

Computational and visual analysis of the development stage of theories in the social sciences: a case in the entrepreneurship field

Ge Qian

The aim of this article is to develop a method combining calculation, visualization and intuitive analysis, which will help social scientists to study the history of a theory in a particular discipline. Considering that the typical type of knowledge-domain map is too complicated and cumbersome for social scientists who possess little knowledge about scientometrics, a new and simplified 3D knowledge-mapping method is designed and illustrated using examples from the entrepreneurship field. The major difference between the orthodox knowledge mapping method and the method we introduce is that co-citation network is not marked, but only co-citation analysis results. Using an alternative method, our map shows the three prominent researchers in the entrepreneurship field as well as three stages: the first from approximately 1920 to 1960, the second from 1960 to 2000, and the third beginning in 2000.

Keywords: Co-citation, entrepreneurship field, knowledge domains, mapping method.

UNLIKE natural scientists, social scientists and scholars usually give special attention to the study of the history of a theory in their discipline, including the examination of how the development stages of a theory can be divided, or how to identify the most prominent researchers in a specific research field. Traditionally, these kinds of discussions are mostly theoretical and qualitative analyses; however, in this article, we try to develop a method that includes calculation, visualization and intuitive analysis – as well as a new type of knowledge mapping method as the carrier – to analyse the above methodological problems in the social sciences, such as how to divide the development stages of a theory or how to identify the most prominent scholars in the entrepreneurship field.

Data and mean

The term ‘mapping knowledge domains’ is used to describe a newly evolving interdisciplinary area of science that aims to chart, mine, analyse, sort, display and enable the navigation of knowledge. This field is aimed at easing access to information, making evident the structure of knowledge, and allowing knowledge seekers to succeed in their endeavours¹. The value of mapping knowledge domains as a scientometric method is beyond doubt. It is an effective knowledge management tool, both for scien-

tific and technological research and technology innovation enterprises.

Like other scientometric methods that aim to study the structure of scientific research, mapping knowledge domains requires an objective basis; it is mainly based on citation and co-citation analysis. Co-citation analysis allows us to identify the groups of scientists and publications from which conclusions can be drawn about the inner structure of research disciplines, schools, or paradigms². There are many co-citation networks and one of them, the document co-citation network, is generated in this study. Document co-citation analysis measures the number of documents that cite any given pair of documents³. It is assumed that the authors whose works are related are repeatedly cited together. Thus, they tend to group together when analysed, whereas authors who are rarely or never cited together do not. Co-citation analysis is dynamic over time, because co-citation frequencies for particular works change as the focus of research efforts in a field changes. For example, if papers *A* and *B* are both cited by paper *C*, they may be said to be related to one another, even though they do not directly cite one another. If papers *A* and *B* are both cited by a number of other papers, they have a stronger relationship, which increases as they are cited by more papers. Overall, co-citation analysis is better than citation analysis or merely looking for the document or author statistics in the ISI *Web of Science (WoS)* database of the Institute for Scientific Information (ISI) in the United States, because co-citation analysis is a more in-depth approach that explores the relationships among citation networks.

Ge Qian is in the Theoretical Teaching and Research Department, Humanities College, Shanghai University of Finance and Economics, Shanghai 200433, China. e-mail: qiange@mail.shufe.edu.cn

The data used in this article come from the *WoS* database. The data retrieval strategy used here is

TS = (Entrepreneurship)

Refined by: Document type = (ARTICLE)

Timespan = All years. Databases = SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH.

Lemmatization = On

The full bibliographic records for 5117 articles – including authors, titles, abstracts and reference lists – were retrieved and downloaded on 8 March 2012. CiteSpace II, developed by Chaomi Chen from Drexel University (<http://cluster.cis.drexel.edu/~cchen/citespace/>) was used to produce the co-citation networks.

Table 1 lists the top 25 most cited documents in the entrepreneurship field, as revealed through the document co-citation network analysis.

The main scientometric and bibliometric indicators used in this article are frequency, burst, centrality, sigma and half-life. Citation frequency is the most used scientometric indicator; it is treated as the indicator of the influence of a document. Moreover, citation frequency is the basis of scientometric analysis. So only the top 25 most cited and most used documents are selected for further analysis. The burst detection algorithm was proposed by Kleinberg⁴ to identify research front concepts, and it is widely treated as the indicator of emergence. Although Freeman's⁵ betweenness centrality metric is used to highlight potential pivotal points, it is also used as an important indicator of a document, following Leydesdorff's⁶ use of the metric as an indicator for measuring the importance of academic journals. The Σ (sigma) index, when used as a combination of the two above indices – centrality and burstness – seems to be the creative indicator described by Chen⁷. The scientific literature can easily be characterized by two distinct types of article-citation half-lives: classical articles that have persistently high citations, and transient articles that have citations which peak over a short period of time⁸.

Design

Figure 1 shows our newly designed 3D knowledge-domain map for the entrepreneurship field, with all the abbreviations from the bubbles described in Table 1.

In this new type of 3D knowledge-domain map, the *X* axis represents the year, while the *Y* and *Z* axes represent citation frequency – the oldest index in scientometrics – and sigma – the newest index for measuring the creativeness of scientific documents – respectively. The remaining three indices also present this information in an intuitive manner. The redder the bubbles, the higher the values of burst, while the bubbles corresponding to documents with no burst value are pure green (Figure 1).

Since Freeman's betweenness centrality metric is used to measure the practical importance, rather than the surface influence, of a document, we use the size of the bubbles to represent the importance of the representative literature; in other words, the larger the bubble, the higher the centrality value. The half-life value is displayed as a blue line, with an arrow that extends from its corresponding bubble. The major difference between the orthodox method for knowledge mapping and our method, is that only the results of co-citation analysis are mapped on the latter map type, while the co-citation network is omitted.

Findings

Our map type shows the three most prominent scholars in the entrepreneurship field. Regarding their chronological order, Schumpeter, J. A., abbreviated as 'SC' in Figure 1, is the first. He has two publications listed in the highly cited documents table: *The Theory of Economic Development. An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*⁹, and *Socialism, Capitalism and Democracy*¹⁰. The former ranks second in citation frequency in the entrepreneurship field and is tied at the first place with the centrality value. The second is McClelland, D. C., abbreviated as 'CL'. His book, *The Achievement Society*¹¹, is not only the most emergent and creative publication, it is also tied at the first place with the most important document because it has the highest values in burst, sigma and centrality. The last researcher is Shane, S., abbreviated as 'SH', who has two highly cited articles: 'The promise of entrepreneurship as a field of research', and 'Prior knowledge and the discovery of entrepreneurial opportunities'; the former is the most-cited document in the entrepreneurship field and related fields.

Our map also presents the three stages of development in entrepreneurship research. The first stage is from about the 1920 to 1960, with Schumpeter, J. A., who represents the founding fathers of entrepreneurship research. According to Schumpeter^{9,10}, an entrepreneur is willing and able to convert a new idea or invention into a successful innovation. Entrepreneurship employs what Schumpeter called 'creative destruction' to replace, in whole or in part, inferior offerings across markets and industries, simultaneously creating new products and new business models. Thus, creative destruction is largely responsible both for the dynamism of an industry and for long-term economic growth. The idea that entrepreneurship leads to economic growth is an interpretation of the residual in endogenous growth theory. Another famous scholar in this phase is Knight¹², abbreviated as 'KN', who thought that entrepreneurship is about taking risks. The feature of this stage is that, although the number of published works listed in the highly cited table is small, they are almost all considered classics, because they all have long half-lives from the scientometric perspective. We may call this

Table 1. The top 25 most cited documents in the entrepreneurship field

Abbreviations	Documents	Frequency	Burst	Centrality	Sigma	Half-life
SH	Shane, S. and Venkataraman, S., The promise of entrepreneurship as a field of research. <i>Acad. Manage. Rev.</i> , 2000, 25 (1), 217–226.	417		0	1	2
SC	Schumpeter, J. A., <i>The Theory of Economic Development. An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle</i> , Transaction Publishers, New Brunswick, NJ, London, 1934.	402		0.12	1	46
BA	Barney, J., Firm resources and sustained competitive advantage. <i>J. Manage.</i> , 1991, 17 (1), 99–120.	263		0	1	9
LU	Lumpkin, G. T. and Dess, G. G., Clarifying the entrepreneurial orientation construct and linking it to performance. <i>Acad. Manage. Rev.</i> , 1996, 21 (1), 135–172.	253		0.01	1	3
KI	Kirzner, I. M., <i>Competition and Entrepreneurship</i> , University of Chicago Press, Chicago, IL, 1973.	239		0.03	1	6
EV	Evans, D. S. and Leighton, L. S., Some empirical aspects of entrepreneurship. <i>Am. Econ. Rev.</i> , 1989, 79 (3), 519–535.	208		0	1	7
NE	Nelson, R. R., <i>An Evolutionary Theory of Economic Change</i> , Harvard University Press, Cambridge, MA, 1982.	208		0	1	12
SH	Shane, S., Prior knowledge and the discovery of entrepreneurial opportunities. <i>Organ. Sci.</i> , 2000, 11 (4), 448–469.	191	4.37	0.01	1.04	3
PE	Penrose, E. T., <i>The Theory of the Growth of the Firm</i> , Shape, New York, 1959.	185		0	1	40
MI	Miller, D., The correlates of entrepreneurship in three types of firms. <i>Manage. Sci.</i> , 1983, 29 (7), 770–791.	183		0.01	1	7
KN	Knight, F. H., <i>Risk, Uncertainty and Profit</i> , Hart, Schaffner and Marx., New York, 1921.	183		0.01	1	58
CH	Cohen, W. M. and Levinthal, D. A., Absorptive capacity: a new perspective on learning and innovation. <i>Admin. Sci. Q.</i> , 1990, 35 (1), 128–152.	179		0	1	11
MA	DiMaggio, P. J. and Powell, W. W., The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. <i>Am. Sociol. Rev.</i> , 1983, 48 (2), 147–160.	178	10.7	0.01	1.11	21
EV	Evans, D. S. and Jovanovic, B., An estimated model of entrepreneurial choice under liquidity constraints. <i>J. Polit. Econ.</i> , 1989, 97 (4), 808–827.	172		0.01	1	6
CL	McClelland, D. C., <i>The Achievement Society</i> , Von Nostrand, Princeton, NJ, 1961.	171	14	0.12	4.75	1
EI	Eisenhardt, K. M., Building theories from case study research. <i>Acad. Manage. Rev.</i> , 1989, 14 (4), 532–550.	171		0	1	16
GR	Granovetter, M., Economic action and social structure: the problem of embeddedness. <i>Am. J. Sociol.</i> , 1985, 91 (3), 481–510.	171		0.01	1	9
VE	Venkataraman, S., The distinctive domain of entrepreneurship research. <i>Adv. Entrepreneurship, Firm Emerg. Growth</i> , 1997, 3 , 119–138.	157		0	1	5
CV	Covin, J. G. and Slevin, D. P., Strategic management of small firms in hostile and benign environments. <i>Strat. Manage. J.</i> , 1989, 10 (1), 75–87.	155	5.15	0	1.01	10
PO	Porter, M. E., <i>Competitive Strategies: Techniques for Analyzing Industries and Competitors</i> , The Free Press, New York, 1980.	152	5.59	0.01	1.05	9
BU	Baumol, W. J., Entrepreneurship: productive, unproductive, and destructive. <i>J. Polit. Econ.</i> , 1990, 98 (5), 893–919.	146		0	1	3
NO	North, D. C., <i>Institutions, Institutional Change and Economic Performance</i> , Cambridge University Press, Cambridge, 1990.	145	7.63	0	1	18
DR	Drucker, P. F., <i>Innovation and Entrepreneurship</i> , Harper and Row, New York, 1985.	141	11.4	0.01	1.09	2
SC	Schumpeter, J. A., <i>Socialism, Capitalism and Democracy</i> , Harper and Brothers, New York, 1942.	140		0.01	1	51
TE	Teece, D. J., Pisano, G. and Shuen, A., Dynamic capabilities and strategic management. <i>Strat. Manage. J.</i> , 1997, 18 (7), 509–533.	138		0	1	5

period the beginning stage of entrepreneurship research. The second stage stretches from about 1960 to 2000, and may be called the rapid development phase of entrepreneurship research. A large number of research results emerged during this stage, which can be divided into three cases: first, studies of entrepreneurship that use new methods and perspectives. These include the work of

McClelland¹¹, who suggested that the key to entrepreneurial behaviour lies in achievement motivation. The need to achieve is a drive to excel, to achieve a goal in relation to a set of standards. A person who is endowed with such a need is expected to spend time considering how to perform a job better, or how to accomplish something that is important to him. McClelland distinguished such persons

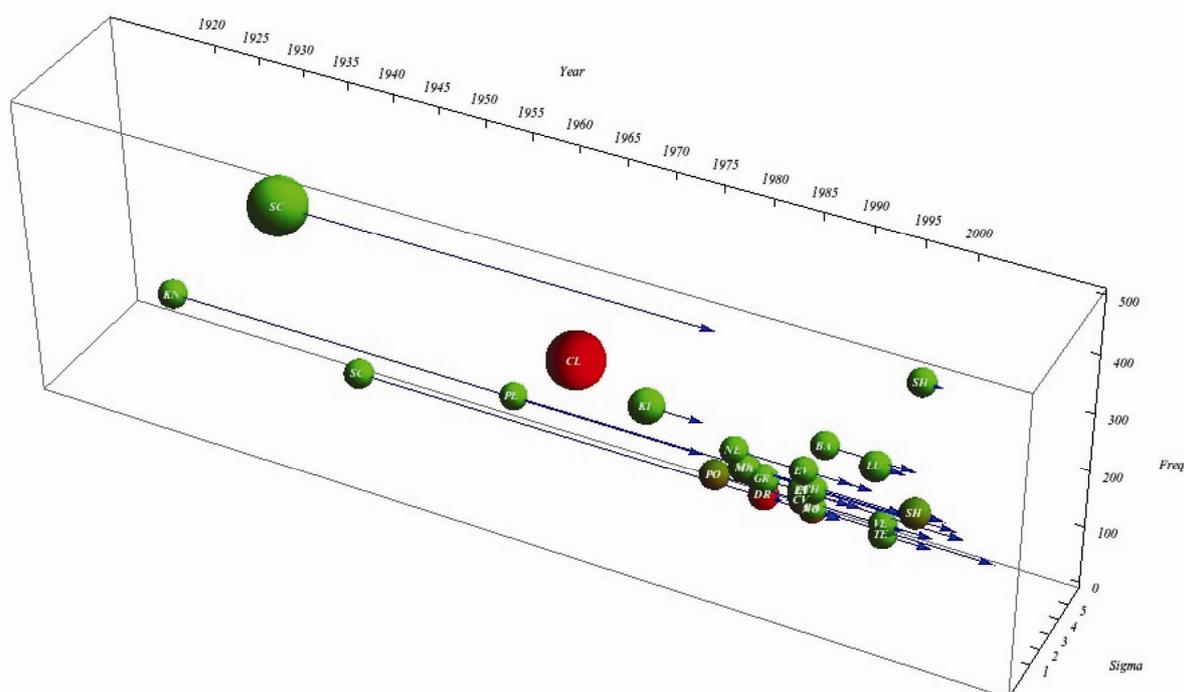


Figure 1. The newly designed 3D knowledge-domain map for the entrepreneurship field.

from the rest of the population, suggesting that members of the former group were high achievers, and were thus driven by that high need for achievement to become entrepreneurs. The second type of research was primarily inherited from the previous view. For example, Drucker¹³, abbreviated as 'DR', considers entrepreneurship as the behaviour of the entrepreneur, and reflects the kind of person who is willing to put his/her career and financial security on the line and take risks to achieve an idea, spending much of his/her time and capital on an uncertain venture. This view obviously draws directly from that of Knight¹². The third type of research stems from debates within classical theory, such as, for example, an alternate description of Schumpeter's theory, mentioned above, offered by Kirzner¹⁴, abbreviated as 'KI'. He suggests that the majority of innovations may actually be largely incremental improvements, such as the replacement of paper with plastic in the production of a drinking straw.

The most recent phase began in 2000, when the contemporary study of entrepreneurship was significantly defined by Shane and Venkataraman¹⁵, in their agenda-setting article 'The promise of entrepreneurship as a field of research'. According to them, entrepreneurship comprises two phenomena – 'enterprising individuals' and 'entrepreneurial opportunities' – and researchers should study the nature of those individuals who respond to these opportunities while others do not, as well as the opportunities themselves and the nexus between individual and opportunity. Since the main method of this article is co-citation analysis as an empirical study, we can only report the beginning of this new stage; we cannot

describe the characters in the third stage because of the time lag in the co-citation analysis.

Discussion

We can now compare our map type with the typical knowledge-domain map. Figure 2 presents one typical type of knowledge-domain map, created by drawing the sample data from the entrepreneurship field with the CiteSpace software. In order to facilitate comparisons, the CiteSpace map is set to its time-zone style.

The major difference between both the methods is that the co-citation network is not marked on our map, which only displays the results of the co-citation analysis. This simplification makes our type of knowledge-domain map clearer and more easily analysable, providing clear usability for social scientists, who may find the complex co-citation networks not only confusing but also useless.

CiteSpace is more successful as a calculation tool than as a 'knowledge domain visualization' tool¹⁶; our map is also based on the calculation results of CiteSpace. However, as a compiled Java program, CiteSpace requires its users to have a high level of skills, but does not allow them to be more imaginative and creative. While many social scientists may not be skillful, they often are more imaginative and creative. I therefore suggest, that knowledge domain visualization software should be developed based on existing and widely used mathematical and statistical graphics software, such as MatLab or Wolfram Mathematica (the map type introduced in this article was

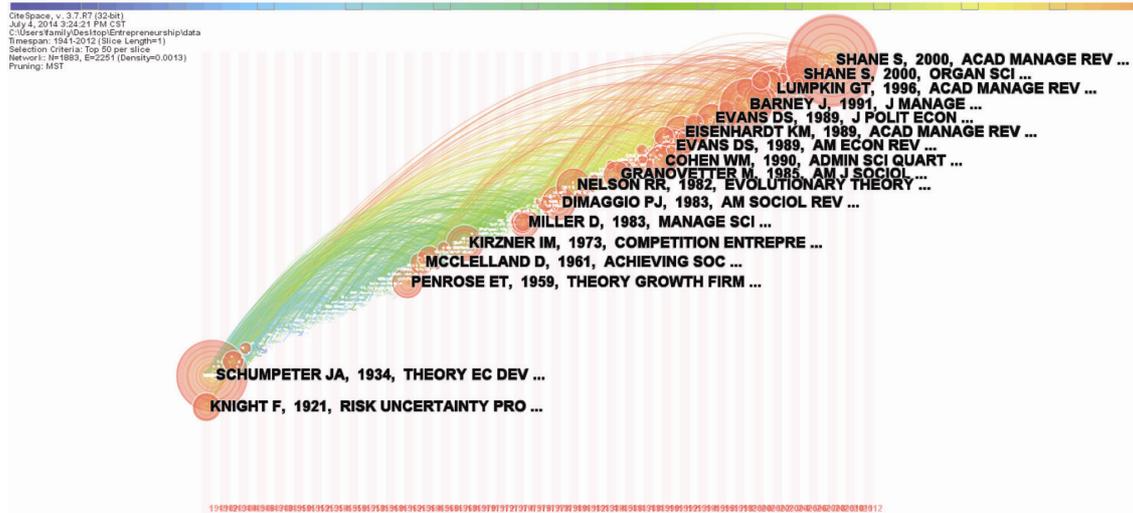


Figure 2. Typical knowledge-domain map drawn by CiteSpace of the entrepreneurship field in time zone style.

drawn using Wolfram Mathematica), instead of going through efforts of developing and establishing new and specific software. In doing so, we may avoid a larger gap between the researchers who draw knowledge-domain maps, usually information scientists, and the potential users of these maps, including social scientists.

In closing, I will give a clear indication of where this method provides insights that are not available from earlier methods. Our map is comprehensive and displays many indicators in the same map, while CiteSpace maps are usually capable of showing fewer indices. In the example presented here, our method helps demarcate the three most prominent researchers, and the corresponding three stages, in entrepreneurial research, while the CiteSpace map cannot provide the same result. McClelland¹¹ – whose work is represented by the large red sphere due to its high centrality, burst, and sigma values – is intuitively displayed as a VIP in the research on entrepreneurship, and as the iconic figure for a new stage of research in the map presented in Figure 1. However, in the CiteSpace map presented in Figure 2, McClelland's position is not outstanding, due to the lower citation frequency of his work.

To summarize, this article seeks to develop a new and simplified knowledge mapping method, which is then tested using data on entrepreneurship research. This exercise is useful and results in some degree of innovation. However, like all other research, this study also has some limitations. One obvious drawback of this work is that no direct evidence is given to justify the methodological choices. In fact, according to Chen⁷, the validity of such studies may be obtained through comments from experts in the field. However, perhaps because the number of famous experts in the entrepreneurship field is quite small, we have always failed to acquire enough comments to support a statistical analysis. Thus, we hope that the publication of this article can facilitate additional discussion on the validity of this method from experts in related fields.

In addition, if this method can be proved effective, it should also be applied to other disciplines in the social sciences, and even to some disciplines in the natural sciences.

1. Shiffrin, R. M. and Börner, K., Mapping knowledge domains. *Proc. Natl. Acad. Sci. USA (Suppl. 1)*, 2004, **101**, 5183–5185.
2. Small, H., Co-citation context analysis and the structure of paradigms. *J. Document.*, 1980, **36**(3), 183–196.
3. Small, H., Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci. Tec.*, 1973, **24**(4), 265–269.
4. Kleinberg, J., Bursty and hierarchical structure in streams. *Data Min. Knowledge Discov.*, 2003, **7**(4), 373–397.
5. Freeman, L. C., Centrality in social networks: conceptual clarification. *Soc. Networks*, 1979, **1**(3), 215–239.
6. Leydesdorff, L., Visualization of the citation impact environments of scientific journals: an online mapping exercise. *J. Am. Soc. Inf. Sci. Technol.*, 2007, **58**(1), 25–38.
7. Chen, C., CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci. Technol.*, 2006, **57**(3), 359–377.
8. Price, D., Networks of scientific papers. *Science*, 1965, **149**(3683), 510–515.
9. Schumpeter, J. A., *The Theory of Economic Development. An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*, Transaction Publishers, New Brunswick, NJ, 1934.
10. Schumpeter, J. A., *Socialism, Capitalism and Democracy*, Harper and Brothers, New York, 1942.
11. McClelland, D. C., *The Achievement Society*, Von Nostrand, Princeton, NJ, 1961.
12. Knight, F. H., *Risk, Uncertainty and Profit*, Hart, Schaffner and Marx, New York, 1921.
13. Drucker, P. F., *Innovation and Entrepreneurship*, Harper and Row, New York, 1985.
14. Kirzner, I. M., *Competition and Entrepreneurship*, University of Chicago Press, Chicago, IL, 1973.
15. Shane, S. and Venkataraman, S., The promise of entrepreneurship as a field of research. *Acad. Manage. Rev.*, 2000, **25**(1), 217–226.
16. Chen, C., Searching for intellectual turning points: progressive knowledge domain visualization. *Proc. Natl. Acad. Sci. USA (Suppl. 1)*, 2004, **101**, 5303–5310.

Received 10 May 2014; revised accepted 24 September 2014