

Comparing citation patterns in entrepreneurship research articles in subject handbooks and Web of Science journals

Fredrik Åström¹

¹*fredrik.astrom@lub.lu.se*

Lund University Libraries, Head Office, P.O. Box 134, 221 00 Lund (Sweden)

Abstract

Not knowing enough about the similarities or differences of citation structures between different types of publications creates problems related to whether citation maps of research fields based on Web of Science/ISI data are representative of research fields as a whole or if they are a representation of how WoS perceives the field, not the least in the humanities and the social sciences. To investigate this problem, the citation structures in entrepreneurship research were analyzed using citation data both coming out of WoS-indexed journal articles and citation data from 12 entrepreneurship research handbooks. The datasets were analyzed by studying the age of references, the co-citation structures and also, citation overlaps. The results show substantial similarities between the two data sets: the distribution of the age of the references is almost identical, the co-citation structures in form of co-citation maps have strong similarities both in terms of identifiable networks as well as which cited authors are grouped together and there is also a strong citation overlap between the two different publication types.

Introduction

Traditionally, visualizations of research areas based different variations of co-citation analyses have been using data from the ISI/Web of Science (WoS) databases, and since the 2000s, also the Scopus database. The consequence of this is co-citation analyses to a large extent building on references coming out of research published in journals, whereas the access to information on research literature published in other forms of publications such as monographs and anthologies has been limited to analyses on WoS non-source items (e.g. Butler & Visser, 2006; Nederhof, van Leeuwen & van Raan, 2010). This means that we are still limited to analyzing references coming out of journal articles. At the same time, we also know that the coverage of the social sciences in WoS is far from complete; and that there are large differences in publication and citation behaviour in different fields of research (e.g. Hicks, 2004; Moed, 2005; van Leeuwen, 2006); and that representations of research fields can vary substantially depending on both selection of material and choice of e.g. level of analysis (Åström, 2002; Åström, 2010; Hellqvist, 2010, Moya-Anegón, Herrero-Solana & Jiménez-Contreras, 2006; Zhao, 2003). When creating maps of research fields in e.g. the social sciences, this creates a problem of lack of control over to what extent the maps based on WoS data are representative of the field we are analyzing as a whole, or if the maps primarily represents subsets of the field publishing research in WoS-indexed journals. Not the least is this question important in terms of the social sciences and the humanities, where a large extent of the research communication is published in other forms of media, such as monographs and anthologies.

The aim of this paper is to investigate the citation structures of one field but in two different kinds of publications, approaching the question of to what extent the structures found in co-citation mapping of a field based on WoS data represents the field or just the WoS representation of the field? Within the framework of the Explore project, a database of references from each individual chapter of 12 entrepreneurship research handbooks was constructed (Landström, Hairichi & Åström, 2010), presenting us with an opportunity to compare the citation structures identifiable in these handbooks with analyses made on WoS data.

Method and material

In entrepreneurship research, a series of handbooks have been published since the early 1980s, presenting the state of the art of the field in handbook chapters authored and edited by the leading scholars in the field. In this series, 12 individual books has been published over the years 1982-2006; and from these handbooks, a database consisting of all references from each individual chapter was constructed. In total, there are 185 chapters containing references; and in these, the total amount of references is 12,781 (Table 1).

Table 1. Entrepreneurship research handbooks

<i>Editor</i>	<i>Title</i>	<i>Year</i>	<i>Chapters</i>	<i>References</i>
Kent et al	Encyclopedia of Entrepreneurship	1982	18	630
Sexton & Smilor	The Art and Science of Entrepreneurship	1986	11	381
Sexton & Kasarda	The State of the Art of Entrepreneurship	1992	22	1,547
Katz & Brockhaus	Advances in Entrepreneurship, Firm Emergence and Growth, Vol. 1	1993	5	335
Katz & Brockhaus	Advances in Entrepreneurship, Firm Emergence and Growth, Vol. 2	1995	8	657
Katz & Brockhaus	Advances in Entrepreneurship, Firm Emergence and Growth, Vol. 3	1997	7	852
Sexton & Smilor	Entrepreneurship 2000	1997	18	907
Sexton & Landström	The Blackwell Handbook of Entrepreneurship	2000	22	1,427
Acs & Audretsch	Handbook of Entrepreneurship Research	2003	19	1,687
Alvarez et al	Handbook of Entrepreneurship Research: Disciplinary Perspectives	2005	11	652
Casson et al	Oxford Handbook of Entrepreneurship	2006	27	2,079
Parker	The Life Cycle of Entrepreneurial Ventures	2006	17	1,627
			185	12,781

To find entrepreneurship research published in WoS-indexed journals, a topic search in Social Science Citation Index was done using the following search string: “entrepreneur* OR small business* OR small firm* OR emerging business* OR emerging firm* OR new venture*” OR emerging venture* OR founder OR founders” (Landström & Åström, in press), restricted to articles published in the same years as the handbooks, resulting in 5,122 documents that was downloaded for analysis. From the WoS data, the information on cited author from the ‘cited reference’ field was extracted.

Thus, we have two sets of data with references from entrepreneurship research articles and handbook chapters published 1982-2006 (Table 2), to be analyzed using the Bibexcel software (Persson, Danell & Schneider, 2009).

Table 2. Dataset properties

	<i>Handbook</i>	<i>WoS</i>
Number of documents	185	5,122
Number of references	12,781	162,530
Number of cited first authors	4,725	66,777

Since the author information in the ‘cited reference’ field from WoS is limited to the first author, the same approach was used on the references from the handbooks. Three different analyses were performed to compare the citation structures of entrepreneurship research in handbooks and WoS journal articles respectively. To provide a general overview of the structures, the age of the references were analyzed, followed by an author co-citation mapping (White & Griffith, 1981) on both datasets. To get a more in-depth view, further analyses were also done in terms of studying the ranking of cited authors and citation overlaps through the share of common references in-between the data sets.

Results

Age of references

The first step was to analyze the age of the cited references in the two datasets by looking at the distribution of references per decade (Figure 1). Not surprisingly, the majority of the references are from the 1980s and onwards, the time when entrepreneurship research started establishing itself as a field of research.

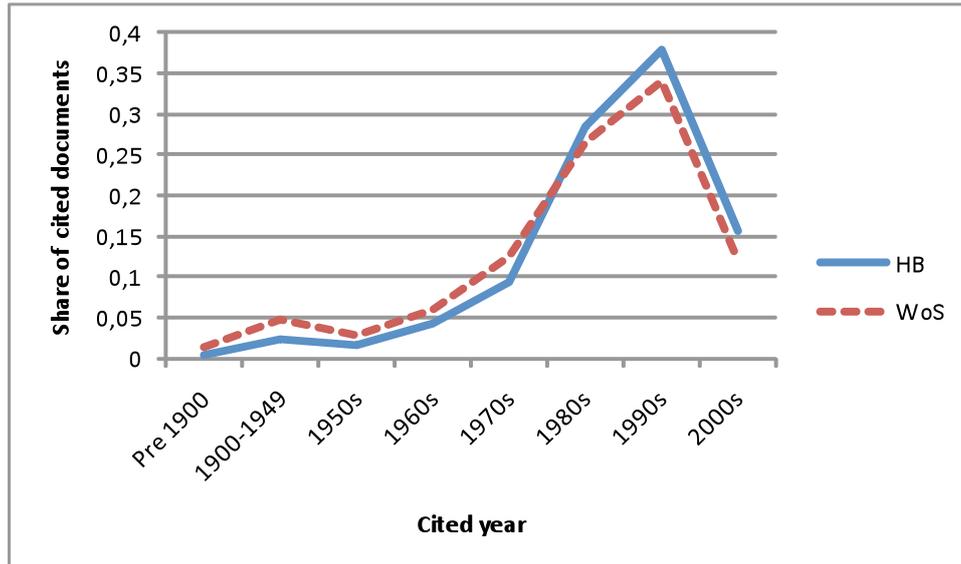


Figure 1. Age of references in entrepreneurship articles in handbooks and Web of Science indexed journals respectively

The variations in terms of age of references between the handbooks and the WoS journal articles are small, with differences ranging between 0.8 and 4%. This shows only structures on a very general level, but can be seen as one indicator on strong similarities between the different forms of publications. One thing that can be noted though, is how the share of older references (from the 1970s and backwards) is slightly higher in the WoS articles, whereas the

share of newer references – from the time entrepreneurship research started establishing itself as a field of research – is slightly higher in the handbooks. One reason for this could be differences between the document sets. Whereas the WoS articles are more heterogeneous, and therefore being more likely to have a wider range of references, the handbook chapters are more directly focused on entrepreneurship research and – not the least – the development of the field itself, therefore having a stronger focus on the literature within the field.

Co-citation analyses

To identify and compare more fine grained structures, a first author co-citation analysis was performed on both datasets. Using Bibexcel (Persson, Danell & Schneider, 2009), the 123 most cited authors from the handbook reference lists and the 118 most cited authors in the WoS journal articles were selected. In both cases, the co-citation frequencies were calculated; and in addition to these, the ‘select strongest links’ option in Bibexcel, where each node is clustered by its strongest link, was used to produce a net-file for Pajek. Furthermore, a clustering routine suggested by Persson (1994), clustering pairs with one unit in common, was used to create partitions, before making the visualization in Pajek (de Nooy, Mrvar & Batagelj, 2005) using the Kamada-Kawai (1989) algorithm and the ‘separate components’ option (Figure 2a & b).

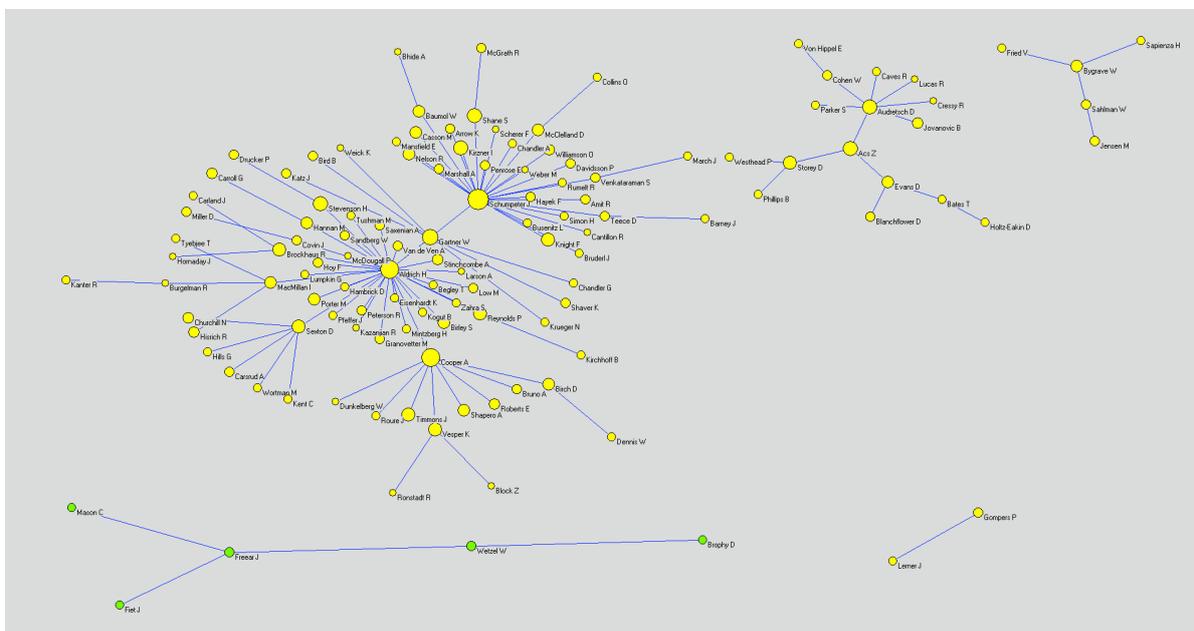


Figure 2a. First author co-citation analysis of entrepreneurship research articles in handbooks: 123 authors with 10 citations or more

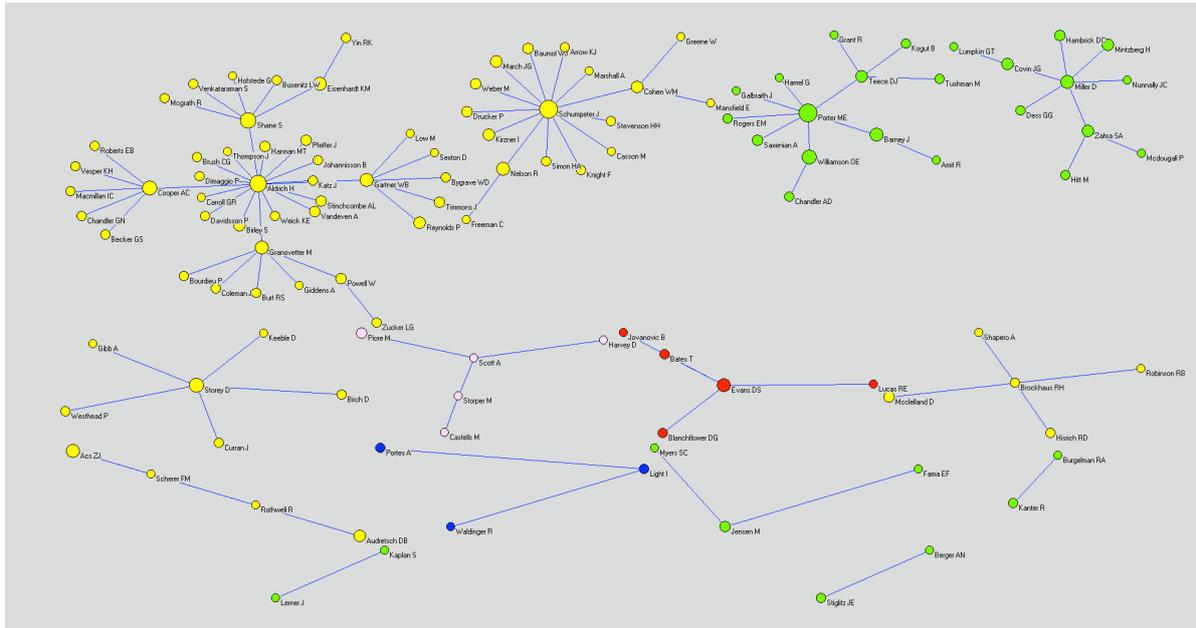


Figure 2b. First author co-citation analysis of entrepreneurship research articles in Web of Science journals: 117 authors with 80 citations or more

Although there are some differences both in terms of graphic display and the Persson (1994) style clustering, the general traits are quite similar: in the handbook map we find one large network with Aldritch and Schumpeter as central nodes in the two major sub-networks, a structure which is also reflected in the two networks on the upper left side of the WoS map. And within these networks, we also find smaller sub-networks around Gartner and Cooper present in both maps. The second largest main network in the handbook map has in the WoS map formed three separate networks: one with Storey, one with Evans and one with Acs and Audretsch, whereas many of the smaller networks in the WoS map contains authors being part of the main network in the handbook map.

The differences in both graphic display – with a larger number of distinct networks in the WoS map – and results of the clustering – where the handbook analysis resulted in two clusters and the WoS analysis in five – can probably be related to the larger citation frequencies in the WoS material. Although the number of authors, as well as the amount of unique co-cited pairs, formed by these authors is relatively similar, the number of co-citation links is significantly higher in the Web of Science analysis (Table 3).

Table 3. Discrete number of co-cited pairs and total number of co-citation links in entrepreneurship handbooks and Web of Science journals respectively

	<i>Co-cited pairs</i>	<i>Co-citation links</i>	<i>Average number of links/ co-cited pair</i>
Handbooks	5,848	19,398	3.32
WoS journal articles	6,528	69,708	10.59

Out of the authors selected for the co-citation analysis, the 20 most cited authors in the handbooks and WoS journal articles respectively have 11 author names in common, whereas seven authors occur in one list only, however ranking high outside the top 20. And in comparison to the ranked list for the selection of ca 120 authors for the co-citation analyses,

only one author per data set is ranked substantially different in the two lists and only one author is only present in one list (Table 4).

Table 4. Most cited authors in entrepreneurship research articles in handbooks and Web of Science journals respectively

<i>Rank</i>	<i>Freq.</i>	<i>Handbook</i>	<i>Freq.</i>	<i>WoS</i>
1	70	Schumpeter J	416	Schumpeter J
2	59	Aldrich H	393	Porter ME
3	54	Cooper A	361	Aldrich H
4	43	Gartner W	301	Shane S
5	39	Stevenson H	271	Cooper AC
6	38	Kirzner I	255	Storey D
7	37	Acs Z	242	Williamson OE
8	35	Audretsch D	234	Miller D
9	34	Shane S	230	Gartner WB
10	33	Sexton D	229	Eisenhardt KM
11	32	Brockhaus R	221	Granovetter M
12	32	Vesper K	218	Acs ZJ
13	32	Reynolds P	213	Nelson R
14	29	Timmons J	210	Barney J
15	29	Knight F	198	Evans DS
16	29	Storey D	190	Teece DJ
17	27	Birch D	180	Audretsch DB
18	27	McClelland D	177	Zahra SA
19	27	Evans D	176	Reynolds P
20	26	Hannan M	173	Kirzner I

Citation overlap

To go into an even more detailed comparison, the citation overlap was analyzed on about 1,400 of most cited authors, looking both into the overlap of discrete author names as well as the citation frequencies related to these author names by calculating the remaining percentage of citations/authors after removing non-overlapping authors. (Table 5).

Table 5. Citation overlap*

	<i>Handbook: citation count</i>	<i>Handbook: cited authors</i>	<i>WoS: citation count</i>	<i>WoS: Handbook: cited authors</i>
Common citations	74% (N=6,210)	53% (N=1,348)	53% (N=50,247)	49% (N=1,453)

* Calculated on the 1,400 most cited authors, giving a citation threshold of 2 citations for handbook references and 10 for Web of Science references.

Of the 6,210 citations to the 1,348 most cited authors in the handbooks, almost 75% of the citations remain after removing citations not overlapping with the WoS set. The general tendency though is that about 50% of both authors and citations remain after the non-overlapping authors have been removed.

Discussion

Having investigated the citation structures in entrepreneurship research as reflected in, on one hand, state of the art presentations of research in the field in a set of handbooks, on the other, in WoS-indexed journal articles, we see substantial similarities. These similarities occur when analysing general structures such as the age of references, when mapping the field using co-citation analysis and when analyzing more fine-grained structures in terms of citation overlaps and comparing highly cited authors in the field.

In terms of the differences that do occur, the tendency is that variations are primarily found among authors with low citation scores, or in ‘the long tail’, whereas among the higher ranked authors, the differences between handbook and WoS citations is very small. This is not entirely surprising: we would expect to find more stability among the higher ranked cited authors where we find many ‘citation classics’ being both predecessors to, and pioneers in, entrepreneurship research, as well as the ones proposing important theories and doing essential empirical investigations. In ‘the long tail’ however, we can expect to find more variations due to citations more related to particular studies in individual papers. This could explain the high number of overlaps when looking at the citation overlap in terms of citation frequencies in the handbooks. Since the handbooks are overviews of research within the field, we would expect less citations being related to particular studies or local investigations, thus making ‘the long tail’ even less significant in terms of citation ranking and its share of the total amount of citations.

To return to the question stated in the introduction to the paper: are maps of research fields based on WoS data representative of the field as a whole or is it only reflecting the WoS version of the field? In the case of entrepreneurship research, the ‘WoS version’ of entrepreneurship research seems to be valid in comparison with how the field is reflected in the citation structures of texts by the leading scholars of the field. Related to the question of WoS based representations of research fields in co-citation analyses is the issue of using WoS data for assessing research performance. Given the similarities found in the analyses presented here, a WoS based analysis of research performance would be likely to produce results that are representative for, although not fully covering, the field. However, it should be considered that the similarities are at the strongest among the most highly cited authors, to a large extent the citation classics like pioneers and predecessors in the field, whereas the differences increase among the less cited authors, where we find much of the cited authors being active today, i.e. the ones that would be of interest to analyze when doing a research performance assessment. It should also be kept in mind, that these results refer specifically to entrepreneurship research; and also, that we are still dealing with two specific genres of publication. Similar investigations needs to be done on other fields in the social sciences and humanities; and also, on an even larger range of publication types.

Acknowledgments

The author wishes to express his gratitude to the anonymous reviewers, who with their comments helped improving the paper.

References

- Åström, F. (2002). Visualizing library and information science concept spaces through keyword and citation based maps and clusters. In H. Bruce at al (Eds.), *Emerging frameworks and methods: CoLIS4* (pp. 185–197). Westport, CT: Libraries Unlimited.
- Åström, F. (2010). The visibility of information science and library science research in bibliometric mapping of the LIS field. *Library Quarterly*, 80(2), 143-159.
- Butler, L. & Visser, M.S. (2006). Extending citation analysis to non-source items. *Scientometrics*, 66(2), 327-343.

- de Nooy, W., Mrvar, A. & Batagelj, V. (2005). Exploratory social network analysis with Pajek. New York: Cambridge University Press.
- Hellqvist, B. (2010). Referencing in the humanities and its implications for citation analysis. *Journal of the American Society for Information Science and Technology*, 61(2), 1-9.
- Hicks, D. (2004). The four literatures of the social science. In H. K. Moed, W. Glanzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 473–495). Dordrecht: Springer.
- Kamada, T. and S. Kawai (1989), An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(12), 7-15.
- Landström, H. & Åström, F. (in press). Who's asking the right questions? Patterns and diversity in the literature of new venture creation. In K. Hindle & K. Klyver (Eds.), *Handbook in New Venture Creation Research*. Cheltenham: Edward Elgar.
- Landström, H., Hairichi, G. & Åström, F. (2010). *Entrepreneurship: Exploring the knowledge base*. Unpublished DIME report. Lund: Lund University, CIRCLE.
- Moed, H. K. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.
- Moya-Anegón, F., Herrero-Solana, V. & Jiménez-Contreras, E. (2006). A connectionist and multivariate approach to science maps: The SOM, clustering and MDS applied to library and information science research. *Journal of Information Science*, 32(1), 63-77.
- Nederhof, A.J., van Leeuwen, T.N. & van Raan, A.F.J. (2010). Highly cited non-journal publications in political science, economics and psychology: A first exploration. *Scientometrics*, 83, 363-374.
- Persson, O., Danell, R. & Schneider, J.W. (2009). How to use Bibexcel for various types of bibliometric analysis. In F. Åström et al, *Celebrating Scholarly Communication Studies A Festschrift for Olle Persson at his 60th Birthday*. International Society for Scientometrics and Informetrics.
- Van Leeuwen, T. (2006). The application of bibliometric analyses in the evaluation of social science research: Who benefits from it, and why it is still feasible. *Scientometrics*, 66(1), 133-154.
- White, H. D. & Griffith B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32, 163-172.
- Zhao, D. (2003). A comparative citation analysis study of web-based and print-journal based scholarly communication in the XML research field. Unpublished doctoral dissertation, The Florida State University, Tallahassee. Retrieved January 10, 2011 from: <http://etd.lib.fsu.edu/theses/available/etd-09232003-012028/>