The Scientific Landscape of Agile Software Development and Methodologies: A Social Network Analysis Perspective

Brian Keith-Norambuena[1] and Vianca Vega-Zepeda[1]

[1] Departamento de Ingeniería de Sistemas y Computación, Universidad Católica del Norte, Av. Angamos 0610, Antofagasta, Chile
{brian.keith, vvega}@ucn.cl

Abstract. Since the publication of the Agile Manifesto at the beginning of the 2000s, the influence of agile methods in the scientific literature of software development and the industry has grown in a considerable manner. Considering the importance of these approaches in software engineering, the purpose of this work is to provide a bibliometric analysis of the publications on agile methods in Web of Science since the beginning of the millennium. In order to accomplish this, several analyses are performed on the co-citation networks of documents and keyword. In particular, this article provides a list of emerging topics and trends using burst analysis and a list of relevant works in the field of agile software development. These results provide a high-level overview of the main topics in current agile software development research.

Keywords: Software engineering, Agile methodologies, Social network analysis, Bibliometric analysis.

1 Introduction

Agile software development is an approach to software development that arises as a reaction to the traditional ways of developing software. Agile methods provide an alternative to the heavy and documentation-driven development processes [1] [2]. While agile methods have existed for a long time, there are still several key challenges that must be addressed by both academia and industry practitioners.

In the past years there have been several systematic literature reviews focusing on different areas of agile software development research, ranging from early adoption issues to newer applications such as product line engineering. Also, significant progress has been made with regards to better connecting agile software development with established domains such a usability, CMMI, and global software engineering [3].

On the other hand, using agile methods as defined in the literature does not work very well and requires further adaptation by practitioners, where both the domain and the process influence the adaptation [4]. However, practitioners are less concerned about adopting agile and more concerned about sustaining agile [5].

The scientific literature on agile software development has struggled to reach a consensus towards Critical Success Factors [6, 7]. An understanding of these factors is vital towards leading agile projects into a successful outcome.

While agile software development is often associated with improved decision making, there are some failings of decision making in an agile setting that must be tackled [8]. One of the main challenges in agile projects is the lack of knowledge of agile practices [9]. In this context, the purpose of this article is to provide a starting point in terms an overview of the relevant literature and terminology for researchers and practitioners who want to tackle such challenges. Also, we provide a list of emerging topics in the field obtained through burst analysis.

Social Network Analysis (SNA) techniques are used to study the contents and interaction of the scholar networks [10]. These techniques seek to analyze the relationships within communities through structured and systematic means by mapping and analyzing relationships among its basic constituents, such as people, groups or organizations [11]. This approach provides a way to find non-trivial network patterns in a social network. It is also capable of handling large volumes of data [12] and identifying critical knowledge gaps in a domain [13]. These analyses also provide a guideline for new research lines. SNA techniques can find regularities in the work of academics and can reveal the underlying structure of knowledge within a research field. Also, it provides a way to identify communities of researches in a network [14] [15]. These techniques can also be applied to perform citation analyses [16] [17].

Bibliometric analysis is regularly applied to understand the landscape of research and innovation in emerging domains [18] [19]. One way of enhancing our understanding of a field is through the identification of emerging and fading trends in the field. To do this, one can perform a popularity-based analysis or a network-based analysis [20]. The works by Ord et al. [21] and Huang [22] study are examples of popularity-based approach. Text mining approaches can be used to analyze keywords as well, this has been applied to patents to identify new technological opportunities [23].

Some examples of network-based approach are: Whitley and Galliers [24] who analyze the citation network formed by research papers on information systems, providing insight into emerging trends; Meyer et al. [25] create a citation network to analyze the different structures in the fields of strategic management, accounting and social simulation; and Khan and Wood [13] provide an emerging themes study through a social network approach in the field of information technology management, our work takes their approach as a general guideline.

Given the relevance of agile methods to the development of software engineering and the relative lack of analyses using social network techniques in this field; this study seeks to fill the extant void of empirical social network analysis in the field of agile software development. Thus, the aim of this work is to apply such techniques and provide a detailed analysis of the corresponding results.

The rest of the paper is organized as follows: first, the methodology applied in this study is presented. Then, the keyword and document networks are presented alongside the emerging trends obtained from the application of burst analysis. Finally, we present a series of relevant conclusions.

2 Materials and methods

In order to retrieve all the studies related to agile development methods a search was performed on the Web of Science (WoS) database on the 27th of February. In order to do this, the following research query was entered into the WoS search engine:

Searched for topic: ((agile OR scrum OR "extreme programming" OR "lean development" OR "lean software development") AND ("software engineer" OR "software engineering" OR "software development" OR "software methodology" OR "software process" OR "requirements engineering" OR "software design" OR "software construction" OR "software integration" OR "software operation" OR "software metrics" OR "software evaluation" OR "software process improvement" OR "agile software" OR "software project" OR "software project management" OR "software engineering education" OR "software developer" OR "software developing")); Years: 2000 – 2018. Coverage: Web of Science Core Collection.
Scrum, XP and Lean were selected due to them being the most prominent and known agile methodologies. The other parts of the string cover a myriad of different parts of the software development process.

The search retrieved 977 documents which had appeared in 326 journals from 2000 to 2017. The h-index as provided by WoS for this set of documents was 47 and the average citations per element was 10.63. The total number of citations was 10387. These documents were authored by 1006 institutions, 79 countries, and 2325 total authors.

VOSviewer [26] was utilized to construct and visualize the different networks used in this work. The main networks used in this work correspond to the keyword co-occurrence network, which are created either from the author supplied keywords or the title keywords appearing in the publications. These networks provide a way to study the knowledge created in a domain [20] [23].

The identification of emerging and fading themes on keywords was done using the burst detection algorithm provided by the Science of Science Tool [27]. Author supplied keywords and the titles of publications are the best place to look for new and fading trends in a research domain [28]. Burst detection has been used in similar studies for trend identification in several research domains [29] [30] [31] [32] [33] [34] [13]. Standard preprocessing was applied before using the algorithm indicated.

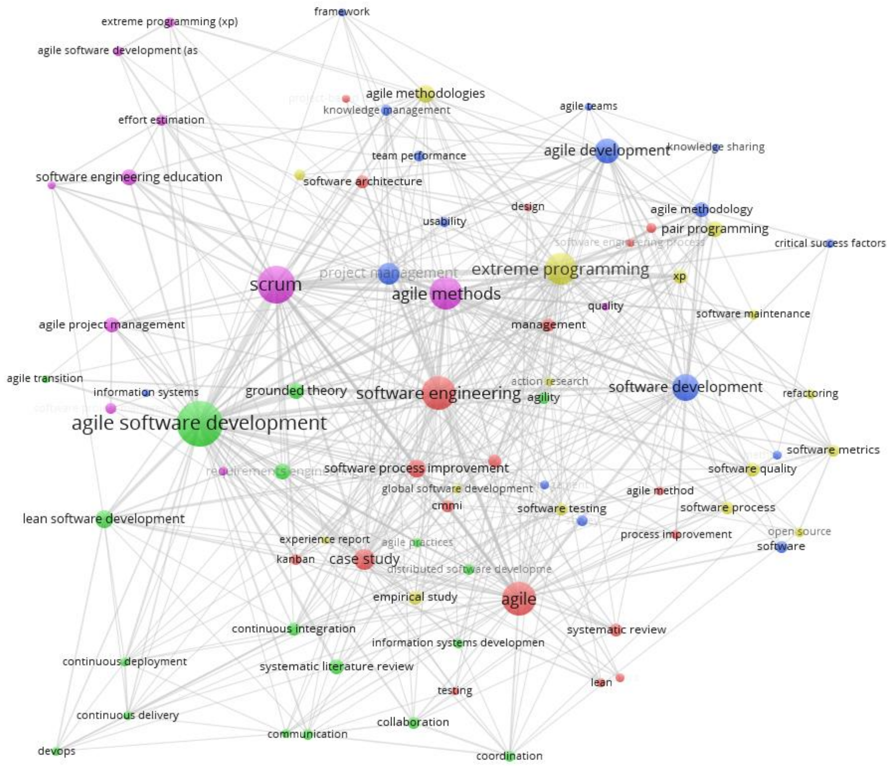
3 Results and discussion

3.1 Author supplied keywords

The network of the top 80 keywords that co-occurred at least 5 times is shown in Figure 1. The keywords are grouped into six clusters as indicated by the colors of the nodes.

Cluster 1 (Red) consists of 19 nodes; its most prominent keywords are "agile" and "software engineering", both with 76 occurrences, followed by "case study" (29) and "software process improvement" (21). This cluster seems to be associated with works that center on the incorporation of agile methods as a software process improvement strategy. Cluster 2 (Green) consists of 18 nodes; its most prominent keyword is "agile software development" itself with 140 occurrences. Other interesting keywords in this cluster are "lean software development" (22), "requirements engineering" (18), "grounded theory" (17) and "continuous integration" (12). There does not seem to be any explicit relationship between these terms. Cluster 3 (Blue) consists of 16 nodes; its most prominent keyword is "software development" (48), followed by "agile development" (43) and "project management" (33). This cluster seems to be generated by keywords from works relating to software project management with an agile approach.

Figure 1: Top 80 author supplied keyword that occurred at least 5 times.



Cluster 4 (Yellow) consists of 16 nodes; its most prominent keyword is "extreme programming" (67), followed by "agile methodologies" (21) and "pair programming" (18). This cluster is centered on extreme programming and in particular it is interesting to note that its most prominent practice appears here. Cluster 5 (Purple) consists of only 11 nodes; its most prominent keyword is "SCRUM" with 94 occurrences, followed by "agile methods" (71) and "software engineering education" far behind with 17 occurrences. This could be related to the interest in using SCRUM for software project management in software engineering courses. Some additional keywords that might be of interest in this cluster are "agile project management" (15) and "effort estimation" (8). This part could be related to complexity of estimating the required efforts with agile methods, which are inherently prone to accepting changes, which would impact the original estimations.

As part of our keyword analysis we present a list of emerging topics, some of which are still relevant today and others which lie in a state of dormancy in terms of publications (see Table 1). These topics were obtained using the burst detection algorithm, which generates a list of the word bursts ranked according to their burst weight. The burst weight corresponds to the intensity of the burst. In terms of author supplied keywords we can see that most of the active topics shown in Table 1 is mostly composed by topics that have emerged this year. It is important to note that some of these terms are trivial or empty for the purposes of this work (e.g. they might be general agile development synonyms or are simply too general).

The currently active terms highlight potential areas of research that are currently the focus of agile methodologies research. While the inactive terms show which areas are not as popular or significant as they were before. It is important to note that while these topics might be currently not as popular since they have only been inactive since recent times they might yet be useful for new research. The words have been stemmed due to preprocessing, thus some terms might refer to different instances of the same root (e.g. program and programming would both have "program" as root).

Table 1: The bursting and disappearing topics extracted from author supplied keywords (words are stemmed due to preprocessing).

Word	Weight	Length	Start	End
object	3.02	8	2001	2008
program	21.87	8	2002	2009
extrem	20.61	6	2002	2007
method	4.02	6	2002	2007
metric	3.56	10	2003	2012
xp	4.95	9	2003	2011
open	3.76	4	2004	2007
sourc	3.38	4	2004	2007
empir	3.02	6	2005	2010
pair	3.81	4	2006	2009
methodolog	4.74	4	2006	2009
adapt	3.18	6	2008	2013
manufactur	3.49	3	2009	2011
theori	3.5	4	2009	2012
framework	3.16	2	2017	-
analyt	3.03	2	2017	-
effort	3.29	2	2017	-
continu	3.82	2	2017	-
devop	3.03	2	2017	-

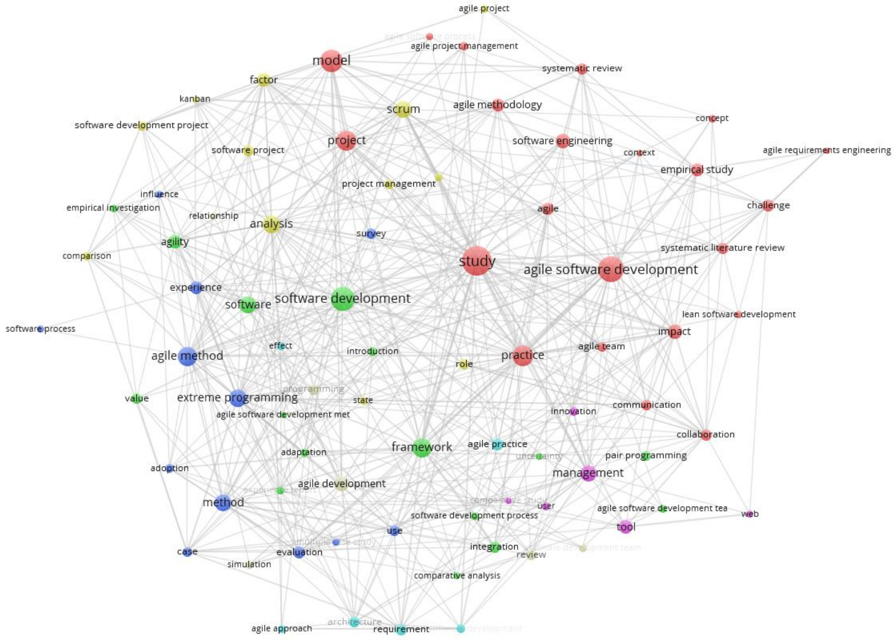
For the purposes of this work, even though there are several words that can be associated with many software engineering concepts (e.g. framework), we assume that in the context of this work they are used in the context of agile methods. Although the idea of software development frameworks has a long history, the results show that in the agile software development community they are currently in vogue. Other keywords that are currently trending according to the burst detection algorithm corresponds to "analyt" (possibly referring to analytics), "effort", which possibly reflects the inherent difficulty of effort estimation with methods that embrace change, "continuous" (probably referring to continuous integration or a similar topic) and "DevOps". Note that all such active trends started in 2017.

On the other hand, there are terms such as “XP” and “extreme” are not currently in vogue with researchers according to the findings of burst detection, with its prime period ending between 2007 and 2011.

3.2 Title supplied keywords

For the purposes of our graphical analysis, the top 80 terms that occurred at least 5 times are included in Figure 2. Node size represents the number of occurrences and the links represent the co-occurrence relationship between the different title keywords. Based on co-occurrence relationships, words are grouped into seven clusters.

Figure 2: Top 80 title keywords that occurred at least 5 times.



Cluster 1 (Red) consists of 22 nodes; its most prominent keyword is "study" with 93 occurrences, followed by "agile software development" (71) and "model" (55). There seems to be no further relationship among these concepts. Other notable keywords in this node relate to "practice" (47) and "project" (44). Cluster 2 (Green) consists of 16 nodes; its most prominent keyword is "software development" itself with 61 occurrences, followed by "framework" (41) and "software" (31). Perhaps the source of the importance of “framework” comes from the fact that the software engineering community is searching for a more detailed guide on how to apply agile methods. Cluster 3 (Blue) consists of 12 nodes; its most prominent keyword is "agile method" with 41 occurrences, followed by "extreme programming" (34) and "method" (30). These relationships are trivial, since XP is an agile method. Cluster 4 (Yellow)

consists of 14 nodes; its most prominent keyword is "analysis" (34), followed by "SCRUM" (31) and "factor" (21). There seems to be no direct relationship between these terms.

The last three cluster seem to be less relevant compared to the previous clusters, taking as a metric for relevance the total occurrences of its constituents. Cluster 5 (Purple) consists of only 6 nodes; its most prominent keywords are "management" (26), "tool" (21) and "innovation" (9). This cluster could be related to the difficulty of managing a project where the scope is subject to change due to the nature of agile methods. Cluster 6 (Cyan) consists of only 6 nodes; its most prominent keywords are "agile practice" (16), "requirement" (15) and "architecture" (13). This cluster could be related to the recent search of a methodology for the application of an agile approach for requirements engineering. Since requirements are so important for software development, it is relevant to clarify them without losing agility. Cluster 7 (Beige) consists of only 6 nodes and its most prominent keywords are "agile development" (25), "programming" (14) and "review" (13). There seems to be no apparent unifying theme in this cluster.

Table 2 shows the emerging and disappearing themes extracted from the titles of the documents, in terms of currently active words the results are quite similar to those found in Table 1 with respect to original keywords, however, the term “estim” (estimation) appears as a bursting topic since 2017.

Table 2: The bursting and disappearing topics extracted from title keywords (words are stemmed due to preprocessing).

Word	**Weight**	**Length**	**Start**	**End**
program	16.59	9	2001	2009
extrem	16.57	6	2001	2006
xp	4.92	3	2003	2005
pair	4.41	4	2006	2009
driven	3.28	4	2010	2013
improve	3.28	1	2010	2010
distribut	3.44	4	2012	2015
systemat	4.49	2	2015	2016
literature	4.58	2	2015	2016
review	5.21	1	2015	2015
challeng	3.19	1	2016	2016
effort	3.07	2	2017	-
continu	3.61	2	2017	-
estim	3.15	2	2017	-

Note that framework does not appear as a bursting nor disappearing topic when using title keywords. Inactive terms show more variation compared to the case of the original author keywords. Some interesting inactive terms are “driven” (2010-2013), possibly referring to test-driven development and of course “XP” and “extreme” referring to extreme programming again as in the case of author keywords. Another interesting example is "pair" and "program", possibly referring to the practice of "pair programming", which according to this stopped being in vogue by 2009.

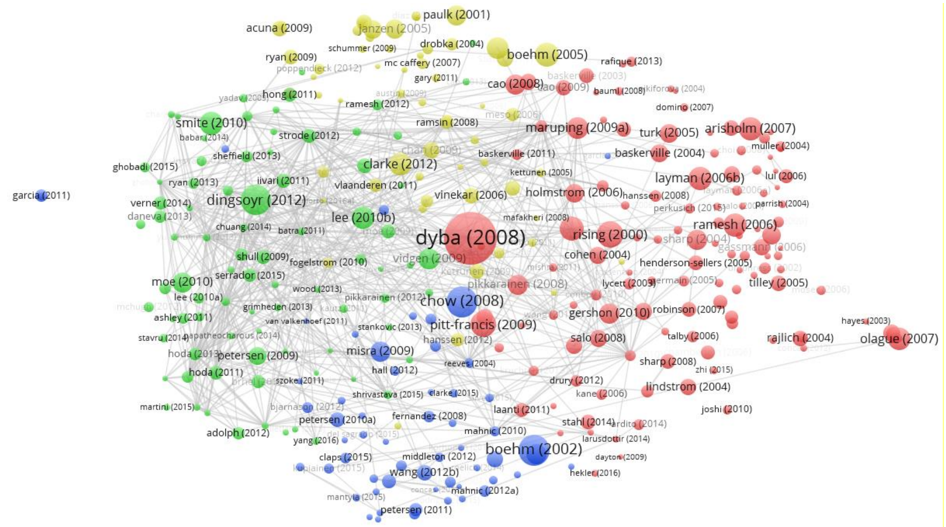
3.3 Relevant document analysis

Finally, we present an analysis of the most relevant documents in terms of citations. First, we show the relation between documents in a graph constructed using the bibliographical coupling among the publications (See Figure 3). Only the top 330 documents are shown in the graph, due to forming the largest connected component. Based on this graph and using the bibliographical coupling of the documents we can identify a total of four main clusters (See Table 3). The publications presented in Table 3 provide a starting point for researchers that are interested in learning about the most influential topics in this area of research.

Table 3: Top 3 publications (number of citations) in each cluster of Figure 3.

Cluster	Title	Authors	Year	Citations
1 (Red)	Empirical studies of agile software development: A systematic review	Tore Dybå & Torgeir Dingsøyr	2008	504
	The Scrum software development process for small teams	Linda Rising & Norman Janoff	2000	127
	Chaste: A test-driven approach to software development for biological modelling	Joe Pitt-Francis et al.	2009	119
2 (Blue)	A survey study of critical success factors in agile software projects	Tsun Chow & Dac-Buu Cao	2008	178
	Get ready for agile methods, with care	Barry Boehm	2002	172
	Identifying some important success factors in adopting agile software development practices	Subhas Chandra Misra, Vinod Kumar & Uma Kumar	2009	72
3 (Green)	A decade of agile methodologies: Towards explaining agile software development	Torgeir Dingsøyr, Sridhar Nerur, VenuGopal Balijepally & Nils Brede Moe	2012	163
	Empirical evidence in global software engineering: a systematic review	Darja Šmite, Claes Wohlin, Tony Gorschek & Robert Feldt	2009	100
	Toward agile: an integrated analysis of Quantitative and qualitative field data on software development agility	Gwanhoo Lee & Weidong Xia	2010	95
4 (Yellow)	Management challenges to implementing agile processes in traditional development organizations	Barry Boehm & Richard Turner	2005	104
	The situational factors that affect the software development process: Towards a comprehensive reference framework	Paul Clarke & Rory V.O'Connor	2011	91
	Agile Modeling, Agile Software Development, and Extreme Programming: The State of Research	John Erickson, Kalle Lyytinen & Keng Siau	2005	88

Figure 3: Bibliographical coupling network formed by the top 330 documents.



4 Conclusions

This article has presented a detailed bibliometric analysis of the agile software development field by querying the Web of Science database. The retrieved document set has been analyzed through SNA techniques. The results of this analysis have provided several insights on the nature of research in the agile software development field. Finally, an overview of some of the most relevant works in scientific landscape of agile software development has been presented.

Through SNA techniques this study has analyzed the keyword networks in the agile software development domain. Also, important documents from the field and the most relevant countries were analyzed by their contribution. The analysis also included the application of the burst detection algorithm to identify emerging and fading themes in academic research published in WoS. In particular, several emerging and disappearing topics were identified, in this context the potential importance of the topics emergent in the previous year (2017) must be highlighted. Topics such as "effort estimation" (or perhaps other kinds of estimation) seem to be in vogue in current research. Other emergent themes that must be considered are frameworks for agile methodologies, analytics and continuous integration and similar topics.

The limitations of the study must be addressed. This work has only considered the main WoS database. Thus, several other sources that publish research on agile software development or its applications are not included in this research. Also, it is important to note that recently, agile scaled approaches or new frameworks based on Scrum and Lean have been proposed. These topics were not included in this analysis. Therefore, it is necessary to be cautious when generalizing the results from this work.

References

1. Cohen, D., Lindvall, M., & Costa, P. Agile software development. DACS SOAR Report, 11, 38 (2003).
2. Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... & Kern, J. Manifesto for agile software development (2001).
3. Hoda, R., Salleh, N., Grundy, J., & Tee, H. M. Systematic literature reviews in agile software development: A tertiary study. Information and Software Technology, 85, 60-70 (2017).
4. Diebold, P., & Dahlem, M. Agile practices in practice: a mapping study. In Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (p. 30). ACM (2014).
5. Gregory, P., Barroca, L., Sharp, H., Deshpande, A., & Taylor, K. The challenges that challenge: Engaging with agile practitioners' concerns. Information and Software Technology, 77, 92-104 (2016).
6. Aldahmash, A., Gravell, A. M., & Howard, Y. A Review on the Critical Success Factors of Agile Software Development. In European Conference on Software Process Improvement (pp. 504-512). Springer, Cham (2017).
7. Senapathi, M., & Srinivasan, A. Sustained agile usage: a systematic literature review. In Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering (pp. 119-124). ACM (2013).
8. Drury-Grogan, M. L., Conboy, K., & Acton, T. Examining decision characteristics & challenges for agile software development. Journal of Systems and Software, 131, 248-265 (2017).
9. Fortunato, C. A., Furtado, F., Selleri, F., de Farias Junior, I., & Júnior, N. L. Quality Assurance in Agile Software Development: A Systematic Review. In Brazilian Workshop on Agile Methods (pp. 142-148). Springer, Cham (2016).
10. Wasserman, S., & Faust, K. Social network analysis: Methods and applications (Vol. 8). Cambridge university press (1994).
11. Cross, R., Parker, A., Prusak, L., & Borgatti, S. P. Knowing what we know. Networks in the knowledge economy, 208 (2003).
12. Khan, G. F., & Park, H. W. The e-government research domain: A triple helix network analysis of collaboration at the regional, country, and institutional levels. Government Information Quarterly, 30(2), 182-193 (2013).
13. Khan, G. F., & Wood, J. Information technology management domain: emerging themes and keyword analysis. Scientometrics, 105(2), 959-972 (2015).
14. Vidgen, R., Henneberg, S., & Naudé, P. What sort of community is the European Conference on Information Systems? A social network analysis 1993–2005. European Journal of Information Systems, 16(1), 5-19 (2007).
15. Krystallis, A., Ormond, R., & Christensen, K. V. Patterns and regularities in the European marketing academic community: A social network analysis of the EMAC annual conferences 2000–2010. In European marketing academy: Conference Proceedings of the 2011 EMAC Conference (2011).
16. Zinkhan, G. M., Roth, M. S., & Saxton, M. J. Knowledge development and scientific status in consumer-behavior research: A social exchange perspective. Journal of Consumer Research, 19(2), 282-291 (1992).
17. Carter, C. R., Leuschner, R., & Rogers, D. S. (2007). A social network analysis

of the Journal of Supply Chain Management: knowledge generation, knowledge diffusion and thought leadership. Journal of Supply Chain Management, 43(2), 15-28.

18. Small, H., Boyack, K. W., & Klavans, R. Identifying emerging topics in science and technology. Research Policy, 43(8), 1450-1467 (2014).
19. Shapira, P., Kwon, S., & Youtie, J. Tracking the emergence of synthetic biology. Scientometrics, 112(3), 1439-1469 (2017).
20. Choi, J., Yi, S., & Lee, K. C. Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. Information & Management, 48(8), 371-381 (2011).
21. Ord, T. J., Martins, E. P., Thakur, S., Mane, K. K., & Börner, K. Trends in animal behaviour research (1968–2002): ethoinformatics and the mining of library databases. Animal Behaviour, 69(6), 1399-1413 (2005).
22. Huang, C. P. Bibliometric analysis of obstructive sleep apnea research trends. Journal of the Chinese Medical Association, 72(3), 117-123 (2009).
23. Yoon, B., & Park, Y. A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. Technological Forecasting and Social Change, 72(2), 145-160 (2005).
24. Whitley, E. A., & Galliers, R. D. An alternative perspective on citation classics: Evidence from the first 10 years of the European Conference on Information Systems. Information & Management, 44(5), 441-455 (2007).
25. Meyer, M., Lorscheid, I., & Troitzsch, K. G. The development of social simulation as reflected in the first ten years of JASSS: a citation and co-citation analysis. Journal of Artificial Societies and Social Simulation, 12(4), 12 (2009).
26. Van Eck, N. J., & Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538 (2010).
27. Team, S. Science of science (Sci2) tool. Indiana University and SciTech Strategies (2009).
28. Leydesdorff, L. The knowledge-based economy: Modeled, measured, simulated. Universal-Publishers (2006).
29. Kleinberg, J. Bursty and hierarchical structure in streams. Data Mining and Knowledge Discovery, 7(4), 373-397 (2003).
30. Mane, K. K., & Börner, K. Mapping topics and topic bursts in PNAS. Proceedings of the National Academy of Sciences, 101(suppl 1), 5287-5290 (2004).
31. Chen, C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the Association for Information Science and Technology, 57(3), 359-377 (2006).
32. Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. Towards an explanatory and computational theory of scientific discovery. Journal of Informetrics, 3(3), 191-209 (2009).
33. Guo, H., Weingart, S., & Börner, K. Mixed-indicators model for identifying emerging research areas. Scientometrics, 89(1), 421-435 (2011).
34. Swar, B., & Khan, G. F. Mapping ICT knowledge infrastructure in South Asia. Scientometrics, 99(1), 117-137 (2014).