluvia map generated based on networks of co-occurring terms in publications related to regenerative medicine. Top 300 most frequently occurred terms are chosen each year	112
Fig. 3.21	An alluvial map of popular tweet topics identified as Hurricane Sandy approaching	113
Fig. 3.22	An alluvial map of co-occurring patterns of chemical compound fragments	113
Fig. 3.23	The simplest procedure of generating an MDS map	115
Fig. 3.24	A geographic map showing 20 cities in the US (Copyright © 1998–2012 USATourist.com, LLC http://www.usatourist.com/english/tips/distances.html)	115
Fig. 3.25	An MDS configuration according to the mileage chart for 20 cities in the US	116
Fig. 3.26	The mirror image of the original MDS configuration, showing an overall match to the geographic map, although Orlando, Miami should be placed further down to the South	116
Fig. 3.27	The procedure of generating an MST-enhanced MDS map of the CRCARS data. Nodes are placed by MDS and MST determines explicit links	117

Fig. 3.28	An MDS configuration of the 406 cars in the CRCARS data, including an MST overlay. The edge connecting a pair of cars is coded in <i>grayscale</i> to indicate the strength of similarity: the darker, the stronger the similarity. The MST structure provides a reference framework for assessing the accuracy of the MDS configuration (Courtesy of http://www.pavis.org/)	118
Fig. 3.29	The procedure of journal co-citation analysis described in Morris and McCain (1998)	118
Fig. 3.30	Cluster solution for SCI co-citation data (Reproduced from Morris and McCain (1998). Note that “Comput Biol Med” and “Int J Clin Monit Comput” belong to different clusters).....	119
Fig. 3.31	SCI multidimensional scaling display with cluster boundaries (Reproduced from Morris and McCain (1998). Note the distance between “Comput Biol Med” and “Int J Clin Monit Comput” to the left of this MDS configuration).....	120
Fig. 3.32	Individual differences scaling results of two red-green color-deficient subjects. The Y axis is not fully extended as normal subjects	122
Fig. 3.33	SCI weighted individual differences scaling display (Reproduced from Morris and McCain 1998)	122
Fig. 3.34	SSCI weighted individual differences scaling display (Reproduced from Morris and McCain 1998)	123
Fig. 3.35	The Swiss-roll data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. <i>Straight lines</i> in the embedding (the <i>blue</i> line in part a) now represent simpler and cleaner approximations to the true geodesic paths than do the corresponding graph paths (the <i>red</i> line in part b) (Reproduced from Tenenbaum et al. (2000) Fig. 3. http://www.sciencemag.org/cgi/content/full/290/5500/2319/F3)	124
Fig. 3.36	Face images varying in pose and illumination (Fig. 1A) (Reprinted from Tenenbaum et al. 2000).....	125
Fig. 3.37	Isomap ($K = 6$) applied to 2,000 images of a hand in different configurations (Reproduced from Supplemental Figure 1 of Tenenbaum et al. (2000) http://isomap.stanford.edu/handfig.html)	126
Fig. 3.38	The color-coding illustrates the neighborhood-preserving mapping discovered by LLE (Reprinted from Roweis and Saul 2000).....	126
Fig. 3.39	The procedure used for concept mapping	128

Fig. 3.40	An MDS-configured base map of topical statements and ratings of importance shown as <i>stacked bars</i>	129
Fig. 3.41	Hierarchical cluster analysis divided MDS coordinates into nine clusters	130
Fig. 3.42	A structural hole between groups <i>a</i> , <i>b</i> and <i>c</i> (Reprinted from Burt 2002)	132
Fig. 3.43	A visualization of a co-citation network associated with research in regenerative medicine. The <i>colors</i> indicate the time of publication.....	138
Fig. 4.1	Three Traveling Salesman tours in German cities: the 45-city Alten Commis-Voyageur tour (<i>green</i>), the Groetschel's 120-city tour (<i>blue</i>), and by far the latest 15,112-city tour (<i>red</i>) (Courtesy of http://www.math.princeton.edu/)	145
Fig. 4.2	Knowledge garden	146
Fig. 4.3	A scene in StarWalker when two users exploring the semantically organized virtual space	146
Fig. 4.4	More users gathering in the scene	147
Fig. 4.5	A site map produced by see POWER. The <i>colored contours</i> represent the hit rate of a web page. The home page is the node in the center (Courtesy of http://www.compudigm.com/)	147
Fig. 4.6	Modeling trails of information foragers in thematic spaces.....	149
Fig. 4.7	Legend for the visualization of foraging tails	154
Fig. 4.8	Relevant documents for Task A in the ALCOHOL space (MST)	155
Fig. 4.9	Overview first: user jbr's trails in searching the alcohol space (Task A).....	155
Fig. 4.10	Zoom in	157
Fig. 4.11	Details on demand	157
Fig. 4.12	Overview first, zoom in, filtering, detail on demand. Accumulative trajectory maps of user jbr in four consecutive sessions of tasks. Activated areas in each session reflect the changes of the scope (clockwise: Task A to Task D)	158
Fig. 4.13	Synthesized trails. The trajectory of the optimal path over the original path of user jbr	159
Fig. 5.1	An inclusion map of research in mass extinction based on index terms of articles on mass extinction published in 1990. The size of a node is proportional to the total number of occurrences of the word. Links that violate first-order triangle inequality are removed ($\epsilon = 0.75$)	169
Fig. 5.2	The co-word map of the period of 1980–1985 for the debate on RISC.....	171

Fig. 5.3	The co-word map of another period: 1986–1987 for the debate on RISC.....	171
Fig. 5.4	A document co-citation network of publications in <i>Data and Knowledge Engineering</i>	173
Fig. 5.5	Citation analysis detected a vital missing citation from Mazur's paper in 1962 to Rydon's paper in 1952	174
Fig. 5.6	A global map of science based on document co-citation patterns in 1996, showing a linked structure of nested clusters of documents in various disciplines and research areas (Reproduced from Garfield 1998).....	176
Fig. 5.7	Zooming in to reveal a detailed structure of biomedicine (Reproduced from Garfield 1998).....	177
Fig. 5.8	Zooming in even further to examine the structure of immunology (Reproduced from Garfield 1998)	178
Fig. 5.9	The specialty narrative of leukemia viruses. Specialty narrative links are labeled by citation-context categories (Reproduced from Small 1986).....	179
Fig. 5.10	A generic procedure of co-citation analysis. <i>Dashed lines</i> indicate visualization options	181
Fig. 5.11	The first map of author co-citation analysis, featuring specialties in information science (1972–1979) (Reproduced from White and Griffith 1981).....	182
Fig. 5.12	A two-dimensional Pathfinder network integrated with information on term frequencies as the third dimension (Reproduced from Chen 1998)	183
Fig. 5.13	A Pathfinder network of SIGCHI papers based on their content similarity. The interactive interface allows users to view the abstract of a paper seamlessly as they navigate through the network (Reproduced from Chen 1998)	184
Fig. 5.14	A Pathfinder network of co-cited authors of the ACM Hypertext conference series (1989–1998) (Reproduced from Chen and Carr 1999)	184
Fig. 5.15	A Pathfinder network of 121 information science authors based on raw co-citation counts (Reproduced from White 2003)	185
Fig. 5.16	A minimum spanning tree solution of the author co-citation network based on the ACM Hypertext dataset (Nodes = 367, Links = 366)	187
Fig. 5.17	The author co-citation network of the ACM Hypertext data in a Pathfinder network (Nodes = 367, Links = 398)	188
Fig. 5.18	The procedure of co-citation analysis as described in Chen and Paul (2001)	188

Fig. 5.19	A Pathfinder network showing an author co-citation structure of 367 authors in hypertext research (1989–1998). The color of a node indicates its specialty membership identified by PCA: <i>red</i> for the most predominant specialty, <i>green</i> the second, and <i>blue</i> the third (© 1999 IEEE)	189
Fig. 5.20	A landscape view of the hypertext author co-citation network (1989–1998). The height of each vertical bar represents periodical citation index for each author (© 1999 IEEE)	190
Fig. 5.21	An annotated historiograph of co-citation research (Courtesy of Eugene Garfield; the original diagram can be found at: http://garfield.library.upenn.edu/histcomp/cocitation_small-griffith/graph/2.html)	191
Fig. 5.22	A minimum spanning tree of a network of 1,726 co-cited patents related to cancer research	193
Fig. 5.23	Landscapes of patent class 360 for four 5-year periods. Olympus's patents are shown in <i>blue</i> ; Sony in <i>green</i> ; Hitachi in <i>green</i> ; Philips in <i>magenta</i> ; IBM in <i>cyan</i> ; and Seagate in <i>red</i> (Reproduced from Figure 1 of Boyack et al. 2000)	194
Fig. 5.24	Map of all patents issued by the US Patent Office in January 2000. Design patents are shown in <i>magenta</i> ; patents granted to universities in <i>green</i> ; and IBM's patents in <i>red</i> (Reproduced from Figure 5 of Boyack et al. 2000)	195
Fig. 5.25	A visualization of the literature of co-citation analysis	196
Fig. 6.1	Paradigm shift in collagen research (Reproduced from Small 1977)	204
Fig. 6.2	The curve of a predominant paradigm	206
Fig. 6.3	An artist's illustration of the impact theory: before the impact, seconds to impact, moment of impact, the impact crater, and the impact winter (© Walter Alvarez)	209
Fig. 6.4	Shoemaker-Levy 9 colliding into Jupiter in 1994. Eight impact sites are visible. From <i>left to right</i> are the E/F complex (barely visible on the edge of the planet), the star-shaped H site, the impact sites for tiny N, Q1, small Q2, and R, and on the far right limb the D/G complex. The D/G complex also shows extended haze at the edge of the planet. The features are rapidly evolving on timescales of days. The smallest features in this image are less than 200 km across. This image is a color composite from three filters at 9,530, 5,550, and 4,100 Å (Copyright free, image released into the public domain by NASA)	210

Fig. 6.5	Interpretations of the key evidence by competing paradigms in the KT debate	210
Fig. 6.6	A paradigmatic view of the mass extinction debates (1981–2001)	211
Fig. 6.7	The location of the Chicxulub crater	212
Fig. 6.8	Chicxulub's gravity field (<i>left</i>) and its magnetic anomaly field (<i>right</i>) (© Mark Pilkington of the Geological Survey of Canada)	213
Fig. 6.9	The periodicity cluster	215
Fig. 6.10	A year-by-year animation shows the growing impact of articles in the context of relevant paradigms. The <i>top-row</i> snapshots show the citations gained by the KT impact articles (<i>center</i>), whereas the <i>bottom-row</i> snapshots highlight the periodicity cluster (<i>left</i>) and the Permian extinction cluster (<i>right</i>)	216
Fig. 6.11	Citation peaks of three clusters of articles indicate potential paradigms	217
Fig. 6.12	Supermassive black holes search between 1991 and 1995. The visualization of the document co-citation network is based on co-citation data from 1981 through 2000. Three paradigmatic clusters highlight new evidence (the cluster near to the front) as well as theoretical origins of the AGN paradigm	221
Fig. 6.13	The visualization of the final period of the AGN case study (1996–2000). The cluster near to the front has almost vanished and the cluster to the right has also reduced considerably. In contrast, citations of articles in the center of the co-citation network rocketed, leading by two evidence articles published in Nature: one is about NGC-4258 and the other is about MCG-6-30-15.....	222
Fig. 6.14	The rises and falls of citation profiles of 221 articles across three periods of the AGN paradigm	223
Fig. 7.1	An evolving landscape of research pertinent to BSE and CJD. The next hot topic may emerge in an area that is currently not populated	228
Fig. 7.2	A Venn diagram showing potential links between bibliographically unconnected literatures (Figure 1 reprinted from Swanson and Smalheiser (1997))	233
Fig. 7.3	A schematic diagram, showing the most promising pathway linking <i>migraine</i> in the source literature to <i>magnesium</i> in the target literatures (C to A3) (Courtesy of http://kiwi.uchicago.edu/)	234

Fig. 7.4	A schematic flowchart of Swanson's Procedure II (Figure 4 reprinted from Swanson and Smalheiser (1997), available at http://kiwi.uchicago.edu/webwork/fig4.xbm)	235
Fig. 7.5	Mainstream domain knowledge is typically high in both relevance and citation, whereas latent domain knowledge can be characterized as high relevance and low citation	237
Fig. 7.6	The 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 article in the process of domain expansion. The expanded domain consists of articles connecting to the seeding article by citation chains of no more than two citation links. Latent domain knowledge is represented through a citation network of these articles	237
Fig. 7.7	An overview of the document co-citation map. Lit-up articles in the scene are Swanson's publications. Four of Swanson's articles are embedded in the largest branch – information science, including information retrieval and citation indexing. A dozen of his articles are gathered in the green specialty – the second largest grouping, ranging from scientometrics, neurology, to artificial intelligence. The third largest branch – headache and magnesium – only contains one of Swanson's articles.....	239
Fig. 7.8	The procedure of visualizing latent domain knowledge.....	241
Fig. 7.9	An overview of the mainstream domain knowledge.....	242
Fig. 7.10	A landscape view of the Pathfinder case. Applications of Pathfinder networks are found in a broader context of knowledge management technologies, such as knowledge acquisition, knowledge discovery, and artificial intelligence. A majority of Pathfinder network users are cognitive psychologists	243
Fig. 7.11	This citation map shows that the most prolific themes of Pathfinder network applications include measuring the structure of expertise, eliciting knowledge, measuring the organization of memory, and comparing mental models. No threshold is imposed.....	245
Fig. 7.12	This branch represents a new paradigm of incorporating Pathfinder networks into Generalized Similarity Analysis (GSA), a generic framework for structuring and visualization, and its applications especially in strengthening traditional citation analysis	248

Fig. 7.13	Schvaneveldt's "exit" landmark in the landscape of the thematic visualization	250
Fig. 7.14	An overview of 379 articles in the mainstream of BSE and vCJD research	251
Fig. 7.15	A year-by-year animation shows the growing impact of research in the connections between BSE and vCJD. <i>Top-left</i> : 1991–1993; <i>Top-right</i> : 1994–1996; <i>Bottom-left</i> : 1997–1999; <i>Bottom-right</i> : 2000–2001	252
Fig. 7.16	Articles cited more than 50 times during this period are labeled. Articles labeled 1–3 directly address the BSE-CJD connection. Article 4 is Prusiner's original article on prion, which has broad implications on brain diseases in sheep, cattle, and human	253
Fig. 8.1	An overview of the structural variation model	263
Fig. 8.2	Scenarios that may increase or decrease individual terms in the modularity metric	267
Fig. 8.3	The structure of the system before the publication of the ground breaking paper by Watts	271
Fig. 8.4	The structure of the system after the publication of Watts 1998	271
Fig. 8.5	The structural variation method is applied to a set of patents related to cancer research. The <i>star</i> marks the position of a patent (US6537746). The <i>red lines</i> show where the boundary-spanning connections were made by the patent. Interestingly, the impacted clusters are about recombination	274
Fig. 8.6	Major areas of regenerative medicine	277
Fig. 8.7	The modularity of the network dropped considerably in 2007 and even more in 2009, suggesting that some major structural changes took place in these 2 years in particular	282
Fig. 8.8	Many members of Cluster #7 are found to have citation bursts, shown as citation rings in <i>red</i> . Chin MH 2009 and Stadtfeld M 2010 at the <i>bottom</i> area of the cluster represent a theme that differs from other themes of the cluster	286
Fig. 8.9	A network of the regenerative medicine literature shows 2,507 co-cited references cited by top 500 publications per year between 2000 and 2011. The work associated with the two labelled references was awarded the 2012 Nobel Prize in Medicine	288
Fig. 8.10	The rate of retraction is increasing in PubMed (As of 3/29/2012)	291
Fig. 8.11	The survival function of retraction. The probability of surviving retraction for 4 years or more is below 0.2	297

Fig. 8.12	An overview of co-citation contexts of retracted articles. Each <i>dot</i> is a reference of an article. <i>Red dots</i> indicate retracted articles. The numbers in front of labels indicate their citation ranking. Potentially damaging retracted articles are in the middle of an area that otherwise free from <i>red dots</i>	299
Fig. 8.13	<i>Red dots</i> are retracted articles. Labeled ones are highly cited. Clusters are formed by co-citation strengths	300
Fig. 8.14	An extensive citation context of a retracted 2003 article by Nakao et al. The co-citation network contains 27,905 cited articles between 2003 and 2011. The <i>black dot</i> in the middle of the dense network represents the Nakao paper. <i>Red dots</i> represent 340 articles that directly cited the Nakao paper (there are 609 such articles in the Web of Science). <i>Cyan dots</i> represent 2,130 of the 9,656 articles that bibliographically coupled with the direct citers	300
Fig. 8.15	69 clusters formed by 706 sentences that cited the 1998 Wakefield paper	302
Fig. 8.16	Divergent topics in a topic-transition visualization of the 1998 Wakefield et al. article	302
Fig. 8.17	The UCSD map of science. Each node in the map is a cluster of journals. The clustering was based on a combination of bibliographic couplings between journals and between keywords. Thirteen regions are manually labeled (Reproduced with permission)	306
Fig. 8.18	Areas of research leadership for China. <i>Left</i> : A discipline-level circle map. <i>Right</i> : A paper-level circle map embedded in a discipline circle map. Areas of research leadership are located at the average position of corresponding disciplines or paradigms. The intensity of the nodes indicates the number of leadership types found, Relative Publication Share (<i>RPS</i>), Relative Reference Share (<i>RRS</i>), or state-of-the art (<i>SOA</i>) (Reprinted from Klavans and Boyack 2010 with permission).....	307
Fig. 8.19	A discipline-level map of 812 clusters of journals and proceedings. Each node is a cluster. The size of a node represents the number of papers in the cluster (Reprinted from Boyack 2009 with permission).....	308
Fig. 8.20	The Scopus 2010 global map of 116,000 clusters of 1.7 million articles (Courtesy of Richard Klavans and Kevin Boyack, reproduced with permission)	309

Fig. 8.21	An overlay on the Scopus 2010 map shows papers that acknowledge NCI grants (Courtesy of Kevin Boyack, reproduced with permission)	310
Fig. 8.22	A global science overlay base map. Nodes represent Web of Science Categories. <i>Grey links</i> represent degree of cognitive similarity (Reprinted from Rafols et al. 2010 with permission)	311
Fig. 8.23	An interactive science overlay map of Glaxo-SmithKline's publications between 2000 and 2009. The <i>red circles</i> are GSK's publications in clinical medicine (as moving mouse-over the Clinical Medicine label) (Reprinted from Rafols et al. 2010 with permission, available at http://idr.gatech.edu/usermapsdetail.php?id=61)	312
Fig. 8.24	A similarity map of JCR journals shown in VOSViewer.....	313
Fig. 8.25	The Blondel clusters in the citing journal map (<i>left</i>) and the cited journal map (<i>right</i>). The overlapping polygons suggest that the spatial layout and the membership of clusters still contain a considerable amount of uncertainty. Metrics calculated based on the coordinates need to take the uncertainty into account	314
Fig. 8.26	Citation arcs from the publications of Drexel's iSchool (<i>blue arcs</i>) and Syracuse School of Information Studies (<i>magenta arcs</i>) reveal where they differ in terms of both intellectual bases and research frontiers	315
Fig. 8.27	h-index papers (<i>cyan</i>) and citers to CiteSpace (<i>red</i>)	315
Fig. 9.1	A screenshot of GeoTime (Reprinted from Eccles et al. 2008)	322
Fig. 9.2	CiteSpace labels clusters with title terms of articles that cite corresponding clusters	323
Fig. 9.3	Citations over time are shown as tree rings. Tree rings in <i>red</i> depict the years an accelerated citation rate was detected (citation burst). Three areas emerged from the visualization	324
Fig. 9.4	A network of 12,691 co-cited references. Each year top 2,000 most cited references were selected to form the network. The same three-cluster structure is persistent at various levels	325
Fig. 9.5	The document view in Jigsaw	326
Fig. 9.6	The list view of Jigsaw, showing a list of authors, a list of concepts, and a list of index terms. The input documents are papers from the InfoVis and VAST conferences ...	327
Fig. 9.7	A word tree view in Jigsaw	328
Fig. 9.8	Tablet in Jigsaw provides a flexible workspace to organize evidence and information	328

Fig. 9.9 Carrot’s visualizations of clusters of text documents.
Top right: Aduna cluster map visualization; *lower middle:* circles visualization; *lower right:* Foam Tree visualization 329

Fig. 9.10 *Left:* The geographic layout of the Western Power Grid (WECC) with 230 kV or higher voltage. *Right:* a GreenGrid layout with additional weights applied to both nodes (using voltage phase angle) and links (using impedance) (Reprinted from Wong et al. 2009 with permission)..... 330

Fig. 9.11 A screenshot of ASE (Reprinted from Dunne et al. 2012 with permission) 332

Fig. 9.12 An ultimate ability to reduce the vast volume of scientific knowledge in the past and a stream of new knowledge to a clear and precise representation of a conceptual structure 337

Fig. 9.13 A fitness landscape of scientific inquires..... 337