Intellectual structure of biomedical informatics reflected in scholarly events

Senator Jeong · Hong-Gee Kim

Abstract The purpose of this paper was to analyze the intellectual structure of biomedical informatics reflected in scholarly events such as conferences, workshops, symposia, and seminars. As analysis variables, 'call for paper topics', 'session titles' and author keywords from biomedical informatics-related scholarly events, and the MeSH descriptors were combined. As analysis cases, the titles and abstracts of 12,536 papers presented at five medical informatics (MI) and six bioinformatics (BI) global scale scholarly event series during the years 1999–2008 were collected. Then, n-gram terms (MI = 6,958; BI = 5,436) from the paper corpus were extracted and the term co-occurrence network was analyzed. One hundred important topics for each medical informatics and bioinformatics were identified through the hub-authority metric, and their usage contexts were compared with the k-nearest neighbor measure. To research trends, newly popular topics by 2-year period units were observed. In the past 10 years the most important topic in MI has been "decision support", while in BI "gene expression". Though the two communities share several methodologies, according to our analysis, they do not use them in the same context. This evidence suggests that MI uses technologies for the improvement of productivity in clinical settings, while BI uses algorithms as its tools for scientific biological discovery. Though MI and BI are arguably separate research fields, their topics are increasingly intertwined, and the gap between the fields blurred, forming a broad informatics—namely biomedical informatics. Using scholarly events as data sources for domain analysis is the closest way to approximate the forefront of biomedical informatics.

Keywords Scholarly event · Conference · Biomedical informatics · Intellectual structure · Co-word analysis · Social network analysis

S. Jeong · H.-G. Kim (✉)
Biomedical Knowledge Engineering Laboratory, Seoul National University, Seoul, Korea
e-mail: hgkim@snu.ac.kr

S. Jeong
e-mail: senator@snu.ac.kr

Introduction

Efforts to define and characterize biomedical informatics were made in several studies (David 2005; Greenes and Shortliffe 1990; Luscombe et al. 2001). The discipline, however, is highly multidisciplinary. Continuous influx of new knowledge and ever-evolving technologies into this field make some results more or less outdated in reflecting the most up-to-date research fronts. As such, defining and scoping biomedical informatics presupposes the understanding of its building blocks. Our objective is to explore the intellectual structure of biomedical informatics in order to better understand the field.

Qualitative and quantitative study

There are a number of qualitative studies on biomedical informatics. Hasman et al. (1996) modeled biomedical informatics. They said the modeling continues to be a major task in research. And education and practice in the domain, and sound formal knowledge as well as a strong engineering background are needed to conduct this type of research (Hasman and Haux 2006, 2007). Maojo and Kulikowski (2003) compared medical informatics (MI) and bioinformatics (BI) and pointed out that they are complementary in that MI features expertise in health care and BI is strong in biological discovery science. These research efforts heavily relied on experts' knowledge, experience, and intuition. Thus they may lack objective and quantitative understanding of the structure of the field.

Mainstream approaches exploring the intellectual structure of a domain are quantitative studies such as co-citation analysis (CCA) and co-word analysis (CWA) which gauge topical closeness, similarity in the level of authors, papers, journals, or disciplines. CCA has been performed to explore the intellectual structure of medical informatics (Andrews 2003; Morris and McCain 1998), biomedical engineering (Garfield 1986), biotechnology (McCain 1995; Rip and Courtial 1984), genetics (McCain 1991), and so on. By conducting CCA for journals, Morris and McCain (1998) found medical informatics' major focus areas of early 1990s. Through CCA of the 50 most-cited American College of Medical Informatics fellows, Andrews (2003) clustered the research field of medical informatics.

The most used method, analysis of words and co-words, CWA, can reveal the overall picture and broad landscape and boundaries of a given field. CWA can overcome some of the limitations of citation analysis for domains in which citing is irregular or absent (Callon et al. 1991). CWA uses patterns of co-occurrence of pairs of words or phrases in a corpus of texts to identify the relationships between ideas within the subject areas presented in texts (He 1999). Since their introduction (Callon et al. 1983), CWA techniques have been used to explore intellectual structure (Bansard et al. 2007; Mane and Börner 2004; Morris 2000), analyze research trends (Bansard et al. 2007; Garfield 1986; He 1999; Pickens and MacFarlane 2006; Rebholz-Schuhman et al. 2007), and generate hypothesis and discover knowledge (Stegmann and Grohmann 2003, 2005; Swanson 1986a, b, 1988). In spite of some problems (Fuller et al. 2004; Lisacek et al. 2005; Pickens and MacFarlane 2006), co-word analysis is the most common technique for extracting themes from the subject content of texts and exploring intellectual structure and global dynamics of knowledge domains.

Research source data

For knowledge domain exploration, the most extensively used data sources have been journals, articles, and citation data. The approach, however, necessarily limits the scope of

analysis. One cannot confirm that analysis results sufficiently represent the updated structure of a domain. This is because the analysis of journal articles has a time lag of 2–5 years between research endeavors and their reflection in the database (Martens and Saretzki 1993). Journal-based metrics also poorly capture topical distinctions in fast-evolving fields.

Let's turn to other scientific communication outlets, i.e., "scholarly events" such as conferences, workshops, symposia, and seminars. Scholarly events demonstrate cutting-edge research fronts, thus provide opportunities to identify updated scientific research results. The use of scholarly event data—calls for papers topics, presentations, program committee member lists, and so on—is highly promising for the analysis of the intellectual structure of a domain. For example, 'call-for-paper topic lists' are indicators of a scientific community's interest and orientation. They are representative of domain topics, which are not simply lists of terms but a potentially valuable data source for domain analysis. In spite of their potential as research tools, there have been few attempts to apply these topics to map the structure of biomedical informatics. The reason is that they are not easily obtained from existing bibliographic databases; they are mainly delivered through scholarly events' websites, mailing lists, or academic societies' websites.

Some informative research efforts have used *scholarly events* as data sources instead of journals, articles, and citation data. Kranakis and Leydesdorff (1989) demonstrated the significance of scholarly events as sources, particularly with respect to the emergence and growth of science. Martens and Saretzki (1993, 1994) demonstrated that analysis of event program records has some advantages over existing bibliometric approaches, because event program records are not restricted by the time lag of bibliographic databases. Synnestvedt and Chen (2003) visualized the domain of medical informatics using the American Medical Informatics Association Symposium Proceedings. Söderqvist and Silverstein (1994) mapped the sub-disciplinary structure of a scientific domain through frequent event participants' data with the assumption that two meetings are more similar than two other meetings if they have more overlapping participants. They argue that their method may better represent the cognitive interests of scientists than co-citation analysis in that the latter does not necessarily reflect cognitive interests. Other studies stress that proceedings have greater importance than articles in knowledge transfer (Glänzel et al. 2006; Godin 1998; Lisée and Larivière 2008; Matsuo et al. 2003). Actually, in biomedical informatics the most current scholarly work tends to appear in events for more rapid knowledge dissemination (Jeong et al. 2009).

In this study, we utilize scholarly events' topics and presentations as the main data source for exploring the intellectual structure of biomedical informatics. The use of scholarly events' topics as the main *input variables* and papers as their *cases* is based on the assumption that the topics covered in scholarly events are building blocks of biomedical informatics, and they reflect forefront research topics communicated by event participants. The following questions guided this research: (1) What are the important topics of biomedical informatics, which comprises medical informatics and bioinformatics, covered in scholarly events? (2) What are the commonalities and differences between the two sub-domains reflected in scholarly events? (3) What are the newly emerging topics?

This paper begins with data and analysis methods to find the intellectual structure of biomedical informatics. Then research findings such as important research topics and their usage contexts, and newly emerging topics are covered. Finally, we discuss the major findings and conclude the paper with further research directions. We use "event" and "conference" interchangeably to refer to a "scholarly event".

Methods

In this study, we adopted a well established co-word analysis protocol. It involves the following steps: (1) select the text corpus for the study (in our case, papers presented at biomedical informatics scholarly events); (2) extract and normalize the terms and get term weights; (3) get a term co-occurrence frequency matrix for the corpus; (4) get term-term relatedness; (5) analyze the term-term relatedness matrix, and visualize it. For this study we developed a java application, called BiKE Text Analyzer (BTA). The co-word analysis workflow and BTA screen shot are shown in Online Supplement, Fig. 1.

Collecting scholarly event data

The time window for the target scholarly events was set as the 10 years from 1999 to 2008. In order to decide on the number of scholarly events for study, we performed the following selection processes. First, we used various event listing databases (e.g., http://conference Alerts.com, http://EventSeer.net), and academic societies' homepages (e.g., *IMIA*, and *International Society for Systems Biology*). Since there is no single authoritative source that comprehensively includes representative scholarly events, we searched and retrieved from the above data sources biomedical informatics-related event series with the following inclusion criteria: (1) longer than 5 years in existence; (2) catering to global audiences; (3) is only parent event series (not child events included in an event series); (4) has a peer-reviewing system such as call-for-papers, and committees. Applying these selection criteria, as Table 1 shows, we ended up with 11 event series in medical informatics and bioinformatics. From

Table 1 Eleven selected scholarly event series in the biomedical informatics

Field	Event series full name	Acronym	Established year	Papers
Medical Informatics	AMIA Annual Symposium	AMIA	1977	3,136
	International Congress of the European Federation for Medical Informatics	MIE	1988	4,665
	World Congress on Health (Medical) Informatics	MedInfo	1974	867
	Conference on Artificial Intelligence in Medicine	AMIE	1985	297
	IEEE International Symposium on Computer-Based Medical Systems	CBMS	1988	895
Bioinformatics	IEEE Symposium on Bioinformatics and Bioengineering	BIBE	2000	757
	Annual International Conference on Computational Systems Bioinformatics	ICSB	2002	602
	Annual International Conference on Intelligent Systems for Molecular Biology	ISMB	1993	490
	Annual International Conference on Research in Computational Molecular Biology	RECOMB	1997	424
	Computational Methods in Systems Biology	CMSB	2003	118
	IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology	CIBCB	2004	285
Total Sum	11			12,536

these event series, we collected the titles and abstracts of 12,536 presented papers from the event web sites, PubMed, BioMed Central, IEEE Explorer, and Springer. The number of papers in medical informatics was 9,860, and bioinformatics 2,676.

Term identification

In a co-word analysis, the critical step is to create a list of terms that constitute the variables for analysis. Our variables were created from a combination of the sources including: (1) event topic lists; (2) author keywords and biomedical informatics keywords from Thomson Web of Science; and (3) MeSH descriptors. We collected event topics from 'call for paper topics' and 'session titles' in all the events that, on top of the selected events, include those events that were excluded by the event selection criteria. The collected terms were appended to the Vocabulary Manager of BTA. The Vocabulary Manager automatically erases duplicate terms and manages n-grams (up to 5-grams), and it allows users to load new vocabularies and input new terms.

Term normalization

Topics are difficult to represent with single words because they often have more than one meaning. In most cases, topics are appropriately described in multi-word phrases, which, especially in research domains, are much more interpretable (Mann et al. 2006). In this study, we view a topic as a multi-word phrase rather than a single word. Before we proceeded to extract phrases and obtain precise variables, words were normalized and their synonyms controlled. We adopted a less strict normalization strategy for words: tokenized words' plural forms were singularized (e.g., records to record) and abbreviations were controlled with synonym lists (e.g., EHR = electronic health record).

Term extraction

From event paper corpora (medical informatics = 9,860; bioinformatics = 2,676), we extracted n-gram terms (from 2- to 5-grams) as variables using BTA. In medical informatics (MI), the total number of extracted terms was 6,958, while bioinformatics (BI) 5,436. The most frequently occurring term in MI was "information system" (1,536), while in BI "gene expression" (720). After excluding the terms that occurred less than 10 times, the term variables for analysis amounted to 1,621 in MI and 815 in BI.

Term weight

Since not all terms may have the same importance in a document, the weight of each term was calculated by multiplying term frequencies (TF) by the inverse document frequency (IDF) for that term:

$$W_{i,j} = \mathrm{TF} \times \mathrm{IDF} = \frac{f_{i,j}}{\sum_k n_{k,j}} \times \log \frac{N}{n_i}$$

where $f_{i,j}$ is the number of times the term i appears in the document j, $\sum_k n_{k,j}$ is the total number of terms in the document d, N is the total number of documents, and n_i is the total number of documents containing the term i.

Term co-occurrence and relatedness matrix

The co-occurrence analysis approach quantifies term co-occurrences in documents. It assumes that the more frequently two terms appear together in the same document, the sooner they will be identified as being closely related (Noyons 2001). BTA generates a term co-occurrence frequency matrix ($1{,}621 \times 1{,}621$ in MI and 825×825 in BI), and then transforms the matrix into a cosine correlation matrix, where each cell indicates the relative closeness of each term pair with a 0–1 range. The cosine measure is defined as the cosine of the angle enclosed between two term vectors x and y:

$$\text{Cosine}(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\left(\sum_{i=1}^{n} x_i^2\right) \times \left(\sum_{i=1}^{n} y_i^2\right)}}$$

Converting matrix into network and visualization

In the current study, the term-term relatedness matrix was converted to a social network showing the binary relationships between any two terms. This network provides a useful medium for representing the topical structure of medical informatics and bioinformatics in a concise and intuitive manner. Pajek software was used for network visualization and analysis. The node size equals to the logarithm of the term frequency, and the thickness of the lines indicates the cosine value between a pair of terms.

Results

Important topics

Some terms have links with many terms; their network of co-occurrences is quite extensive and occupies a central position in a field. To identify the important topics both in MI and BI, authority and hub scores were calculated for each topic. In social network analysis, if a vertex points to many good authorities, it is a good hub. And if a vertex is pointed to by many good hubs, it is a good authority (Kleinberg 1999). The authority scores and hub scores of topics are rendered as:

$$\text{Authority Score}(T_i) = \sum_{T_j \to T_i} \text{Hub Score}(T_j)$$

$$\text{Hub Score}(T_i) = \sum_{T_i \to T_j} \text{Authority Score}(T_j)$$

The authority score of a topic i (T_i) equals the sum of the hub scores of all topics (T_j) that point to it. The hub score (T_i) of a topic i (T_i) equals to the sum of the authority scores of all topics that it points to. Authority scores mutually reinforce hub scores. We extracted the 100 most important topics each for MI and BI during the past 10 years (1999–2008). Some of the important topics in MI (grouped into 40 components) were "clinical decision support system", "clinical information system", "computerized order entry", "electronic health record", and so on as shown in Online Supplement, Fig. 2. In BI (grouped into 45 components), "gene expression data (profile)", "protein (RNA) secondary structure", "gene selection", "microarray data analysis" were included in the important topics as shown in Online Supplement, Fig. 3.

Usage contexts

Usage contexts of important topics

The most important topic in MI is "decision support" (hub weight = 0.199; authority weight = 0.199), while in BI "gene expression" (hub weight = 0.199; authority weight = 0.199). How are these topics used in each of the two domains? To answer the question we analyzed their usage contexts by extracting their *k-nearest neighbors* in the network. Here, a neighbor is a vertex (topic) adjacent to another vertex in a network. The usage context of "decision support" in MI is "semantic network", "unified medical language system", "practice pattern", "grid infrastructure", "quality improvement" and so on (Online Supplement, Fig. 4). In BI, "decision support", which is not included in the 100 important topics, is connected to "system biology", "medical image", "high throughput data" via "system level", and is also linked to "alzheimer disease" via "computational framework" (Online Supplement, Fig. 5). The topic "decision support" in MI is related to terminology system and quality of care, while, in BI, to an application of biological research result.

The most important topic in BI, "gene expression", is surrounded by "gene expression profile", "gene expression pattern", "gene regulation", etc. (Online Supplement, Fig. 6). While in MI, it is the center for the topics "genetic information" and "clinical diagnosis" (both via "microarray technology"), and is the background knowledge for "informatics education" and "decision process" (Online Supplement, Fig. 7).

Usage contexts of methodological topics

MI and BI share several methodology-related topics: "machine learning", "support vector machine", and "data mining". But their usage contexts are different. For example, in the MI context (Online Supplement, Fig. 8), "machine learning" is adjacent to diseases (i.e., "diabetes", "breast cancer", "chronic hepatitis") and clinical settings (i.e., "clinical data repository"). Whereas, in the BI context (Online Supplement, Fig. 9), it neighbors "structural information" connected to "sequence pattern", and "information retrieval" and computational and statistical topics such as "clustering method", "false discovery rate", and "classification accuracy".

Newly emerging topics

To investigate newly emerging topics, for each 2-year period, we calculated raw term frequencies and identified the topics which represented the lowest 10% in the low frequency group in the preceding period(s), and which also remained in the highest 10% in the high frequency group in the following periods. Take "snomed ct" for example, it belonged to the lowest 10% at 1999–2000, but since 2001–2002, it had manifested itself as one of the most highly mentioned topics. During 2001–2002, as shown in Online Supplement, Table 1, newly emerging topics in MI were "snomed ct", "e health", "record linkage", "description logic", "conceptual framework", etc., while in BI were "monte carlo", "time series", "biological sequence", "genomic data", etc. Since 2003–2004, in MI, "web service", "microarray data", "feature extraction", etc. were frequently used, while in BI "gene ontology", "system biology", "gene selection", etc. were extensively mentioned. Since 2005, MI has been extensively using "grid technology", whereas BI has been visiting "ppi network", "prostate cancer" and "network motif". During 2007–2008,

"categorical structure" and "process mining" were frequently mentioned in MI, while in BI "intrinsically disordered protein", "active contour", "peak detection", "bone remodeling", "clinical trial", etc. were frequently mentioned. In general, the most frequent emerging topics in the MI domain were "snomed ct" ($tf = 267$), "e health" ($tf = 214$) and "patient safety" (2000, $tf = 139$). In the BI domain they were "interaction network" ($tf = 127$), "feature selection" ($tf = 120$) and "time series" ($tf = 105$).

Discussion

The networks produced by the term co-occurrence pattern analysis of biomedical informatics scholarly events seem to produce quite clear and interpretable results. We identified, during the past 10 years, the major topics in MI and BI. The induced networks of the important topics suggest that both fields have their own research orientations. The major topics in MI (e.g., "healthcare information system", "electronic health record", "clinical decision support system", etc.) provide evidence that MI mainly deals with the resources, devices, and methods to optimize information flow for the purpose of better healthcare. The major research topics in BI (e.g., "gene expression", "gene regulation", "protein and RNA structure", etc.) suggest that BI mainly deals with molecular biological processes. Though several algorithmic methods were shared by the two communities, they are used in different contexts in our analysis. In MI, they are associated with diseases and clinical practice, whereas in BI, they are instrumental in scientific discovery. The prevalence of methodological topics in BI such as "monte carlo", "neural network", "clustering algorithm", and "dynamic programming" entails that BI takes more computational and statistical approaches than MI does. This evidence supports our belief that MI uses computational technologies to improve the productivity in clinical settings, while BI uses them as tools for scientific biological discovery.

   Such topics as "expression data" (since 2001–2002), "microarray data" (since 2003–2004), "sequence alignment" and "protein ontology" (since 2005–2006), and "transcription factor" (since 2007–2008)—all usually considered as BI topics—became more prevalent in MI. Similar phenomena were noticed in BI: such MI topics as "prostate cancer" (since 2005–2006), "bone remodeling", "clinical trial", and "eeg signal" (since 2007–2008) began more prevalent. Though MI and BI are arguably separate research fields, their topics are increasingly intertwined, and the gap between the two fields are blurred, forming a unified informatics—namely biomedical informatics. This phenomenon is supported by other research efforts. Malin and Carley (2007) investigated how editorial boards of MI and BI journals integrate researchers from the two communities. Rebholz-Schuhman et al. (2007) showed a similar result by analyzing journal articles as data sources. In our case, however, there is a time gap in adoption of counterpart's topics. The emergence of BI topics in MI started from 2001 to 2002, whereas that of MI topics in BI started only from 2005 to 2006: in other words, MI was faster in bringing scientific results from BI in the clinical investigation (from early 2000s) than BI which didn't pay attention to MI topics until 2005.

   Several contributions of this research are worth mentioning. First, this study provides topic networks for better understanding of the intellectual structure of biomedical informatics, and to help gain insight into the main boundaries of its sub fields, MI and BI. The results are useful to novices and researchers who are switching from MI to BI and vice versa. Our usage context analysis of the research topics provides more contextualized understanding of biomedical informatics. Secondly, the use of *n*-grams as analysis

variables (in our case, the longest *n*-grams is five words in length) is more advantageous than the case of single terms or bi-grams being used only (Rebholz-Schuhman et al. 2007), which can tend to be ambiguous and less able to capture to represent detailed topics. For example, a central concept, "decision support" can be specialized with modifiers such as "clinical" and "system" to make "clinical decision support" or "clinical decision support system". Thirdly, we used scholarly events as data sources for domain analysis in order to more precisely capture the up-front status of biomedical informatics. While journal literature citation analysis could identify intellectual linkage among highly cited literature (McCain and Silverstein 2007) and emerging research topics history (Synnestvedt et al. 2005), it may not as easily capture upfront research states, since the data is unavailable until several years after publication time. Our method also materializes more contextualized and deeper domain analysis in comparison to the use of controlled vocabularies such as MeSH and INSPEC descriptors (Morris 2001). This argument comes from the assumption that subject headings (codes) unavoidably limit the number of topics which ends up oversimplifying the domain structure, and has difficult capturing newly generated topics; our method can reflect new or specialized knowledge which is not yet fully matured. Thus, by exploiting scholarly events as an analysis data source, future studies can use our method to explore the intellectual structure of any ever evolving discipline. The implication of this study is also significant from scholarly event organization and education perspectives. The methodology to track newly emerging topics can help scholarly event organizers to shape their meeting scopes and agenda for forthcoming events. And the findings can be of help to decide which technologies and themes should be included in biomedical informatics curriculum to meet learners' needs.

Note that there are technical limitations to our study. Several advantages of using the *n*-grams as a text analysis unit can also be viewed as a drawback: an *n*-gram may not catch important topics that are single words (i.e., 'terminology'). In our study, however, even the combination of single-word and *n*-gram topics did not result in a desirable product since they lost the highly pre-coordinated meaning of the topics. We do not regard the topics covered in this study as representatives of all the topics in biomedical informatics, nor did we exhaust all the scholarly events in biomedical informatics. The events which are not included in our work exist, and if desired, they could be included as a data source by using sourcing tools such as a scholarly event data crawler, which is under development. However, we are confident that the 11 scholarly events are sufficient enough to ascertain our research goals.

Conclusion

We investigated the intellectual structure of biomedical informatics by analyzing social networks of research topics reflected in scholarly events. We identified the major topics of both medical informatics and bioinformatics. And we found that shared methodology-related topics were used in different contexts. We were able to find newly emerging topics in biomedical informatics, which can allow for a better understanding of the forefront biomedical informatics.

The use of scholarly events as data sources opens further research questions: (a) How rapidly does the agenda of scholarly events in a given field change? (b) Which events focus on which topics and how topically similar are they? (c) Are there any differences between the results of subject analysis on scholarly events and scientific journals?

References

Andrews, J. E. (2003). An author co-citation analysis of medical informatics. *Journal of the Medical Library Association, 91*(1), 47–56.

Bansard, J. Y., Rebholz-Schuhmann, D., Cameron, G., Clark, D., van Mulligen, E., Beltrame, E., et al. (2007). Medical informatics and bioinformatics: A bibliometric study. *IEEE Transaction on Information Technology in Biomedicine, 11*(3), 237–243.

Callon, M., Courtial, J., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics, 22*(1), 155–205.

Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information, 22*(2), 191–235.

David, F. (2005). Introduction to bioinformatics. *Journal of the American Society for Information Science and Technology, 56*(5), 440–446.

Fuller, S., Revere, D., Bugni, P., & Martin, G. (2004). A knowledgebase system to enhance scientific discovery: Telemakus. *Biomedical Digital Libraries, 1*(1), 2.

Garfield, E. (1986). Mapping the world of biomedical engineering: Alza lecture (1985). *Annals of Biomedical Engineering, 14*(2), 97–108.

Glänzel, W., Schlemmer, B., Schubert, A., & Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. *Scientometrics, 68*(3), 457–473.

Godin, B. (1998). Measuring knowledge flows between countries: The use of scientific meeting data. *Scientometrics, 42*(3), 313–323.

Greenes, R. A., & Shortliffe, E. H. (1990). Medical informatics. An emerging academic discipline and institutional priority. *JAMA, 263*(8), 1114–1120.

Hasman, A., & Haux, R. (2006). Modeling in biomedical informatics—an exploratory analysis (part 1). *Methods of Information in Medicine, 45*(6), 638–642.

Hasman, A., & Haux, R. (2007). Modeling in biomedical informatics: An exploratory analysis part 2. *International Journal of Medical Informatics, 76*(2–3), 96–102.

Hasman, A., Haux, R., & Albert, A. (1996). A systematic view on medical informatics. *Computer Methods and Programs in Biomedicine, 51*(3), 131–139.

He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends, 48*(1), 133–159.

Jeong, S., Lee, S., & Kim, H.-G. (2009). Are you an invited speaker? A bibliometric analysis of elite groups for scholarly events in bioinformatics. *Journal of the American Society for Information Science and Technology, 60*(6), 1118–1131.

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM, 46*(5), 604–632.

Kranakis, E., & Leydesdorff, L. (1989). Teletraffic conferences: Studying a field of engineering science. *Scientometrics, 15*(5), 563–591.

Lisacek, F., Chichester, C., Kaplan, A., & Agnes, S. (2005). *Discovering paradigm shift patterns in biomedical abstracts: Application to neurodegenerative diseases*. Paper presented at the first international symposium on semantic mining in biomedicine.

Lisée, C., & Larivière, V. (2008). Conference proceedings as a source of scientific information: A bibliometric analysis. *Journal of the American Society for Information Science and Technology, 59*(11), 1776–1784.

Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine, 40*(4), 346–358.

Malin, B., & Carley, K. (2007). A longitudinal social network analysis of the editorial boards of medical informatics and bioinformatics journals. *Journal of American Medical Informatics Association, 14*(3), 340–348.

Mane, K. K., & Börner, K. (2004). Mapping topics and topic bursts in PNAS. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 5287–5290.

Mann, G. S., Mimno, D., & McCallum, A. (2006). *Bibliometric impact measures leveraging topic analysis*. Paper presented at the 6th ACM/IEEE-CS joint conference on digital libraries.

Maojo, V., & Kulikowski, C. A. (2003). Bioinformatics and medical informatics: Collaboration on the road to genomic medicine? *Journal of American Medical Informatics Association, 10*(6), 515–522.

Martens, B., & Saretzki, T. (1993). Conferences and courses on biotechnology. Describing scientific communication by exploratory methods. *Scientometrics, 27*(3), 237–260.

Martens, B., & Saretzki, T. (1994). Quantitative analysis of thematic structures in the field of biotechnology: A study on the basis of conference data. *Scientometrics, 30*(1), 117–128.

Matsuo, Y., Tomobe, H., Hasida, K., & Ishizuka, M. (2003). Mining social network of conference participants from the web. In *Proceedings of the international conference on web intelligence* (pp. 190–194).

McCain, K. W. (1991). Core journal networks and cocitation maps: New bibliometric tools for serials research and management. *Library Quarterly, 61*(3), 311–336.

McCain, K. W. (1995). Biotechnology in context: A database-filtering approach to identifying core and productive non-core journals supporting multidisciplinary R & D. *Journal of the American Society for Information Science, 46*(4), 306–317.

McCain, K. W., & Silverstein, S. M. (2007). *Tracing persistent highly visible research themes in medical informatics*. Paper presented at the 2006 AMIA spring congress. Retrieved from http://hdl.handle.net/1860/1594.

Morris, T. A. (2000). *Structural relationships within medical informatics*. Paper presented at the AMIA 2000 annual symposium.

Morris, T. A. (2001). *Structural relationships within medical informatics: A classification/indexing co-occurrence analysis*. Philadelphia: Drexel University.

Morris, T. A., & McCain, K. W. (1998). The structure of medical informatics journal literature. *Journal of the American Medical Informatics Association, 5*(5), 448–466.

Noyons, E. (2001). Bibliometric mapping of science in a policy context. *Scientometrics, 50*(1), 83–98.

Pickens, J., & MacFarlane, A. (2006). *Term context models for information retrieval*. Paper presented at the 15th ACM international conference on information and knowledge management.

Rebholz-Schuhman, D., Cameron, G., Clark, D., van Mulligen, E., Coatrieux, J.-L., Del Hoyo Barbolla, E., et al. (2007). SYMBIOmatics: Synergies in medical informatics and bioinformatics—exploring current scientific literature for emerging topics. *BMC Bioinformatics, 8*(Suppl 1), S18.

Rip, A., & Courtial, J. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics, 6*(6), 381–400.

Söderqvist, T., & Silverstein, A. M. (1994). Participation in scientific meetings: A new prosopographical approach to the disciplinary history of science—the case of immunology 1951–72. *Social Studies of Science, 24*, 513–548.

Stegmann, J., & Grohmann, G. (2003). Hypothesis generation guided by co-word clustering. *Scientometrics, 56*(1), 111–135.

Stegmann, J., & Grohmann, G. (2005). Transitive text mining for information extraction and hypothesis generation. Retrieved July 10, 2008, from http://arxiv.org/abs/cs/0509020.

Swanson, D. (1986a). Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine, 30*(1), 7–18.

Swanson, D. (1986b). Undiscovered public knowledge. *Library Quarterly, 56*(2), 103–118.

Swanson, D. (1988). Migraine and magnesium: Eleven neglected connections. *Perspectives in Biology and Medicine, 31*(4), 526–557.

Synnestvedt, M. B., & Chen, C. (2003). *Visualizing AMIA: A medical informatics knowledge domain analysis*. Paper presented at the AMIA annual symposium. Retrieved from http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1479953&blobtype=pdf.

Synnestvedt, M. B., Chen, C., & Holmes, J. H. (2005). *Visual exploration of landmarks and trends in the medical informatics literature*. Paper presented at the 2005 AMIA annual symposium. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1560559.