Transactions in GIS

Transactions in GIS, 2010, 14(5): 689–708

Research Article

The Landscape of GIScience Publications 1997–2007: An Empirical Investigation with Latent Semantic Analysis

David A. Parr
Department of Geography, Texas State University

Yongmei Lu
Department of Geography, Texas State University

Abstract

This article reports on an empirical study of the trends and patterns of research activities in Geographic Information Science (GIScience) during the years 1997–2007. The GIScience research priorities identified by the University Consortium of Geographic Information Science (UCGIS) were used as guidelines to examine the 985 research articles published in six well-recognized academic journals. Latent Semantic Analysis (LSA) was employed to investigate the association among the different GIScience research themes. The spatial and temporal patterns of the association between the publications and the different GIScience themes were examined to show the development of GIScience research during the study period. Furthermore, correlation analyses between the publications were conducted following the LSA results to reveal GIScience research networks, including the networks of the published articles and those formed by the research places. In this article, we applied an approach that was developed within information science to depict *what* GIS research activities were conducted *when* and *where* and *how* they connect to each other through sharing common research themes. The related findings pave the way for future efforts to describe the paradigm of GIScience as well as the pattern of GIScience research.

1 Motivation and Research Design

"The rapid rise and diffusion of Geographic Information Science is one of the most remarkable features of the academic landscape in the last decade of the twentieth century" (Berry 2004 p. 444). Geographic Information Science (GIScience) was first proposed as a scientific field in the early 1990s (Goodchild 1990, 1992). It refers to the

Address for correspondence: Yongmei Lu, Department of Geography, Texas State University, San Marcos, 601 University Drive, San Marcos, TX 78666, USA. E-mail: yl10@txstate.edu

© 2010 Blackwell Publishing Ltd
doi: 10.1111/j.1467-9671.2010.01228.x

theories and methodologies addressing issues related to the development, use, and application of geographic information systems and related technologies. Following Kuhn (1996), a clear definition of the paradigm of GIScience should include a distinct set of theories and methodologies that reflect a world view of the members of GIScience community. However, with GIScience's astounding development, an effort to define its paradigm is as risky as an attempt to predict its future development (Goodchild 2009). Yet we can better prepare for the future of GIScience by looking back at its history to understand its growth patterns and trends. This article reports on an empirical effort that examines the research activity landscape represented by the selected GIScience publications during the period 1997–2007.

The University Consortium of Geographic Information Science (UCGIS) was formed by "a broad representation of universities and professional associations"; its mission includes "serv(ing) as an effective, unified voice for the geographic information science research community"; its goals incorporate "provid(ing) ongoing research priorities for advancing theory and methods in Geographic Information Science" (http://www.ucgis.org/aboutucgis/mission_goals.htm). The UCGIS is recognized for its role in advocating GIScience research (Elmes 2005). The Consortium publishes a series of white papers on GIScience research priorities every few years (UCGIS 1996, 1998, 2000). Usery and McMaster (2005) provided a detailed description on the evolving process of the UCGIS research priorities. To what degree have these priorities guided GIScience research activities during the fast-growing years from 1997 to 2007? How are these priorities reflected by the GIScience publications during this period? Answers to these questions will not only enhance our understanding of the structure and paradigm of GIScience, they will also help UCGIS to reflect on its research priorities in order to better serve the GIScience community. Table 1 summarizes the research themes advocated by UCGIS and investigated by this study.

By using the UCGIS research priorities as reference to examine the research themes of the publications, this study may be open to a critique of possible bias towards the nomenclature of GIScience activities in North America. However, we argue that, through supporting international membership and including GIScientists from Europe and other parts of the world in its many activities, UCGIS serves GIScience communities beyond the physical boundary of the US. Moreover, the ubiquitous Internet access has minimized the impact of distance on the distribution of ideas and information (Cairncross 2001). The extremely low cost of information flow facilitates high frequency and increasing volume of intellectual capital flow (Batty 2007). Therefore, GIScientists from other parts of the world have, theoretically, similar access to the UCGIS white papers. Furthermore, a large number of GIScientists in the US came from other places; many of them maintain active collaboration with people and institutions in their countries of origin and act as a channel to connect GIScience activities in the US with those outside. Therefore, in spite of being instituted in the US, the UCGIS is well-positioned to promote GIScience beyond the US. Nevertheless, further empirical studies will certainly benefit from a more comprehensive investigation to include GIScience research themes that are promoted by more organizations worldwide.

Starting from the UCGIS research themes, the first part of this study depicts the knowledge domain of GIScience that is reflected by the themes and their relationships. To reveal how these themes are related to the GIScience research activities, this study examines how the different research themes have been pursued across space and time. Latent semantic analysis (LSA) is used to analyze the text of the publications in order to

© 2010 Blackwell Publishing Ltd

Table 1 GIScience research priorities identified by the UCGIS research white papers

UCGIS research priority white paper	Year published	Keywords
Spatial data acquisition and integration (SD)	1998	Acquisition, integration
Distributed computing (DC)	1998	Distributed
Extensions to geographic representations (GR)	1998	Representation
Cognition of geographic information (C)	1998	Cognition
Interoperability of geographic information (IO)	1998	Interoperability
Scale (S)	1998	Scale
Spatial analysis in a GIS environment (SA)	1998	Analysis
The future of spatial information infrastructure (SI)	1998	Infrastructure
Uncertainty in geographic information and GIS-based analyses (U)	1998	Uncertainty
GIS and society (GS)	1998	Society
Geospatial data mining and knowledge Discovery (DM)	2000	Mining
Ontological foundations for GIS (O)	2000	Ontological, ontology
Geographic visualization (GV)	2000	Visualization
Remotely acquired data and information in GIScience (RD)	2000	Remote

assess the subscription of each article to the different themes. Summarizing by research themes, we provide a quantitative description of the spatial patterns of GIS publications on the different research themes and the changes in these patterns through time.

The second part of this study focuses on depicting the GIScience research networks as reflected by the publications. Both the connections among researchers and those among places are examined. The text of the publications is examined using LSA to create a similarity index measuring the association between two articles. The articles are then grouped by their authors and authors' affiliations. Similarity indices of GIS publications were then generated between different authors and between different places. Using the similarity indices to indicate the strength of connections, we describe how GIS research themes have linked different researchers and different places.

2 Related Approaches

Rarely is any research conducted without connection to the remainder of the world. The connection exits through various links between researchers and across institutions; it is also present due to the inherent association between different themes or topics. The investigation of such connections can reveal a landscape of research activities to depict what research is conducted where, how research activities are connected across space and time, and how research of different themes converge to define the knowledge domain of a scientific discipline (e.g. Skupin 2004a, b).

© 2010 Blackwell Publishing Ltd

2.1 Social Network Theory

Social networks represent the pattern of human interaction present in a social structure (Newman 2000). A social network is made of vertices of actors and linkages representing acquaintances between vertices. Actors are most often individuals, while the linkages are exchanges of financial, informational, or technological means, markers of the spread of disease, and organizational structures of formal or informal groupings. The structure of a social network is of particular interest because it has implications for the spread of phenomena or information.

Revealing the structure of the social network through which research ideas and fronts spread can help in understanding the construction and development of a scientific discipline. However, it is a challenging task to depict the social network that facilitates the spread of scientific research. Not all social connections are appropriate for the communication of scientific ideas. Moreover, the rapid development of the Internet and other communication technologies has greatly decreased the friction of distance; this makes the establishment and maintenance of research networks more "intangible" and different from the traditional social networks. As a result, co-authorship and citation information is widely used to approximate the scientific connection between scientists.

2.2 Co-Citation and Co-authorship

Citations were first used in the late 19th century (Kessler 1963). The Scientific Citation Index (SCI) was published in the early 1960s as an index of paper-to-paper citations. Citations reflect binary relations between documents and are limited for assessing the connection between articles. Bibliographic coupling was originated by Kessler (1963) to measure the amount that different papers cite the same sources. Co-citation was proposed by Marshakova (1973) and Small (1973) independently to measure the relatedness of two articles by the frequency of them being cited together by a third article. These two indices can describe the strength of links between articles on a continuous scale. The first co-citation map was created by White and Griffith (1981). Co-citation analysis was used to determine the subject similarity and the association between the key ideas of publications (Small 1973). It was also used to analyze the specialty structure of science (e.g. Chen 1999, 2006). However, both bibliographic coupling and co-citation analysis depict the relatedness between documents indirectly – based on the frequency of two documents citing a same third document or being cited by a same third document; they fail to differentiate between critical and marginal citations. The content-based connection between articles is not accounted for.

Another method for detecting the linkage between research is co-authorship. A social structure of science can be described through examining the co-authorship linkages in a science authorship network, a social network where vertices are authors and linkages (or edges) are co-authorship status on publications. Co-authorship networks can be interpreted as a structural representation of the collaborative nature of scientific research (White 2003). The nature of the edge can be defined clearly for these studies (Newman 2000, 2001). A challenge to co-authorship study is that, while reflecting the collaboration between researchers, it does not speak to the specific contribution of a publication to a scientific field. There is no assessment on how a researcher's scientific endeavor advances the structure or knowledge domain of a discipline. Furthermore, similar to

© 2010 Blackwell Publishing Ltd

co-citation analysis, co-authorship study only reflects the research connections that resulted in successful publications.

2.3 Latent Semantic Analysis

Latent semantic analysis (LSA) (Deerwester et al. 1990) examines the semantics of the text in a publication to quantify its association with a target text, which can be a description of a research theme, a set of key words, or another publication. One major advantage of LSA over co-citation analysis and co-authorship study is that LSA assesses the association between two entities using content-based information. Through semantic analysis, one publication can be evaluated as being more or less connected to a research theme or a target document than another publication. Not only can the strength of the connection between any two publications be described (compared with binary assessment), but also the association of a publication to different themes in a scientific field can be quantified. The latter enables a depiction of the structure of a scientific field and its development. Seeing the advantages of semantic analysis over the traditional citation or authorship network analysis, researchers have recently started to promote the use of semantic association for assessing the development of scholarly knowledge and expertise in a discipline (e.g. Börner 2006, 2007). The remainder of this section explains the technical process of LSA in general terms; the detailed explanation of LSA for this empirical study is included in the next section of this article.

LSA starts with creating a term-document matrix to report weighted word counts in each document for the selected terms. Singular value decomposition (SVD) is then performed to decompose this term-document matrix into three matrices: an orthonormal matrix representing term vectors, an orthonormal matrix representing document vectors, and a diagonal matrix containing singular values to represent the strength of relationship between documents and terms (see Figure 1). These three matrices decompose the term-document matrix into linearly independent factors (equation 1).

$$M = U * \Sigma * V^T \tag{1}$$

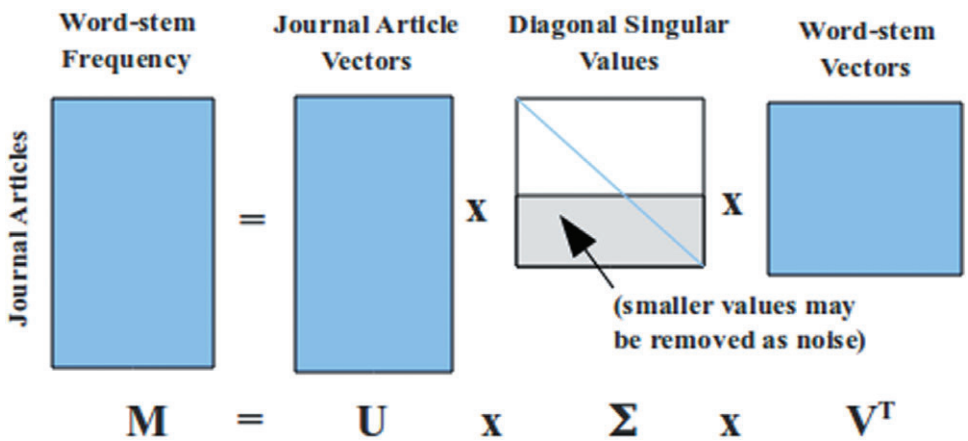


Figure 1 An illustration of singular value decomposition for latent semantic analysis

© 2010 Blackwell Publishing Ltd

where U and V^T have orthogonal columns and Σ is a diagonal matrix of singular values representing the strength of the relationship between U and V^T.

Many of the factors represented in the singular value matrix Σ are small and may be ignored as noise. Keeping only the largest *k* singular values and replacing the small ones with zeros converts the original singular value matrix from Σ to χ. The matrices U and V^T must also be adjusted slightly to maintain orthogonality, and these become T and D^T respectively. By applying dot product to the three matrices, T, χ and D^T, a new *k*-dimensional term-document matrix is created to represent the relative strength of the association between each term and each document (equation 2).

$$X_k = T * \chi * D^T \tag{2}$$

The reduced term-document matrix X_k exhibits characteristics that will be used to extract meaning and relationship from the corpus. Since the term vector space and the document vector space are orthonormal, the dot-product correlation (or cosine) can be computed easily for any two terms or any two documents. Performing dot-product on all terms (or all documents) creates a correlation matrix of term-term (document-document) relationships with each value between –1 and 1. These correlation matrices provide a means to quantify the semantic association between terms (or documents). Furthermore, when aggregating the publications by researchers or their affiliations, the semantic association between research activities by different GIScience researchers and different places can be revealed.

3 Data and Methodology

3.1 Data Preparation

The research articles from *Transactions in GIS*, the *International Journal of Geographical Information Science*, *Cartography and Geographic Information Science*, *GeoInformatica*, and the GIS research papers published in the *Annals of the Association of American Geographers* and *The Professional Geographer* form the corpus of GIS publications investigated in this study. A total of 985 research articles between 1997 and 2007 are included. In addition to the text part of these articles, the date of publication, the names of the authors, and their affiliations were extracted. These articles were published by 823 different first authors, whose affiliations are located in 398 distinguishable places worldwide.

Author affiliations, article titles, and dates of publication were entered into a database using several small programs to check for completeness and accuracy. Common discrepancies include capitalization in names, small differences in the name of author's affiliation, and the choice of complete name versus abbreviation or nickname. One author may have multiple affiliations and multiple authors may be listed for one publication. For this study, the first author and the first affiliation for the first author were extracted to represent the person who conducted the research and the place where the research was conducted. This is based on the assumption that the first author contributes the most to the publication. While this is not always the case, to assess the contribution of each author using information revealed in the publications is beyond the scope of this paper. A potential technique that may be used in the future to address the issue of multiple research locations

© 2010 Blackwell Publishing Ltd

and multiple authors is higher dimensional LSA. For this study, the first author's affiliation was geocoded to represent the location of a published research.

3.2 Latent Semantic Analysis

A LSA module written by Fidolin Wild[1] in the R programming language was used to generate the text-document matrix. This module examines each article to identify words of interest as well as their variations. A total of 602 common words, such as articles, pronouns, common verbs, and prepositions, were ignored. Using an **R** module written by D. T. Lange,[2] words were broken down into word-stems, or roots of words. Using stemming allows LSA to find plurals and alternate words for each word of interest. Lange's R module uses a list of over 60 word endings, logic and rules to decide how and when to stem; it also has a list of exceptions to ensure no excessive stemming is performed. For example, the words "integration", "integrate", "integrates" and "integrator" are stemmed to "integra-", while such words as "succeed" and "ugly" are not reduced to "succe-" and "ug-".

The text-document matrix was further processed to account for the variation of article length. Each word-stem in the matrix is weighted by a "term frequency multiplied by the inverse document frequency method", or TF*IDF method (Salton 1989). "Term frequency" is the number of times a term appears in a given document divided by the total of times the term appears in all documents. "Inverse document frequency" is the log of the number of documents divided by the number of documents the word appears in. The text-matrix generated after this standardization process reflect how intensively each word of interest is used by the 985 articles.

3.3 Variations of LSA Text-Matrices

Variation of the above text-matrix can be generated to reveal more aspects of GIS research activities. First, the "words of interest" for LSA can adopt the keywords of the GIScience research themes. The association of the publications with these keywords reflects how each theme was pursued by the GIS community. Table 2 summarizes the many text-matrix variations. Using research keywords, we can interrogate the thematic substructure of the GIScience publications. Alternatively, by grouping the articles according to authorship, author's affiliation, research places, or publication time, the text-matrices that are generated can be used as input to investigate the association among the publications by different authors, from different places, and at different time. This allows the depiction of the connections among GIScience publications during the investigated years.

Sixteen keywords are identified from the UCGIS research priorities (see Table 1). Each keyword is stemmed to a root-word. With LSA correlation analysis, the words (referred to as "stemmed words") that are closely related to each root-word of the research keyword are identified. For example, Table 3 lists stemmed words that are highly correlated to root-word "uncertain-" (connoting the research keyword "mobile"). The stemmed words with Pearson correlation of 0.5 or above with the root-words are used to generate weighted keywords text-matrices. In this example, the usage of keyword "uncertainty" in an article is the sum of the number of occurrences of "uncertainti-" plus the number of occurrences of "uncertain" weighted by 1.0 plus the number of occurrences of "signifi-" weighted by 0.66, and so on. This way the appearance of each UCGIS keyword in an article is quantified as the accumulated weighted count of all the selected

© 2010 Blackwell Publishing Ltd

Table 2 Summary of the input matrices for Latent Semantic Analysis (the numbers in parentheses indicate the count of entities of the type)

Unit of Investigation	Words (50,371)	Stemmed Words (34,290)	Keywords (16)	Weighted Keywords (16)
Articles (985)	A1	A2	A3	A4
Affiliations (398)	B1	B2	B3	B4
Authors (823)	C1	C2	C3	C4
Years (11)	D1	D2	D3	D4
Places (142)	E1	E2	E3	E4

Table 3 Stemmed words (with a 0.5 or greater Pearson correlation) to the root-word, "uncertain"

Word-stem	Correlation to "uncertain"
"Uncertainti-"	1
"Uncertain"	1
"Signifi-"	0.66
"Discord-"	0.61
"Extrens-"	0.59
"Nonspecif-"	0.55
"Indefin-"	0.55
"Depict-"	0.54
"Creator-"	0.53
"Vectoris-"	0.51

stemmed words. This weighted count for each keyword is further standardized using the TF*IDF method.

After generating the singular value matrices, we create correlation matrices using the pair-wise similarity index between the rows. The results denote the intensity of association between paired articles or paired groups of articles. For instance, a similarity index between two articles represents the thematic association between the articles measured by the common occurrence of certain words that are examined. For article-article, author-author, year-year, affiliation-affiliation, and location-location correlation matrices, we used stemmed words to determine the correlation index. We chose stemmed

© 2010 Blackwell Publishing Ltd

words over whole words because we believe stemmed words better represent the semantic meaning of a term, and that stemmed words produce more connections between subjects.

4 UCGIS Research Priorities as Reflected by the GIScience Publications

4.1 Relationship between the UCGIS Research Themes

To show how the UCGIS research priorities are related among themselves, the correlation between each keyword to every other keyword is investigated. Language is not mathematically precise. As a semantic analysis, LSA tends to reveal a certain degree of association between virtually all entities. Therefore, it is necessary to select certain links from the LSA results in order to reveal the valid connections. Our trial-and-error experiments with the data showed that a 0.8 cut-off value identifies the correlation links that are abundant and meaningful. Only the most salient links – the association with a correlation coefficient of 0.8 or higher – are included in this study to describe the GIS research networks.

The overall findings echo those by Usery and McMaster (2005) – these research themes are intertwined and overlapping frequently, and the relationships between and among themes are rarely reciprocal. However, Usery and McMaster (2005) conducted their evaluation of the interrelations among the UCGIS research themes based on experts' assessment of the concepts that are related to the research themes. This study follows an empirical approach by using LSA to reveal how the published articles that are associated with each of the research themes are related to each other. Put another way, Usery and McMaster (2005) followed a top-down approach to investigate how the research themes are supposed to connect to each other according to expert knowledge about the themes; the present study takes a bottom-up approach to depict how the different GIS themes were revealed to connect to each other by the investigated publications.

Table 4 reports our findings on the connections among the 14 UCGIS priorities. The numbers indicate how two research themes are correlated to each other based on the LSA analyses of the GIS publications. A high correlation coefficient between two research themes indicates that the publications that are associated with one research theme (as assessed by the LSA analysis) tend to be associated with the other theme as well. The correlation coefficients of > 0.5 are highlighted in the table. It can be seen that some areas of research are tightly coupled with others. The *Spatial Analysis* theme is closely related to ten other UCGIS research themes. The *Remotely Acquired Data and Information*, *Geographic Representation*, and *Geospatial Data Mining and Knowledge Discovery* themes are each closely related to nine other UCGIS research priorities. At the other end of the spectrum, research themes such as *Spatial Information Infrastructure* and *GIS & Society* are only closely related to one other research theme (i.e. *Geospatial Data Mining and Knowledge Discovery* and *Spatial Analysis*, respectively). The research themes, *Spatial Data Acquisition and Integration* and *Ontological Foundations of GIS* are each closely related to two other research themes. Moreover, it is interesting to note that some research themes are negatively related to each other, although all of the negative coefficients are relatively small. For example, *the Future of Spatial Information Infrastructure* theme appears to be negatively related to the *Geographic Visualization*, *Geographic Representation*, and *GIS & Society* themes, indicating that the publications closely related to *Spatial Information Infrastructure* tend to be less related to the papers

© 2010 Blackwell Publishing Ltd

Table 4 Connection between the different research themes based on LSA analysis. (Please refer to Table 1 for the abbreviation definitions for the themes.)

	SD	SA	C	DC	SI	IO	O	RD	GR	S	GS	U	GV	DM
	(2)	(10)	(8)	(5)	(1)	(7)	(2)	(9)	(9)	(7)	(1)	(7)	(5)	(9)
SD	1	0.31	**0.67**	0.12	0.27	0.34	0.16	0.52	0.2	0.18	−0.03	0.27	0.29	**0.57**
SA		1	**0.58**	**0.89**	0.35	**0.77**	0.04	**0.94**	**0.76**	**0.78**	**0.52**	**0.72**	**0.68**	**0.78**
C			1	0.22	0.07	**0.8**	**0.56**	**0.65**	**0.63**	0.29	0.32	0.47	**0.64**	**0.86**
DC				1	0.39	0.49	−0.1	**0.84**	**0.7**	**0.9**	0.32	**0.78**	0.37	0.07
SI					1	0.32	0.04	0.2	−0.11	0.08	0.64	−0.01	−0.25	**0.82**
IO						1	**0.67**	**0.72**	**0.83**	**0.6**	0.46	**0.72**	0.42	0.1
O							1	0.06	0.45	0.08	0.05	0.34	−0.14	0.4
RD								1	**0.82**	**0.85**	0.24	**0.82**	**0.69**	**0.89**
GR									1	**0.86**	0.05	**0.94**	**0.54**	**0.9**
S										1	−0.05	**0.96**	0.35	**0.72**
GS											1	−0.11	0.32	0.1
U												1	0.35	**0.83**
GV													1	**0.7**
DM														1

Note: The correlation coefficients larger than 0.5 are highlighted in bold. The numbers in parentheses indicate the frequency that each theme is highly correlated to other themes (by larger than 0.5)

© 2010 Blackwell Publishing Ltd

on the themes of *Geographic Visualization*, *Geographic Representation*, and *GIS & Society*.

Two issues should be kept in mind when interpreting these results. First, the relationship between the different UCGIS research themes only represents the patterns embedded in the publications that are investigated by this study. Therefore, any specially identified focuses or flavors of these six journals during the years 1997–2007 would have impacted the revealed patterns. Second, LSA relies highly on the meaning and context of the words in publications. The lack of a relatively uniform meaning for any words could degrade the algorithm's ability to reveal accurate correlations.

4.2 Research Trends for the UCGIS Themes

Using the matrices listed in the last row of Table 2, LSA was conducted to examine the spatial distribution of GIS publications. By linking all the investigated articles with the cities where the primary authors' affiliations are, the connection between places (or cities) and research themes can be established. Although we cannot for certain infer the innovative sites for the different research themes, a high correlation between a place and a theme signify the presence of active research on the theme in that area. Therefore, the key centers of GIScience research on specific themes can be identified. The ten placess with the most counts of weighted keywords were identified for the UCGIS research priorities (Table 5). Some well-known places for GIS research appear under multiple research themes, including Santa Barbara, CA (11 themes), London (11 themes), Hong Kong (10 themes), State College, PA (9 themes), and Melbourne, Australia (9 themes). This confirms the significant role these places hold in advancing GIScience. On the other end of the spectrum, some locations only appear under one research theme, such as Zurich (under "uncertainty") and Stuttgart (under "visualization"), signifying the contribution by these places to the specific GIScience research themes. Geographically, outside of North America and Europe, GIScience research is active in Austria, Brazil, China (Beijing and Hong Kong), New Zealand, South Africa, South Korea, and Taiwan.

Through linking the articles in each year to the research themes, the thematic distribution of the GIS publications investigated is revealed (Table 6). Several trends can be observed. When checking the GIScience themes on which the publications were focused, some themes were well reflected throughout this study period, including *spatial analysis*, *geographic representation*, *cognition of geographic information,* and *GIS and society*. Other themes have not received enough attention or are not well-represented by the publications, such as *data mining*, *scale*, and *interoperability*, Still others are fairly represented by the publications, including *spatial data acquisition and integration*, *visualization*, and *distributed computing*. Furthermore, examining the sustainability of the GIScience themes, it can be seen that some have become more active while others less active. Themes such as *ontology*, *visualization*, *uncertainty*, and *interoperability* have seen increased publications through the years, while the numbers of publications on *spatial analysis*, *GIS and society*, and *spatial data infrastructure* have dropped considerably. The findings on the dynamics of the GIScience themes throughout the study period can be used as guidance by the organizations such as UCGIS and the disciplines sharing an interest in GIScience to identify strategies and directions to encourage, monitor, and facilitate GIS research activities. They also provide information for researchers to relate their research interests and career plans to the development of GIScience. However, it is important to note that due to the dynamic and imprecise

© 2010 Blackwell Publishing Ltd

Table 5 The ten places with the most GIS publications for different research themes

Cognition	Data Mining	Distributed Computing	Spatial Data Infrastructure
Burnaby, BC	Beijing, China	Canberra, Australia	Columbus, OH
Columbia, SC	Burnaby, BC	Edinburgh, UK	Enschede, Netherlands
Columbus, OH	Hong Kong, China	Hong Kong, China	Hong Kong, China
Enschede, Netherlands	Ispra, Italy	Leuven, Belgium	Leeds, UK
Hong Kong, China	London, UK	London, UK	London, UK
Leeds, UK	Madison, WI	Madison, WI	Melbourne, Australia
Madison, WI	Orono, ME	Melbourne, Australia	San Diego, CA
Melbourne, Australia	Santa Barbara, CA	Orono, ME	Santa Barbara, CA
Munster, Germany	State College, PA	San Diego, CA	Seattle, WA
State College, PA	Vienna, Austria	Santa Barbara, CA	State College, PA

Remote Acquired Data	Representation	Scale	GIS and Society
Canberra, Australia	Hong Kong, China	Canberra, Australia	Camden, NJ
Edinburgh, UK	Ispra, Italy	Edinburgh, UK	Columbus, OH
Hong Kong, China	Leeds, UK	Enschede, Netherlands	Enschede, Netherlands
London, UK	London, UK	Hong Kong, China	London, UK
Madison, WI	Madison, WI	London, UK	Melbourne, Australia
Melbourne, Australia	Orono, ME	Madison, WI	Minneapolis, MN
Orono, ME	Richmond, BC	Melbourne, Australia	San Diego, CA
Santa Barbara, CA	Santa Barbara, CA	San Diego, CA	Santa Barbara, CA
State College, PA	State College, PA	Santa Barbara, CA	Seattle, WA
Vienna, Austria	Vienna, Austria	State College, PA	State College, PA

Interoperability	Data Acquisition and Integration	Uncertainty	Visualization
Burnaby, BC	Columbia, SC	Canberra, Australia	Columbia, SC
Enschede, Netherlands	Columbus, OH	Edinburgh, UK	Columbus, OH
Hong Kong, China	Enschede, Netherlands	Hong Kong, China	Enschede, Netherlands
Leeds, UK	Hong Kong, China	Leuven, Belgium	Hong Kong, China
London, UK	Leeds, UK	London, UK	London, UK
Melbourne, Australia	London, UK	Orono, ME	Madison, WI
Munster, Germany	Melbourne, Australia	Richmond, BC	Melbourne, Australia
Orono, ME	Santa Barbara, CA	Santa Barbara, CA	Santa Barbara, CA
Santa Barbara, CA	University Park, PA	Vienna, Austria	State College, PA
State College, PA	Utrecht, Netherlands	Zurich, Switzerland	Stuttgart, Germany

Spatial Analysis	Ontology
Adelaide, Australia	Cagliara, Italy
Anyang, South Korea	Frederichshafen, Germany
Beijing, China	Gent, Belgium
Buffalo, NY	Kalgoorlie, Australia
Champaign, IL	Research Triangle Park, NC
Durham, NC	Redlands, CA
Kirksville, MO	Thessaloniki, Greece
Peterborough, ONT	Toronto, ONT
Winnipeg, Manitoba	Waterloo, ONT
Yangsan, South Korea	West Long Branch, NJ

© 2010 Blackwell Publishing Ltd

Table 6 The dynamics of the UCGIS research priorities reflected by the six journals

GIS Themes	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Spatial Analysis	**18.3**	**18.5**	**14.7**	**18.4**	**13.5**	**15.9**	**13.4**	**13.3**	**11.7**	**13.5**	**12.2**
GIS and Society	**14.4**	**15.8**	**12.0**	**14.5**	**10.1**	**11.7**	**9.7**	**9.4**	**7.7**	**10.1**	**8.2**
Geog. Representation	12.3	12.4	12.6	12.4	12.5	12.5	12.5	12.5	12.6	12.6	12.6
Cognition	10.9	11.5	10.3	11.1	10.0	10.2	9.8	9.7	9.2	10.0	9.5
Acquisition & Integration	8.8	8.7	9.4	8.8	9.7	9.3	9.7	9.7	10.0	9.7	10.0
Visualization	**8.4**	**7.7**	**10.3**	**8.3**	**11.3**	**10.1**	**11.5**	**11.7**	**12.9**	**11.3**	**12.4**
Distributed Comp.	7.7	8.0	7.4	7.9	7.0	7.3	7.0	6.9	6.6	7.0	6.6
Remote	4.7	4.9	4.6	4.7	4.4	4.5	4.4	4.4	4.2	4.5	4.3
Ontology	**3.9**	**2.4**	**7.1**	**3.7**	**8.9**	**6.9**	**9.4**	**9.7**	**11.7**	**8.9**	**10.9**
Spatial Data Infrastructure	**3.9**	**4.5**	**2.9**	**4.0**	**2.4**	**2.9**	**2.3**	**2.2**	**1.5**	**2.4**	**1.8**
Uncertainty	**3.0**	**2.4**	**4.7**	**2.6**	**5.7**	**4.6**	**5.8**	**5.9**	**6.9**	**5.5**	**6.6**
interoperability	**2.0**	**1.6**	**2.5**	**1.8**	**3.0**	**2.5**	**3.0**	**3.1**	**3.4**	**3.0**	**3.4**
Scale	1.0	1.1	0.8	1.1	0.8	0.9	0.8	0.8	0.7	0.8	0.7
Data Mining	0.7	0.7	0.7	0.7	0.8	0.7	0.8	0.8	0.8	0.7	0.8
Total number of articles	53	58	93	104	88	92	96	87	125	109	80

Note: The numbers indicate the percentage of GIS publications on each theme in each year

© 2010 Blackwell Publishing Ltd
Transactions in GIS, 2010, 14(5)

nature of language, the meaning of a word or a phrase for GIScience can change over time, which might add noise to the patterns revealed.

5 Research Networks Presented in the GIScience Publications

5.1 The Network of GIScience Publications

Running correlation analyses between the LSA indices for the different articles reveal how strongly the articles are related to each other. Visualizing only the significant connections between the articles, Figure 2 is a visual representation of the network formed by the 985 articles. Each edge in the network indicates a correlation coefficient of 0.8 or larger. The figure is intended not to enable the readers to identify which article corresponds to which node in the network, but to illustrate the ubiquitous connections among the published articles when the 14 UCGIS research themes are applied to them. However, it is important to notice that there are a number of self-contained sub-networks in Figure 2. The articles in each sub-network share similar subscription to the research theme(s); they might indicate a possible route of GIScience knowledge diffusion. Further investigation focusing on the sub-networks is warranted, as the spatial and temporal sequence of these articles might reveal how GIS research has "traveled" over space and time.

For illustration purpose, we take a medium-sized sub-network from Figure 2 to show the construction of a sub-network. This sub-network consists of five related articles (Table 7). Two articles (BC2005 and JO2006) cited a third article (F1998) in the same sub-network. The others did not cite each other, but they do hold common citations in their references. Applying LSA and correlation analysis, the connection among the articles are analyzed and their strength reported in Figure 3. It is relatively clear that the

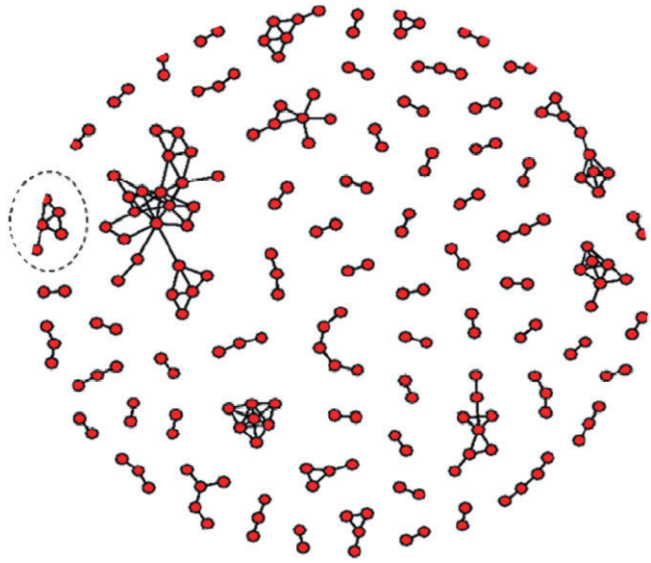


Figure 2 Network representation of the investigated GIS articles. (Note: the oval highlights the sub-network that is further investigated and illustrated in Figure 3)

© 2010 Blackwell Publishing Ltd

Table 7 Description of the five related articles

BC2005	Title: Modeling the Spatial Distribution of DEM Error Author: Bruce H. Carlisle (University of Northumbria) Journal: Transactions in GIS, 2005 Location: Newcastle, UK
JO2006	Title: Uncovering the statistical and spatial characteristics of fine toposcale DEM error Authors: J. Oksanen, T. Sarjakoski (Finnish Geodesic Institute) Journal: IJGIS, 2006 Location: Masala, Finland
PF1998	Title: Improved Modeling of Elevation Error with Geostatistics Author: Peter Fisher (University of Leicester) Journal: GeoInformatica, 1998 Location: Leicester, UK
GA1999	Title: Error Propagation Modeling in Raster GIS: Adding and Ratioing Operations Authors: Giuseppe Arbia, Daniel Griffith, Robert Haining Journal: Cartography and GIS, 1999 Location: Pescara, Italy
WS2003	Title: Modeling error propagation in vector-based buffer analysis Authors: Wenzhong Shi (Hong Kong Polytechnic), Chui Cheung, Changqing Zhu Journal: IJGIS, 2003 Location: Hong Kong, China

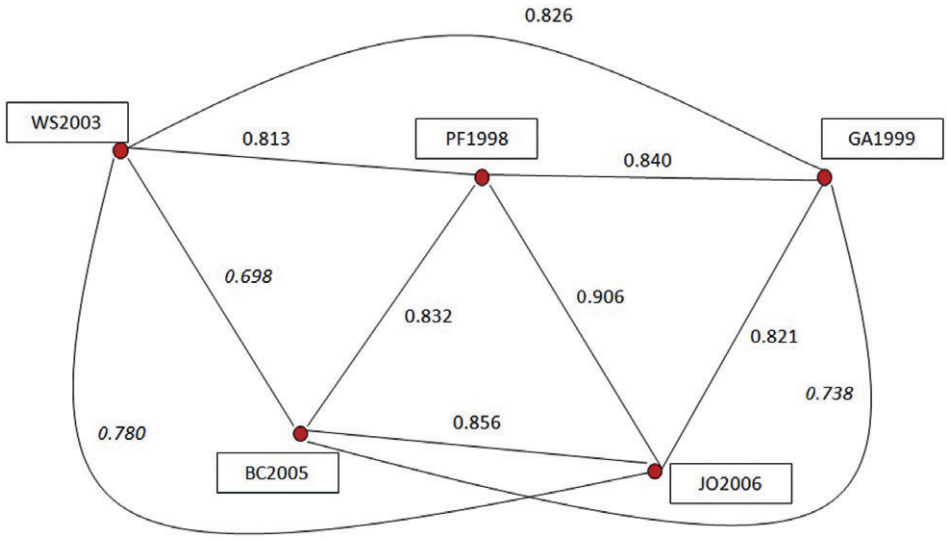


Figure 3 Network representation of the five related articles. (Note: The numbers indicate the correlation value between the paired articles. The links with correlation value lower than 0.8 are denoted in italic fonts)

© 2010 Blackwell Publishing Ltd

PF1998 article has strong connections to the other four articles, indicating that this article might have important impact on the other published research activities. Further investigations can be conducted by following the publication time and research locations to examine if and how these publications can possibly reflect the diffusion of GIScience research related to "error and uncertainty".

5.2 The Network of GIScience Research over Space

As indicated in the last row of Table 2, the information derived from the LSA matrices can be used to derive the correlation between different places. The correlation reveals how the places are connected in their roles to support GIScience research on the different themes. The coefficient for two places is high if the articles published by researchers from these places show similar subscription to the GIScience themes. Figure 4 is a world map illustrating how the different places are connected based on the association between the published articles. Two places are connected on the map if the articles published from them show a high correlation coefficient. The map shows some striking patterns. Many connections exist between places in North America and Europe, both within and between the two continents. The connections from North American or European to other places tend to point to a limited number of geographical regions. These include countries such as Australia, Brazil, China, Japan, New Zealand, Singapore, and South Africa. Some specific places in these countries are more related to the places in North America and Europe than to the other places in their regions. For example, Hong Kong and Beijing are more connected to the West than to the many large cities in China and Asia. However, Australia is a special case – despite its close connection to North America and Europe, two links connect places (Brisbane with Melbourne and Townsville with Newcastle) within this country. These patterns suggest that the major GIScience research

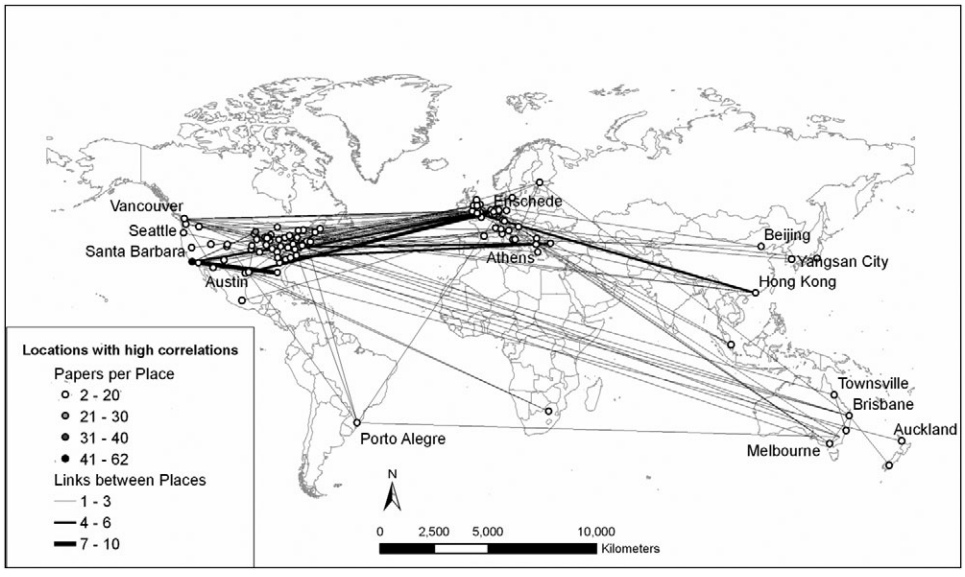


Figure 4 Connections among places measured by the linkage between GIS publications

© 2010 Blackwell Publishing Ltd

activities exist mostly in the developed countries. While some major cities in the developing world are connected to this developed-country research network, there is a lack of connection among places in developing countries. This could suggest either weak communication between the researchers within the developing world or the lack of leading researchers and research in these countries and regions.

It is important to point out that, despite the interesting patterns revealed by the analyses, these results could be biased towards publications in English. As English is the only language used by the journals investigated in this study, non-English speakers and their GIScience publications are underrepresented. Furthermore, the language barrier can be significant for the diffusion of knowledge. Given that English is less popular in many developing countries than in North America and Europe, the lack of connection between the GIScience research activities in these two groups of countries may reflect the restricted information flows. Lastly, institutional funding and resources may present challenges for GIScience research in developing countries. North American and European institutions may be better funded or have better access to GIScience literature.

6 Concluding Remarks

6.1 Conclusions

LSA provides a method to identify and quantify the relationship among published research works. While it may not speak to the quality of the research, it can establish semantic linkages between research themes, publications, publication years, and publication places. Previous academic research has focused on the citation network as the structure of scientific connectivity (Newman 2000, 2001, Small 1973, White and Griffith 1981). In this research, we show that LSA can be used to derive content-based similarity indices between the GIScience publications or other grouping units for publications. A content-based similarity index provides more information than one based on citation information since it describes the semantic intensity of the connection between two entities. LSA is used in this study to examine the relationship between researchers and publications based on their subscription to the GIScience themes. The results reveals the connections between GIScience researchers who share the same research interests through their publications, regardless of their citation, co-citation, or co-authorship relation.

By linking GIS publications to UCGIS research priorities, we depict an empirical picture of the landscape of published GIScience research. The structure of knowledge indicated by UCGIS research priorities is described. The analyses also reveal the dynamics of GIScience research foci during the 11 years. These findings can serve as a reference for UCGIS to revisit their research priorities in order to encourage or de-emphasize specific research themes. Furthermore, the spatial and temporal patterns of GIScience publications on the different themes as well as the network formed by the major GIScience research places, as revealed by this study, have implications for understanding the specialty structure of GIScience and its dynamics. They laid out a good start for future research to follow on in order to understand better how research has changed, and when and where different research themes were pursued.

© 2010 Blackwell Publishing Ltd

6.2 Limitations of Current Study

The publications examined were restricted to the years 1997 to 2007. This was mainly because digital copies of full journal papers published before 1997 are not as easily available for certain journals. Furthermore, the findings reported in this article are likely to be affected by the nature and tastes of the six selected journals. GIScience articles are published in a wide range of journals, including those outside the field of geography. By limiting our source material, we have omitted research in applied GIScience and other fields. Furthermore, the patterns of GIS publications as identified by this empirical study are limited to the themes covered by the UCGIS research priorities. We recognize that there are GIScience concepts and themes that are important for the development of the GIScience research agenda but that were not included in this investigation.

Similar to most other empirical studies of structure and development of a scientific field, the analyses and conclusions are limited to the published GIScience research. However, this does not mean that the unpublished scientific activities are not important or that they are not an integral part of GIScience research.

Lastly, the imprecise nature of language affects the patterns revealed. LSA infers semantic context, but language is never exact or static. Words have multiple meanings and multiple words can share a single meaning; the meaning of a term may change over time. The looseness of language does add noise to the data and affect the quality of the results. Using stemmed words and weighted inputs, we have endeavored to limit the introduction of imprecision and error into the analysis.

6.3 Future Research

At the foundation of all the analyses in this article is the identification of the words closely related to the UCGIS research priorities. As discussed in the opening section of this article, while we acknowledge the limitation of choosing the research priorities of UCGIS, we wanted to point out that UCGIS being a US-centric organization does not necessarily mean that its view for GIScience development excludes non-US research communities. Furthermore, similar limitation would be present should the GIScience research agenda advocated by another institution be chosen. Therefore, future research should explore how to define the most appropriate set of research themes shared by the whole GIScience community. The technical approach presented in this article can be followed by future investigations when a more rigorous GIScience research agenda is applied.

Future research can expand on the network analysis of LSA results. Options include optimizing the correlation cut-off values, finding community structure within the network (Chen 2006, Newman 2006), or measuring the differences among multiple models of network types. Further research may focus on investigating the links between certain places, or the geographic network of a GIScience theme. Knowing the relationship between the research locations and the themes could assist in describing the diffusion patterns of GIS research (Chen et al. 2007, Singh 2005).

LSA is based on SVD, which uses two dimensions of analysis. Higher-order Singular Value Decomposition (HOSVD) (Omberg et al. 2007) analyzes matrices in three or higher dimensions. A higher-order LSA could explore multiple dimensions of analysis, incorporating authorship, research location, publication time, or other measures into the analysis. Doing so would provide a method for removing noise from multiple

© 2010 Blackwell Publishing Ltd

2-dimension analyses. Using thematic keywords in a higher-order LSA to examine the publications in GIScience will remove some publication frequency noise.

Endnotes

1 http://cran.r-project.org/src/contrib/Descriptions/lsa.html; accessed Tuesday, January 29, 2008.
2 http://www.omegahat.org/Rstem/; accessed Tuesday, January 29, 2008.

References

Batty M 2007 The real-time academy: Anyplace, anywhere, anytime. *Environment and Planning B: Planning and Design* 34(6): 947–48

Berry B 2004 Epilogue: Spatial analysis in retrospect and prospect. In Goodchild M F and Janelle D G (eds) *Spatially Integrated Social Science*. New York, Oxford University Press: 443–46

Börner K 2006 Semantics association networks: using semantics web technology to improve scholarly knowledge and expertise management. In Geroimenka V and Chen C (eds) *Visualizing the Semantics Web*. Berlin, Springer: 183–98

Börner K 2007 Making sense of mankind's scholarly knowledge and expertise: Collecting, interlinking, and organizing what we know and different approaches to mapping (network) science. *Environment and Planning B* 34: 808–25

Cairncross F (ed) 2001 *The Death of Distance, How the Communications Revolution Is Changing Our Lives*. Boston, MA, Harvard Business School Press

Chen C 1999 Visualizing semantic spaces and author co-citation networks in digital libraries. *Information Processing and Management* 35: 401–20

Chen C 2006 *Information Visualization: Beyond the Horizon*, 2nd edn. London, Springer

Chen C, Zhu W, Tomaszewski B, and MacEachren A 2007 Tracing conceptual and geospatial diffusion of knowledge. In Schuler D (ed) *Online Communities and Social Computing*. Berlin, Springer-Verlag: 265–74

Deerwester S, Dumais S T, Furnas G W, Landauer T K, and Harshman R 1990 Indexing by latent semantic analysis. *Journal of the American Society of Information Science* 41: 391–407

Elmes G 2005 The University Consortium for Geographic Information Science: Shaping the future at ten years. *Transactions in GIS* 9(3): 273–76

Goodchild M F 1990 Spatial information science. In *Proceedings of the Fourth International Symposium on Spatial Data Handling*, Zurich, Switzerland: 3–12

Goodchild M F 1992 Geographical information science. *International Journal of Geographical Information Systems* 6: 31–45

Goodchild M F 2009 Geographic information systems and science: Today and tomorrow. *Annals of GIS* 15: 3–9

Kessler M M 1963 Bibliographic coupling between scientific papers. *American Documentation* 14: 10–25

Kuhn T S 1996 *The Structure of Scientific Revolutions*, 3rd edn. Chicago, IL, University of Chicago Press: 174–210

Marshakova I V 1973 A system of document connection based on references. *Scientific and Technical Information Serial of VINITI* 6: 3–8

Newman M E J 2000 Who is the best connected scientist? A study of scientific coauthorship networks. WWW document, http://arxiv.org/PS_cache/cond-mat/pdf/0011/0011144v2.pdf

Newman M E J 2001 The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America* 98: 404–09

Newman M E J 2006 Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* 103: 8577–82

Omberg L, Golub G H, and Alter O 2007 A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. *Proceedings of National Academy of Sciences of the United States of America* 104: 18371–76

© 2010 Blackwell Publishing Ltd

Salton G 1989 *Automated Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Reading, MA, Addison-Wesley Publishing

Singh J 2005 Collaborative networks as determinants of knowledge diffusion patterns. *Management Science* 51: 756–70

Skupin A 2004a The world of geography: Visualizing a knowledge domain with cartographic means. *Proceedings of National Academy of Sciences of the United States of America* 101(suppl. 1): 5274–78

Skupin A 2004b A picture from a thousand words. *Computing in Science and Engineering* 6: 84–8

Small H 1973 Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science* 24: 265–69

University Consortium of Geographic Information Science 1996 Research priorities for geographic information science. *Cartography and Geographic Information Systems* 23: 115–27

University Consortium of Geographic Information Science 1998 Research Priorities for Geographic Information Science. WWW document, http://www.ucgis.org/

University Consortium of Geographic Information Science 2000 Research Priorities for Geographic Information Science. WWW document, http://www.ucgis.org/

Usery E L and McMaster R 2005 Introduction to the UCGIS Research Agenda. In Robert McMaster R and Usery E L (eds) *A Research Agenda for Geographic Information Science*. Boca Raton, FL, CRC Press: 1–15

White H D 2003 Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology* 54: 423–34

White H D and Griffith B C 1981 A cocitation map of authors in judgment and decision research. *Journal for the American Society of Information Science* 32: 163–72

© 2010 Blackwell Publishing Ltd