Journal of Informetrics xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Diffusion of latent semantic analysis as a research tool: A social network analysis approach

Yaşar Tonta [a], Hamid R. Darvish [b,*]

[a] Department of Information Management, Faculty of Letters, Hacettepe University, 06800 Beytepe, Ankara, Turkey
[b] Department of Computer Engineering, Faculty of Engineering and Architecture, Çankaya University, 06530, Balgat, Ankara, Turkey

ARTICLE INFO

Article history:
Received 23 August 2009
Received in revised form 25 October 2009
Accepted 9 November 2009

Keywords:
Latent semantic analysis
Social network analysis
Co-authorship analysis
Cluster analysis

ABSTRACT

Latent semantic analysis (LSA) is a relatively new research tool with a wide range of applications in different fields ranging from discourse analysis to cognitive science, from information retrieval to machine learning and so on. In this paper, we chart the development and diffusion of LSA as a research tool using social network analysis (SNA) approach that reveals the social structure of a discipline in terms of collaboration among scientists. Using Thomson Reuters' Web of Science (WoS), we identified 65 papers with "latent semantic analysis" in their titles and 250 papers in their topics (but not in titles) between 1990 and 2008. We then analyzed those papers using bibliometric and SNA techniques such as co-authorship and cluster analysis. It appears that as the emphasis moves from the research tool (LSA) itself to its applications in different fields, citations to papers with LSA in their titles tend to decrease. The productivity of authors fits Lotka's Law while the network of authors is quite loose. Networks of journals cited in papers with LSA in their titles and topics are well connected.

1. Introduction

The technique of latent semantic analysis (LSA) was patented on June 13, 1989 by Deerwester et al. (1989). LSA is a fully automatic mathematical/statistical technique for extracting meaning and inferring relations of expected contextual usage of words in passages of discourse. It is not a traditional natural language processing or artificial intelligence program, as it uses no humanly constructed dictionary, knowledge bases, semantic networks, grammars, syntactic parsers, or morphologies. Instead, LSA "uses singular value decomposition [SVD], a general form of factor analysis, to condense a very large matrix of word-by-context data into a much smaller, but still large, typically 100–500 dimensional representation" (Kitajima, Kariya, Takagi, & Zhang, 2005).

One of the very first applications of LSA has been in information retrieval. The formal description of LSA was first published in an information science journal in the context of indexing (Deerwester, Furnas, Landauer, & Harshman, 1990). In fact, the inventors of LSA published two papers on latent semantic indexing before their seminal paper and before they were awarded the patent (Deerwester, Dumais, Landauer, Furnas, & Beck, 1988; Lochbaum & Streeter, 1989). Whereas Boolean or vector space models are based entirely on the strict matching of terms that appear in users' queries with those in the bibliographic records or full-texts of documents, indexing by LSA does not necessarily rely on the occurrence or absence of certain terms. LSA can detect the meaning even though the terms in the user's query are absent in the text or are described using different

* Corresponding author.
E-mail addresses: tonta@hacettepe.edu.tr (Y. Tonta), darvish@cankaya.edu.tr (H.R. Darvish).

1751-1577/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2009.11.003

terms. LSA overcomes the synonymy (different words with the same meaning, e.g., automobile-car) and polysemy (the same word with different meanings, e.g., apple as fruit and apple as computer) problems in information retrieval by capturing the latent semantic relations between terms (Deerwester et al., 1990; Landauer, Foltz, & Laham, 1998).[1]

LSA has quickly become a popular research technique and has been put to use in different fields. In addition to information retrieval, LSA has been used in cognitive science, knowledge acquisition, machine learning, intelligent tutoring systems, and computational biology (for remote homology detection between protein sequences), among others. LSA has been instrumental in the study of knowledge acquisition, induction and representation, which is called "Plato's problem" and was tackled earlier by many psychologists, linguists, and computer scientists (e.g., Angluin & Smith, 1983; Chomsky, 1991; Jackendoff, 1992; Michalski, 1983; Pinker, 1990; Shepard, 1987; Vygotsky, 1968). Landauer and Dumais (1997) approached Plato's problem with LSA and analyzed a large corpus of natural text and generated a representation that captures the similarity of words and text passages. They proposed that LSA constitutes a fundamental computational theory of acquisition and representation of knowledge and explained how the LSA modeling technique imitates the human knowledge acquisition and induction process.

The Landauer and Dumais study sparked an interest and set the infrastructure for scholarly works in a variety of scientific fields using the LSA technique. In addition to hundreds of articles on LSA and citations thereof, the original patent of Deerwester et al. (1989) was referenced by 147 different patents in the USPTO[2] database since 1989. In this paper, we attempt to chart the development and diffusion of LSA as a research tool by combining bibliometric and social network analysis techniques such as citation analysis, co-authorship analysis and cluster analysis. We investigate the collaboration patterns of scientists doing research on LSA. What follows are the preliminary findings of our exploratory study.

2. Literature review

Bibliometrics is defined as "the application of mathematical and statistical methods to books and other media of communication" (Pritchard, 1969). For example, the productivity of authors is tested using Lotka's Law, which states that the number of authors contributing n papers would constitute $1/n^2$ of those contributing one paper and that the proportion of authors contributing only one paper is about 60% of all authors (Hertzel, 1987, p. 159). Thus, about 60% of authors studying in a certain field would publish just one article, 15% two articles, 6.6% three articles, and so on. Lotka's Square Law can be defined mathematically as a function $f(n) = C/n^{\alpha}$, where $f(n)$ is the frequency function and C and α are constants ($C > 0$ and $\alpha \geq 0$). The number of authors publishing n papers is determined by the law of diminishing returns (Egghe, 2005, p. 14). Citation and co-authorship analyses measure the impact of authors' contributions and identify their scientific collaboration patterns, respectively (Price, 1970). Scientometricians use co-authorship patterns to predict new trends in scientific fields (Glänzel, 2002).

Social network analysis (SNA), on the other hand, has become a widely accepted tool to reveal and map the structures of social networks. SNA consists of actors (or nodes) and ties, actors being persons, teams or companies and ties being friendship between several people, collaboration between teams and business relationships between companies (Newman, 2004). SNA is based on graph theory and uses terms such as density (connectedness of the graph) and centrality measures (relationships between nodes in terms of degree, closeness and betweenness) to conceptualize social structures as networks (Otte and Rousseau, 2002). The density of a network is the number of actual connections between members divided by the number of possible connections (Scott, 2000). The centrality of the network, on the other hand, measures the degree to which it approaches the configuration of a "star" network (Scott et al., 2005). Measuring of a node's centrality reveals the importance of the node's position in a network (Chen, 2006). Degree centrality is the number of direct relationships that a node has. Betweenness centrality is an indicator of a node's ability to make connections to other nodes in a network while closeness centrality measures how quickly a node can access more nodes in a network (Sentinel Visualizer, 2009). Betweenness centrality is a widely used centrality metric (Freeman, 1977).

White, Wellman, and Nazer (2004) tested longitudinally if social and intellectual ties among 16 members of an interdisciplinary research group had an impact on their citing behaviors of each other's work. They found that intellectual ties based on shared-content did better as predictors of intercitation behavior than social ties and that members being cocited tend to cite each other's work more often. Newman (2001) used SNA techniques in three repositories (MEDLINE, arXiv and NCSTRL) to construct collaboration networks among scientists in different fields (medicine, physics and computer science, respectively). Similarly, Hou, Kretschmer, & Liu (2008) used SNA to illustrate the structure of social network collaboration in the journal *Scientometrics*. Leydesdorff (2007) showed that betweenness centrality is a measure of interdisciplinarity of scientific journals in local citation environments whereas closeness provides a global measure of multidisciplinarity within a journal set.

SNA techniques enable researchers to visualize scholarly collaboration in different scientific fields (Otte and Rousseau, 2002). From the standpoint of network visualization and citation analysis, network nodes are classified into three, namely, landmark nodes, hub nodes, and pivot nodes:

[1] For more information on LSA, see http://lsa.colorado.edu.

[2] The search was performed on 21 January 2009 in the USPTO database. See http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2Fsearch-adv.htm&r=0&f=S&l=50&d=PALL&Query=ref/4839853.

Please cite this article in press as: Tonta, Y., & Darvish, H.R. Diffusion of latent semantic analysis as a research tool: A social network analysis approach. *Journal of Informetrics* (2009), doi:10.1016/j.joi.2009.11.003

> A highly cited article tends to provide an important landmark regardless of how it is cocited with other articles. . .A hub node has a relatively large node degree; a widely cocited article is a good candidate for significant intellectual contributions. . .Pivot nodes are joints between different networks; they are either the common nodes shared by two networks or the gateway nodes that are connected by internetwork links. (Chen, 2004, p. 5305)

Small (2006) used cocitation clusters over three 6-year periods to track the emergence and growth of research areas. Chen (2006) applied "cluster labeling" to co-cited network graphs to reveal new scientific trends. Cluster labeling is achieved by selecting words from co-cited articles in the social network graphs using the CiteSpace software package. Words thus identified tend to lead to new themes and discoveries in scientific fields. Moreover, CiteSpace makes use of the LSA method in this process to list the top ranked terms in each network cluster (Chen, 2006).

3. Data and methods

Data on LSA comes from Thomson Reuters' Web of Science (WoS) database. We searched WoS on January 19, 2009 to identify the literature on LSA by entering the keyword "latent semantic analysis". We restricted our keyword search to Titles and Topics (1990–2008) and obtained a total of 315 papers from WoS: 65 papers with LSA in their titles only; and an additional 250 papers with LSA in their topics (but not in their titles). The latter search retrieved records with LSA in the following fields: abstract, author keywords and keywords plus. Full bibliographic records (including their reference lists) of all papers were downloaded. Bibexcel[3] was used to analyze each paper along with its reference list to carry out citation, co-authorship and cluster analyses. Pajek[4] was used to calculate the density, betweenness and closeness of the structure of social network of LSA. CiteSpace[5] was used to depict the structure of social network as well as to identify the cluster labels in the network of journals cited in papers with LSA in their titles and topics (Chen, 2006). Lotka's Law was used to see if the productivity of authors contributing to the LSA literature fits this regularity. Co-authorship analysis was performed to see the collaboration between scholars using LSA. Cluster analysis was employed to cluster authors as well as journals publishing papers on LSA. Density and centrality measures (closeness centrality and betweenness centrality) were calculated for the social network of LSA. As mentioned earlier, network density "is an indicator for the general level of connectedness of the graph" while the closeness centrality is an indicator of the cohesion of the network and the betweenness centrality measures how nodes facilitate the flow in the network (Otte and Rousseau, 2002, pp. 442–443). Mathematical formulae of these measures are given in Otte and Rousseau (2002). (See also Rousseau and Rousseau, 2000.)

In addition to providing descriptive statistics on LSA in terms of its evolution within the last 20 years, we addressed the following research questions: (1) How fast did LSA as a research tool diffuse and become a part of the regular scientific discourse in different fields? (2) As time passes, an innovation/method or discovery becomes less interesting and scholars tend not to cite the original contributions. Is this also the case for LSA? As LSA becomes a more mainstream research tool, does the number of papers on LSA decrease? To state somewhat differently, do fewer papers with LSA in their titles get published while the number of papers with LSA in their topics increase? We try to address these research questions using bibliometric and SNA techniques.

4. Findings and discussion

The number of papers published between 1990 and 2008 with LSA in their titles and topics is given in Table 1, along with the number of times they were cited. It is clear that the number of papers with LSA both in their titles and topics has increased over the years. The number of papers with LSA in their titles went up from one article in 1990 to 13 articles in 2006, the average being 3.4 papers per year. The corresponding figures for papers with LSA in their topics were one and 46, average being 13 papers. Papers with LSA in their titles were cited a total of 3049 times between 1990 and 2008 while papers with LSA in their topics were cited 1659 times between 1998 and 2008.

Although LSA was patented by Deerwester et al. (1989), the very first journal article by the same authors entitled "Indexing by Latent Semantic Analysis" was published in the *Journal of the American Society for Information Science* in 1990 (Deerwester et al., 1990). Note that no other paper was published on LSA in the next five years. This paper received a total of 1400 citations from journals indexed in Web of Science. The citation figure is well over 4000 when citations from journals that are not indexed in WoS are added. The second important paper on LSA by Landauer and Dumais (1997) was published in *Psychological Review*. It generated a total of 615 citations. Landauer et al. (1998) have also authored an introductory paper on LSA and generated a total of 455 citations. Garfield (2004) considers papers that were cited more than 400 times as "citation classics". These three papers received a total of 2625 citations, two-thirds of all citations (3049) generated by 65 papers.

As the use of LSA as a research tool has increased in other disciplines starting from the late 1990s, the number of papers with LSA in their topics has also increased tremendously. Three times more papers with LSA in their topics have appeared

[3] http://www.se/inforsk/Bibexcel.
[4] http://vlado.fmf.uni-lj.si/pub/networks/pajek/.
[5] http://cluster.cis.drexel.edu/~cchen/citespace.

Please cite this article in press as: Tonta, Y., & Darvish, H.R. Diffusion of latent semantic analysis as a research tool: A social network analysis approach. *Journal of Informetrics* (2009), doi:10.1016/j.joi.2009.11.003

Table 1
Number of publications with "latent semantic analysis" in their titles and topics and number of citations thereof.

Years	# of papers with/citations to LSA in titles		# of papers with/citations to LSA in topics	
	Papers	Times cited (1990–2008)	Papers	Times cited (1998–2008)
1990	1	1,400	0	0
1991	0	0	0	0
1992	0	0	0	0
1993	0	0	0	0
1994	0	0	0	0
1995	0	0	0	0
1996	1	41	0	0
1997	1	615	0	0
1998	5	636	5	2
1999	2	8	4	3
2000	2	11	15	18
2001	4	155	7	22
2002	4	22	20	40
2003	3	26	21	92
2004	8	64	27	117
2005	9	24	32	185
2006	13	32	46	217
2007	5	8	35	341
2008	7	7	38	462
Total	65	3,049	250	1,659

Table 2
Annual distribution of citations received by three citation classics.

Years	Deerwester et al. paper (1990)	Landauer & Dumais paper (1997)	Landauer, et al. paper (1998)
1990	1		
1991	3		
1992	8		
1993	1		
1994	5		
1995	12		
1996	9		
1997	13	1	
1998	29	17	4
1999	22	18	4
2000	38	23	11
2001	26	24	7
2002	53	31	28
2003	71	40	25
2004	81	59	44
2005	97	49	36
2006	120	67	57
2007	74	64	39
2008	82	57	49
Total	765	450	304

Note: Figures are based on WoS. Not all citations are shown.

in the literature in the late 2000s than papers with LSA in their titles. Concomitantly, the number of citations to papers with LSA in their topics has also increased (1659).

Figures can be interpreted as such that the incubation period for LSA lasted about five years. Once LSA was noticed as a novel tool that can be used in a wide variety of applications, it picked up quickly and several papers employing LSA appeared in other disciplines starting from 1998. LSA has become a part of regular scientific discourse within about a decade.

Papers with LSA in their titles generated twice as many citations in total than those with LSA in their topics. However, this statement is misleading in that more than two-thirds of citations to papers with LSA in their titles were generated by three citation classics only, whereas citations to papers with LSA in their topics are more evenly distributed. Note that the number of citations to papers with LSA in their topics has quadrupled within the last five years. As the emphasis moves from the research tool (LSA) itself to its applications, citations to three seminal papers seem to have slowed down in recent years (Table 2, Fig. 1).

We performed a network analysis on authors contributing to the LSA literature. Using CiteSpace, we first identified clusters of researchers including their research fields whose articles contained LSA in their titles and then, using CiteSpace, drew the network structure of LSA researchers. For the sake of clarity, we rearranged the social network graphs. Fig. 2

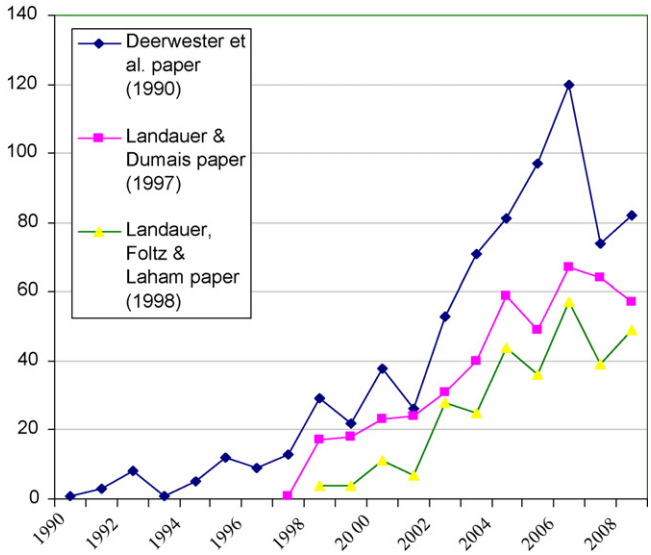

Fig. 1. Citations to seminal LSA papers.

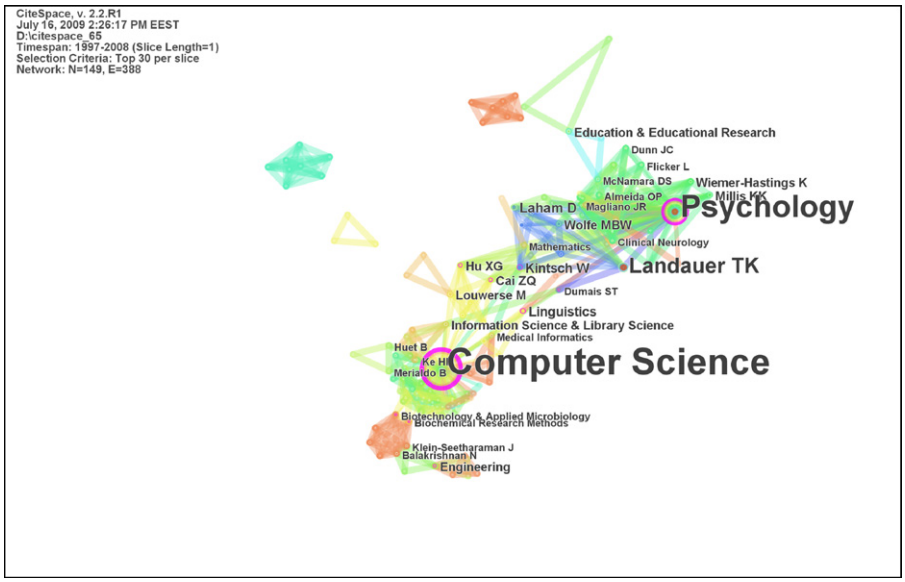

Fig. 2. The network of latent semantic analysis researchers and their research areas.

shows 13 clusters with 132 nodes.[6] The landmark nodes of Computer Science and Psychology are the most crowded clusters containing the most prolific authors. Some of the well-known LSA researchers are in the Psychology cluster, however (e.g., Landauer, Kintsch, and Laham) (Table 3). The hub node Linguistics links the Computer Science and Psychology nodes. The pivot nodes Biotechnology & Applied Microbiology and Biochemical Research Methods perform as a gateway between the Computer Science and Engineering clusters. The Information Science & Library Sciences node is located near the Computer Science cluster.

We used cluster analysis to find out if the structures of networks of journals cited in papers with LSA in their titles and topics differ from each other. Papers with LSA in their titles cited 275 different journals while papers with LSA in topics cited in 1001 journals. Using Bibexcel, we calculated the density, closeness and betweenness centrality measures for both networks of journal sets (Table 4). The structure of the journals network for papers with LSA in titles is slightly more connected (e.g., denser), more cohesive and more flowing (e.g., with journals connecting different groups). The closeness centrality measures

[6] The number of clusters and nodes calculated by Pajek and CiteSpace software packages were almost the same.

Table 3
The most prolific LSA researchers.

Author	# of papers	Author	# of papers
Landauer T.K.	10	Millis K.K.	3
Kintsch W.	4	Hu X.G.	3
Foltz P.W.	4	Louwerse M.	3
Laham D.	4	Dumais S.T.	3
Wiemer-Hastings K.	3		
Cai Z.Q.	3	24 authors	2
Wolfe M.B.W.	3	122 authors	1

Table 4
Centrality measures for journals cited in papers with LSA in titles and topics.

	# of journals	Density	Betweenness	Closeness
LSA in title	275	0.31840	0.06155	0.71958
LSA in topic	1001	0.29466	0.02866	0.69763

for journals cited in papers with LSA in their titles and topics journals are quite high (0.72 and 0.70, respectively). This is an indicator of LSA being a multidisciplinary research area. The betweenness centrality measure for journals cited in papers with LSA in their titles (0.06) is twice as high as that for journals cited in papers with LSA in their topics (0.03), suggesting that the former group is made up of a more interdisciplinary set of scientific journals than the latter one.

In both networks of journals cited in papers with LSA in their titles and topics, the journal *Discourse Processes* occupies the central place because it is the first journal that introduced the LSA method that defined a coherent process of induction theory (Landauer et al., 1998) (see Figs. 3 and 4). The *Journal of the American Society for Information Science (JASIS)* comes next, followed by the *Psychological Review*. *JASIS* has published the seminal article on indexing by LSA (Deerwester et al., 1990) while the article on Plato's problem appeared in the *Psychological Review* (Landauer & Dumais, 1997). The key journals appear in the center of both network graphs, although the places of some journals tend to vary. The network graphs are somewhat dissimilar, however. The three journals mentioned above constitute the landmark nodes in Fig. 3 while this is not the case in Fig. 4.

CiteSpace was used to select labels for co-cited clusters in the social network graphs. The two network graphs were configured in the same way. CiteSpace calculated 30 co-cited clusters for the network of journals cited in papers with LSA in titles, whereas there were 18 co-cited clusters for the network of journals cited in papers with LSA in topic. In the former the terms represented by term numbers 3, 7, and 13 appeared 3 times, term numbers 18, 22, 29 and 28 appeared twice. Fifty eight percent of the overall terms were repeated (see Fig. 5). In the latter one only 0.11 percent of the themes were repeated (see Fig. 6).

Using CiteSpace, we also calculated the top ranked terms per cluster using the LSA method (the clustering algorithm used was "Mutual Information"). The term "Latent Semantic Analysis" occurred in most of the co-cited clusters

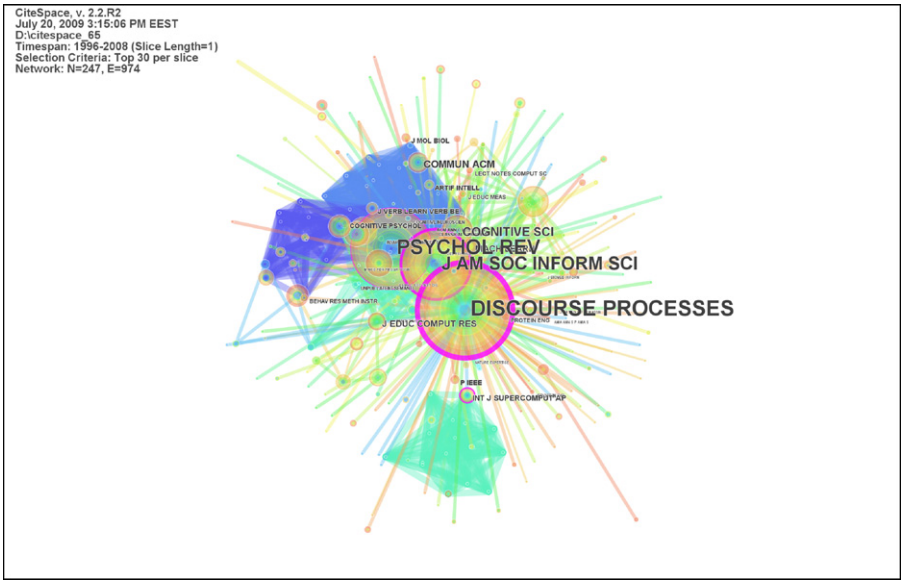

Fig. 3. The network of journals cited in papers with LSA in titles.

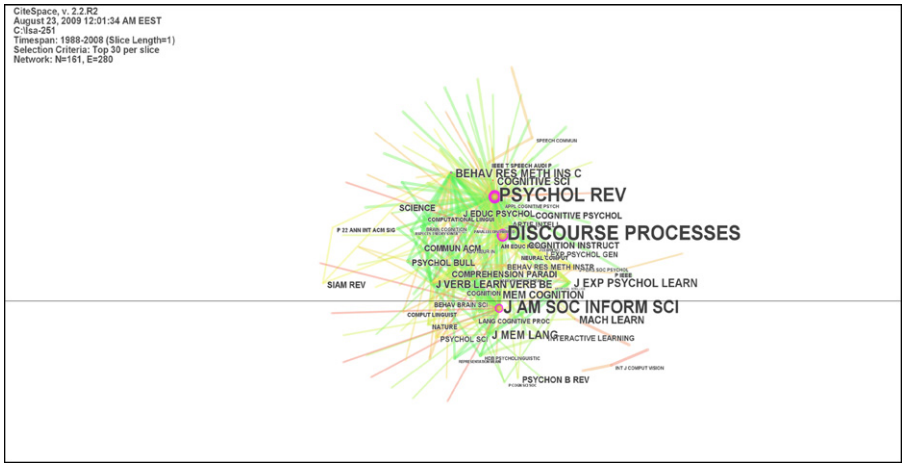

Fig. 4. The network of journals cited in papers with LSA in topic.

in papers with LSA in titles. The term occurred less frequently in the network of journals cited in papers with LSA in topic. Instead, new and somewhat related terms such as speech, intelligent, entropy, schizotypy, visualizing, citation, indicator-assisted, recognition, topographic, animated, and pronouns occurred more often. Although the clusters' labels changed on the basis of the clustering algorithm used (e.g., weighted term frequency (tf/idf), log-likelihood ratio, and mutual information), the top ranked terms produced by the LSA method for all clusters were the same.

The betweenness centrality measure for the journal *Discourse Processes* was the highest for the network of journals cited in papers with LSA in titles, while the *Psychological Review* had the highest betweenness centrality measure for the network of journals cited in papers with LSA in topic. *JASIS* had the second highest measure of betweenness centrality in both networks. The journal *Cognitive Science* had scored a similar centrality values in both network graphs. In summary, the above pattern shows that three journals provide a consistent structure for both social network graphs.

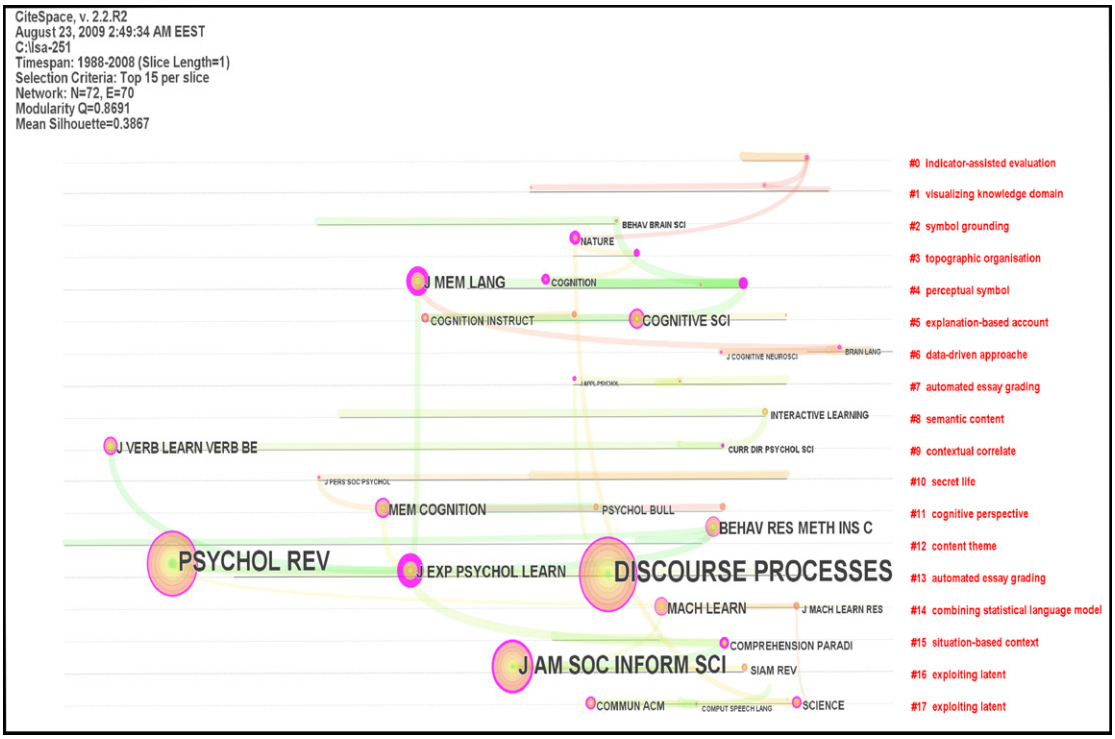

Fig. 5. Number of the terms in co-cited papers with LSA in titles.

Please cite this article in press as: Tonta, Y., & Darvish, H.R. Diffusion of latent semantic analysis as a research tool: A social network analysis approach. *Journal of Informetrics* (2009), doi:10.1016/j.joi.2009.11.003

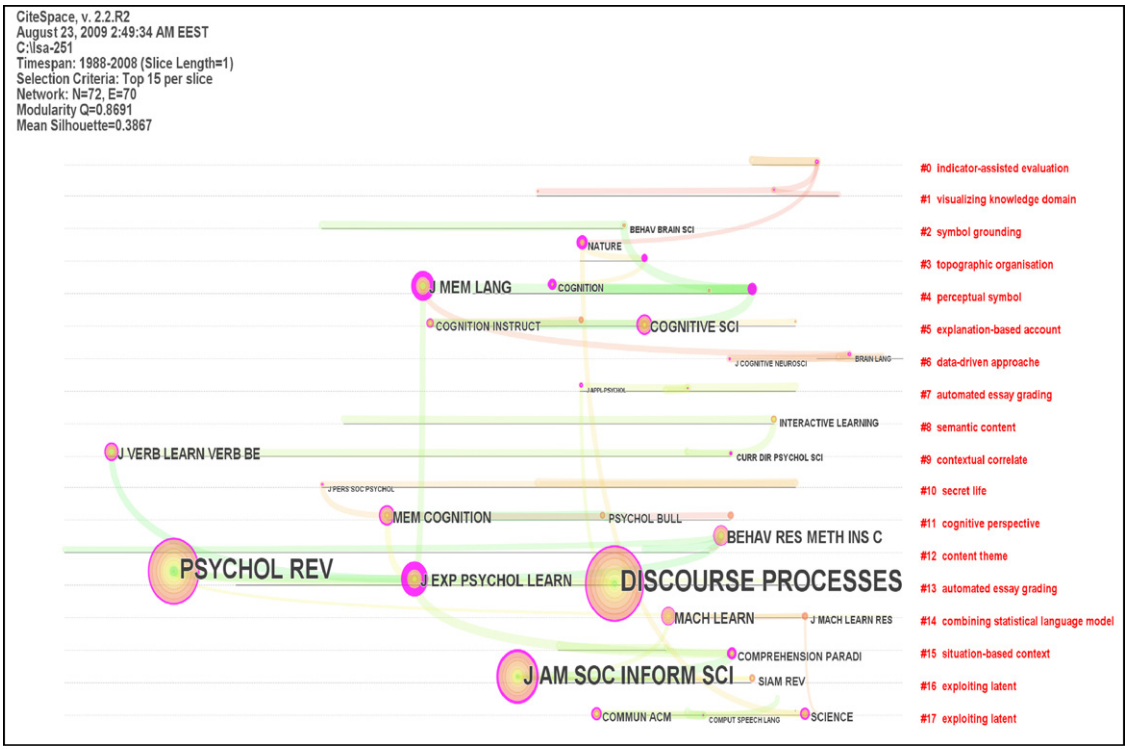

Fig. 6. Number of the terms in co-cited papers with LSA in topic.

5. Conclusion

LSA as a research tool has been used by many scientists of various scientific fields in the last decade. We used social network analysis (cluster methods and centrality measures), co-occurrence analysis on authors and journals, bibliometric methods (Lotka's Law) on selected bibliographic data on LSA downloaded from WoS. We identified the most prolific LSA researchers. While the networks of journals cited in papers with LSA in their titles and topics are well connected, this is not the case for the network of LSA researchers. We used CiteSpace to investigate the co-cited themes in network graphs. Themes were consistent with those of LSA researchers. Pivot points showed the turning points from one research field to the other. The LSA method was the main theme in all clusters. The betweenness centrality measures calculated indicate that LSA is a multidisciplinary method. Data fits Lotka's Law in that the majority of LSA researchers published just one paper while a relatively few prolific authors published several. It appears that the diffusion of LSA is still ongoing and LSA has become an indispensable part of the scientific discourse.

Acknowledgement

Helpful comments of anonymous reviewers are gratefully acknowledged.

References

Angluin, D., & Smith, C. H. (1983). Inductive inference: Theory and methods. *Computing Surveys*, *15*(3), 237–269.

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the USA*, *101*(Suppl. (1)), 5303–5310. From http://www.pnas.org/content/101/suppl.1/5303.full.pdf Retrieved 11.09.09

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*, 359–377.

Chomsky, N. (1991). Linguistics and cognitive science: Problems and mysteries. In A. Kasher (Ed.), *The Chomskyan turn* (pp. 26–53). Cambridge, MA: Basil Blackwell.

Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Beck, L. (1988). Improving information retrieval with latent semantic indexing. In *Proceedings of the 51st annual meeting of the American society for information science, Vol. 25* (pp. 36–40).

Deerwester, S. T., Furnas, G. W., Harshman, R. A., Landauer, T. K., Lochbaum, K. E., & Streeter, L. A. (1989). Computer information retrieval using latent semantic structure. United States Patent http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PTXT&s1=4839853.PN.&OS=PN/4839853&RS=PN/4839853 - h0#h0http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PTXT&s1=4839853.PN.&OS=PN/4839853&RS=PN/4839853 - h2#h24,839,853 (June 13, 1989). USPTO Patent Full-text and Image Database. Retrieved 15.04.09, from http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PTXT&s1=4839853.PN.&OS=PN/4839853&RS=PN/4839853.

Deerwester, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*, 321–407.
Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Amsterdam: Elsevier.
Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41. From http://moreno.ss.uci.edu/23.pdf Retrieved 11.08.09
Garfield, E. (2004). Historiographic mapping of the knowledge domain literature. *Journal of Information Science*, *30*, 119–145.
Glänzel, W. (2002). Coauthorship patterns and trends in the science (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends*, *50*(3), 461–473.
Hertzel, D. H. (1987). Bibliometrics, history of the development of ideas. *Encyclopedia of library and information science* (pp. 144–211). New York: Marcel Dekker.
Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of Collaboration networks in Scientometrics. *Scientometrics*, *75*(2), 189–202.
Jackendoff, R. S. (1992). *Languages of the mind*. Cambridge, MA: MIT Press.
Kitajima, M., Kariya, N., Takagi, H., & Zhang, Y. (2005). Evaluation of website usability using Markov chains and latent semantic analysis. *IEICE Transactions on Communications*, *E88B*(4), 1467–1475.
Landauer, T. K., & Dumais, S. K. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. From http://lsa.colorado.edu/papers/plato/plato.annote.html Retrieved 11.08.09
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.
Leydesdorff, L. (2007). Betweenness centrality" as an indicator of the "interdisciplinarity" of scientific journals. *Journal of the American Society for Information Science and Technology*, *58*, 1303–1319.
Lochbaum, K. E., & Streeter, L. A. (1989). Comparing and combining the effectiveness of latent semantic indexing and the ordinary vector-space model for information retrieval. *Information Processing & Management*, *25*, 665–676.
Michalski, R. S. (1983). A theory and methodology of inductive learning. *Artificial Intelligence*, *20*, 111–161.
Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the USA*, *98*(2), 404–409. From http://nicomedia.math.upatras.gr/courses/mnets/mat/Newman_StructureOfScientificCollaborationNets.pdf Retrieved 12.09.09
Newman, M. E. J. (2004). Who is the best connected scientist? A study of scientific coauthorship networks. In E. Ben-Naim, H. Frauenfelder, & Z. Toroczkai (Eds.), *Complex networks* (pp. 337–370). Berlin: Springer. From http://www-personal.umich.edu/~mejn/papers/cnlspre.pdf Retrieved 11.08.09
Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of information Science*, *28*, 443–455.
Pinker, S. (1990). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
Price, D. J. de Solla. (1970). Citation measures of hard science, soft science, technology and nonscience. In C. E. Nelson, & D. Pollock (Eds.), *Communication among scientists and engineers* (pp. 3–22). Lexington, MA: D.C. Heath & Co.
Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, *25*, 348–349.
Rousseau, B., & Rousseau, R. (2000). LOTKA: a program to fit a power law distribution to observed frequency data. *Cybermetrics*, *4*(1). From http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html Retrieved 11.08.09
Scott, J. (2000). *Social network analysis: A handbook* (second ed.). Thousand Oaks, CA: Sage Publications.
Scott, J., Tallia, A., Crosson, J. C., Orzano, A. J., Stroebel, C., DiCicco-Bloom, B., et al. (2005). Social network analysis as an analytic tool for interaction patterns in primary care practices. *Annals of Family Medicine*, *3*, 443–448. From http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1466914 Retrieved 11.08.09
Sentinel Visualizer. (2009). Social network analysis (SNA). From http://www.fmsasg.com/SocialNetworkAnalysis/ Retrieved 11.08.09.
Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*(September (4820)), 1317–1323.
Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, *68*(3), 595–610.
Vygotsky, L. S. (1968). *Thought and language (1934) (A. Kozulin, Trans.)*. Cambridge, MA: MIT Press.
White, H. D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure? Longitudinal evidence from the "Globenet" interdisciplinary research group. *Journal of the American Society for Information Science and Technology*, *55*, 111–126.

Please cite this article in press as: Tonta, Y., & Darvish, H.R. Diffusion of latent semantic analysis as a research tool: A social network analysis approach. *Journal of Informetrics* (2009), doi:10.1016/j.joi.2009.11.003