

Bibliometric Analysis of IT/IS Investment Research

Michal Krčál¹, Ahad Zareravasan²

Department of Corporate Economy, Faculty of Economics and Administration, Masaryk University, Brno,
Czech Republic

Abstract.

The topic of IT/IS investment has been intensely researched for the last four decades. This study presents a bibliometric analysis of IT/IS investment literature published from 1983 to 2020. Totally 766 journal papers published in 419 Scopus-indexed journals by 1,600 unique authors, were analyzed using R Programming language. Both impact and content analysis, including significant, influential aspects (e.g., authors, affiliations, countries, keywords, and journals), research streams, and themes, were included in the analysis.

Keywords: Bibliometric Analysis, Citation Analysis, Content Analysis, Information System (IS) Investment, Information Technology (IT) Investment

1. Introduction

Firms invest heavily in their Information Technology or Information system (IT/IS) infrastructure (hardware, software, networks, and storage) to boost their performance and to respond to a turbulent environment properly (Lu & Ramamurthy, 2011). While organizations engage in different types of strategic decisions, those related to IT/IS investments are essential, yet, complicated (Bernroider & Schmöllerl, 2013) because they encompass different strategic, financial, social, managerial, and technological aspects. During the last four decades, IT/IS scholars have explored many aspects of IT/IS investment from different points of view.

To identify the trends of IT/IS investment research in different fields, researchers have to review a large number of sources and databases containing relevant papers. Bibliometric analysis is an option that brings initial insights into the vast number of published studies. Bibliometric analysis is defined as the application of mathematical and statistical methods to papers, books, and other means of communication used in the analysis of scientific publications (Repanovici, 2011). Bibliometric methods have been used to report scientific measures in many disciplines of science and technology. They are a standard research instrument for the systematic analysis of IT/IS literature (e.g., Kalantari et al., 2017; Miao & Yang, 2018; Mortenson & Vidgen, 2016).

The bibliometric analysis can significantly improve the quality of literature review by introducing a transparent, systematic, and reproducible review process. It provides means for mapping the set of publications and identifying influential studies without relying on the researcher's intuitiveness (Ellegaard & Wallin, 2015). The bibliometric analysis benefits researchers by discovering top-cited publications, authors, research trends, and emerging topics (Fardi et al., 2011). It can guide researchers to find the best field or best journal to succeed in their publication. Utterly, it can also help researchers to obtain a preliminary idea about the articles and research that has an impact on a particular area of interest (Ale Ebrahim et al., 2014; Gomez-Jauregui et al., 2014). In the subsequent second section, we present the method used for the bibliometric analysis. Then, we present the findings from the impact and content analysis in the third section. The fourth section contains conclusions and limitations.

2. Method

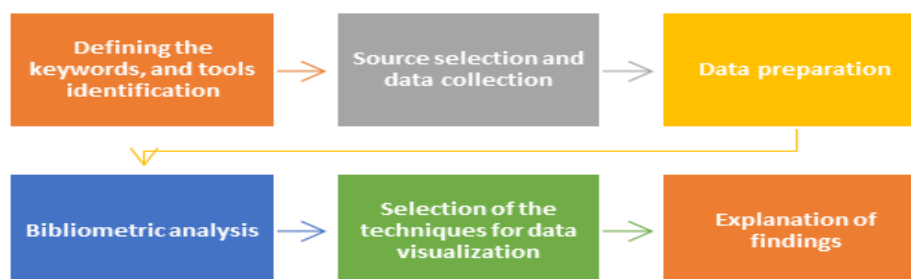
Bibliometric analysis is among the oldest research methods in library and information science (Wilson, 2013). The bibliometric analysis contains various applications from information science, sociology, history of science, and research evaluation. This method is used to evaluate, monitor, and visualize the structure of scientific fields (Rouzbahani et al., 2020). Bibliometric methods can be divided into two parts: general instructions and publication analysis. For general instructions, researchers show how to avoid possible sources of error in the search process by showing how to search articles using a search engine. The publication analysis includes impact factor, citations, publisher, and country statistics.

Citation analysis helps to identify core publications, journals, countries and shows a relationship between citing and cited works in the analyzed area. Three primary bibliometric data sources for searching the literature are Web of Science (WoS), Scopus, and Google Scholar. In this paper, the Scopus data were exported and analyzed because of the width of coverage, quality of data, (almost) standardized format, and ease of extraction made Scopus an ideal choice for our bibliometric analysis.

To ensure that all the relevant articles to the research topic were included in our final dataset, the Title, Abstract, and Keywords of the published papers in Scopus were queried using Wildcard (*), a Boolean operator (OR), and proximity operator W/n (within the n words). The final query was: *TITLE-ABS-KEY ("Information * technolog*" OR "Information * system*") W/1 TITLE-ABS-KEY (investing OR investment*)*. The preliminary results included 1,329 papers. Limiting the papers to journal articles yielded to 801 papers and excluding non-English papers resulted in 766 unique journal papers. There were no limits for the period of the published papers. Then, using the Scopus feature, we downloaded all the metadata as a single CSV file. We analyzed the data in R using *stringr*, and *tm* (text mining) packages to prepare the data for further analysis. We used 'Biblioshiny' which is a tool designed for non-coders to provide means for complete scientometric and bibliometric analysis offering numerous options divided into categories of sources, documents, authors, conceptual structure, social structure,

and intellectual structure. It allows obtaining multiple results in the shape of tables and graphs, which are not common in other software (Aria & Cuccurullo, 2017). Figure 1 represents the six steps to complete the process of bibliometric analysis of the literature.

Figure 1. The main research steps



In the data preparation phase, we used standard steps of the natural language processing (NLP):

1. Removing keywords from the title, abstract, author's keyword, and index (Scopus) keywords, as it is expected to see them frequently in the text. These words included: information technology, IT, information and communication technology, ICT, information system, IS, investment, and investing.
2. Turning the Abstract, Author keywords and index keywords to lowercase (because R is case-sensitive).
3. Removing the stopwords (short function words such as "I", "of", "and", "is", and "it"), punctuation and non-alphabetical characters.
4. Stemming of words (reducing each word to its shortest stem such as "technology", "technologies", "Technological" and alike are all reduced to their shortest stem, "technolog"). This means that all words that share a stem are treated as the same in the analysis.

3. Findings

This section presents the results of the bibliometric analysis. Totally 766 journal papers published in 419 unique journals, by 1,600 unique authors were investigated for the analysis purpose. 187 out of total 766 papers were single-authored documents. The authors per document index equals 2.09, and the total number of author's keywords were 1,688. To analyze the publication, citation analysis, and content analysis as the two widely used bibliometric methods is presented in the next two subsections.

3.1 Impact Analysis

Within the collected data, the number of publications was increasing steadily until 2000, while there is a fluctuation afterwards as shown in Figures 2, with a peak in 2015 and 2019.

The top ten journals within our dataset based on the journal H-index are also illustrated in Table 1. This table ranks the articles based on h-index, total citation (TC), and net production (NP) and publication starting year (PY_start). Journal of Management Information Systems, Information and Management, Information systems Research, and MIS quarterly are the sources with the highest impact.

Figure 2. Number of published articles by years

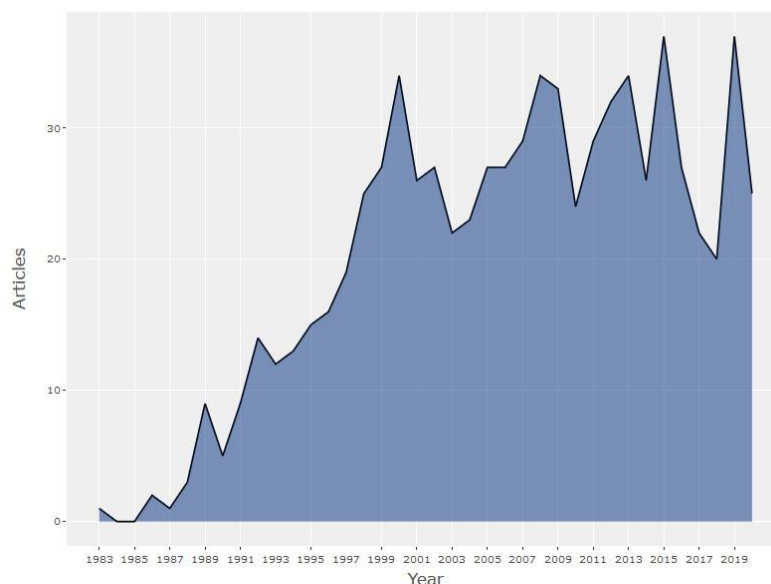


Table 1. Top ten journals with the highest H-index

Journal title	h-index	TC	NP	PY
Journal of Management Information Systems	26	4208	30	1990
MIS Quarterly: Management Information Systems	19	4491	20	1989
Information Systems Research	17	3394	22	1992
Information and Management	16	1223	19	1989
Management Science	13	1835	14	1994
International Journal of Information Management	11	259	12	1986
Journal of Strategic Information Systems	10	533	12	1992
Decision Support Systems	9	262	9	1992
Journal of Computer Information Systems	7	230	12	1998
Journal of Information Technology	7	438	8	1990

To find the core journals, we use source impact and Bradford Law. Table 2 represents Bradford's law, which divides the journals into three zones. Zones 1 represents our core sources, representing journals having significant publications. We found that 25 journals (out of 419) belong to core zone 1 (publishing one-third of all papers), the rest of the journals are in zone 2 and zone 3.

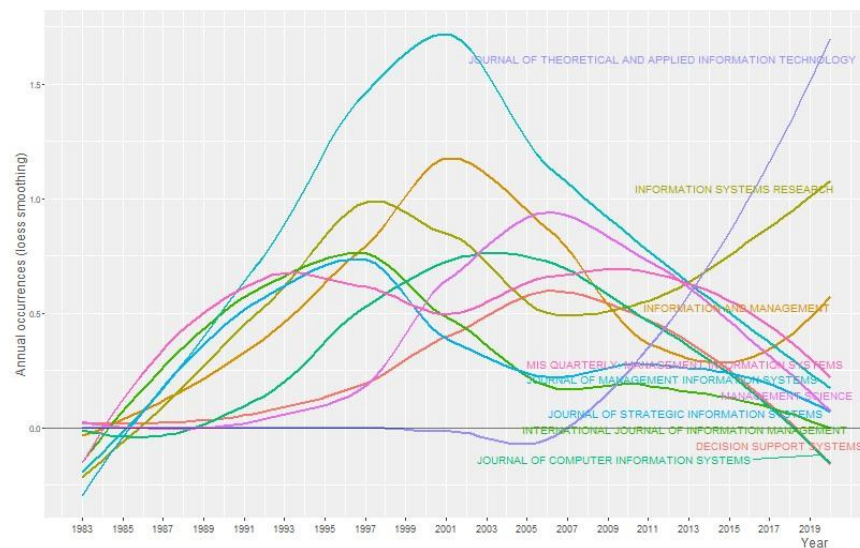
Table 2. Top journals based on Bradford's law

Journal title	Rank	cumFreq
Journal Of Management Information Systems	1	30
Information Systems Research	2	52
MIS Quarterly: Management Information Systems	3	72
Information and Management	4	91
Management Science	5	105
International Journal of Information Management	6	117
Journal of Computer Information Systems	7	129
Journal of Strategic Information Systems	8	141
Journal of Theoretical and Applied Information Technology	9	151
Decision Support Systems	10	160
International Journal of Accounting Information Systems	11	168
International Journal of Business Information Systems	12	176
Journal Of Information Technology	13	184
IEEE Transactions on Engineering Management	14	191
Information Systems Frontiers	15	198
Information Systems Journal	16	205
European Journal of Information Systems	17	211
Industrial Management and Data Systems	18	217
Information Systems Management	19	223
Journal of Enterprise Information Management	20	229
Technovation	21	235
Information Technology for Development	22	240
Journal of Organizational Computing and Electronic Commerce	23	245
Journal of The Operational Research Society	24	250
Advanced Science Letters	25	254

Moreover, we plotted the trend of the number of publications per year per top ten sources. Figure 3 shows the growth in the publication by top journals. We use the loess smoothing technique that is the locally weighted smoothing use regression analysis to demonstrate the smooth line with the help of a time plot or scatter plot. Loess smoothing helps to understand trends through time (Royston, 1992). As can be seen from Figure 3, most of the journals are ceasing to publish IT/IS investment research. However, the Journal of Theoretical and Applied Information Technology (Scopus indexed, Q4 ranked journal) has been increasingly publishing IT/IS investment research. Another two Q1 ranked journals (Information Systems Research,

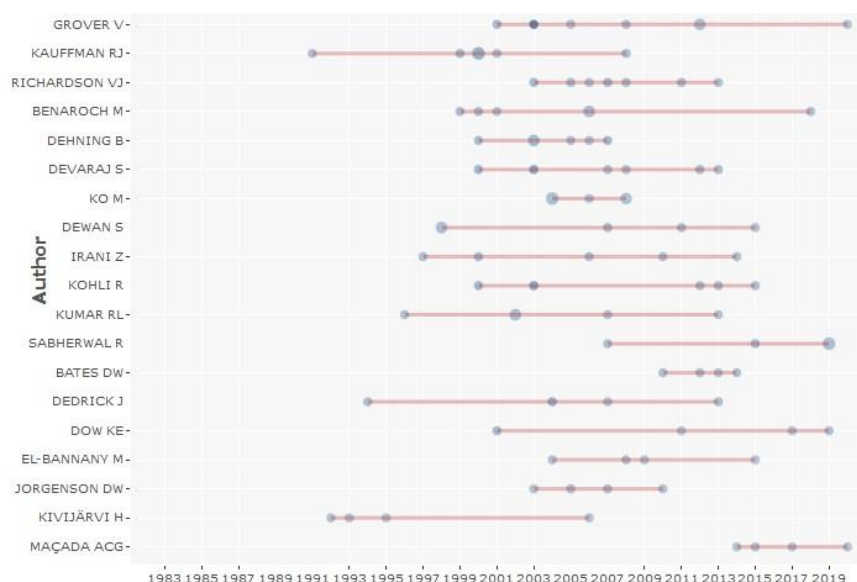
Information and Management) also increased publications in the IT/IS investment field contrary to other respected journals such as MIS Quarterly or Journal of Strategic Information Systems

Figure 3. Source growths based on the number of publications per year



Next, top authors based on H-index, number of publications or number of citations can be discovered. We plot the number of publications of the top 20 authors throughout the period in figure 4. It shows that most of the prominent authors are classic ones which have been publishing in the field and very few of them such as Grover_V or Dow_Ke, have published till recently. Besides, there is only one emerging author (Macada_Acg). Further, Lotka's law was derived to describe the publication frequency by authors in this given field. Lotka's study indicated that 88.9% of authors make a single publication during a given period. As the number of publications by a single author increases, the number of authors decreases. The general formula says $x^n y = \text{constant}$, which represents the inverse power law of author productivity (Miau & Yang, 2018). Based on Lotka's law, only 8%, and 1.6% of authors made two and three publications, respectively.

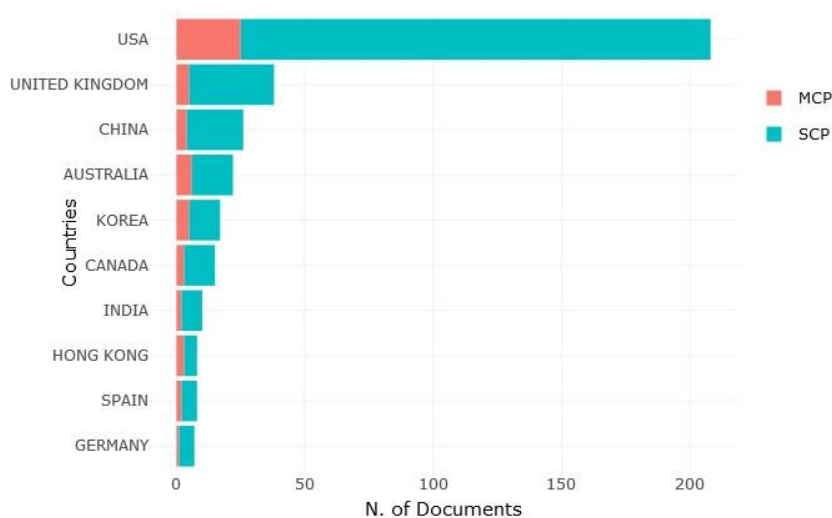
Figure 4. Top 20 authors' production over time



The most relevant affiliations consist of the University of California, University of Arkansas, Bina Nusantara, National University of Singapore, and the University of Brunel ranked the top five universities with the highest number of publications during the period.

Alike, the corresponding authors' countries are depicted in Figure 5, which shows the number of publications based on single country publications (SCP) and multiple country publications (MCP). MCP refers to at least one co-author from a foreign country. However, regarding the number of citations, the USA scores the highest citations (16,452) followed by UK (1,594) and Australia (1,305).

Figure 5. Top ten corresponding authors' countries



3.2 Quantitative content analysis

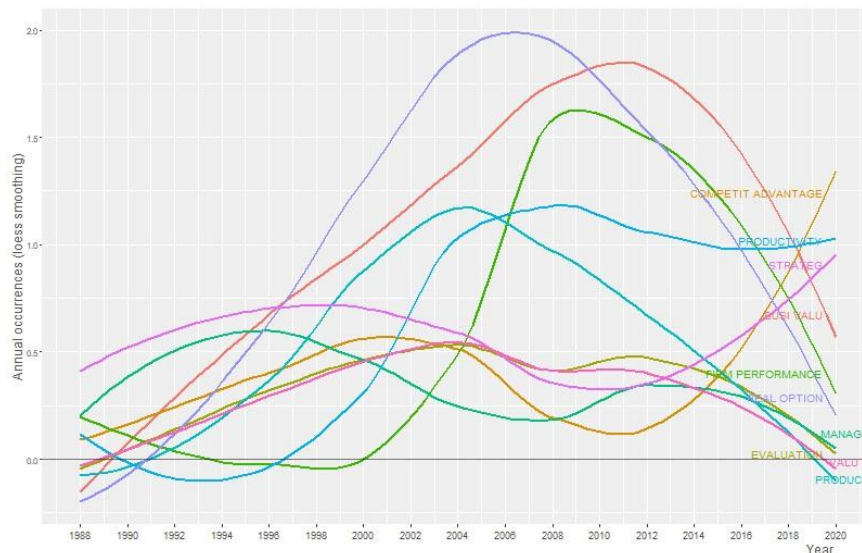
We analyzed the most frequent author's keywords in the stemmed or root version of each keyword. Accordingly, "busi valu", "real option", "firm performance", "productivity", and "strateg" were the top five author keywords. We also visualized the top 40 most frequent author's keywords using a word cloud in Figure 6.

Figure 6. WordCloud of top 40 (stemmed) author's keywords



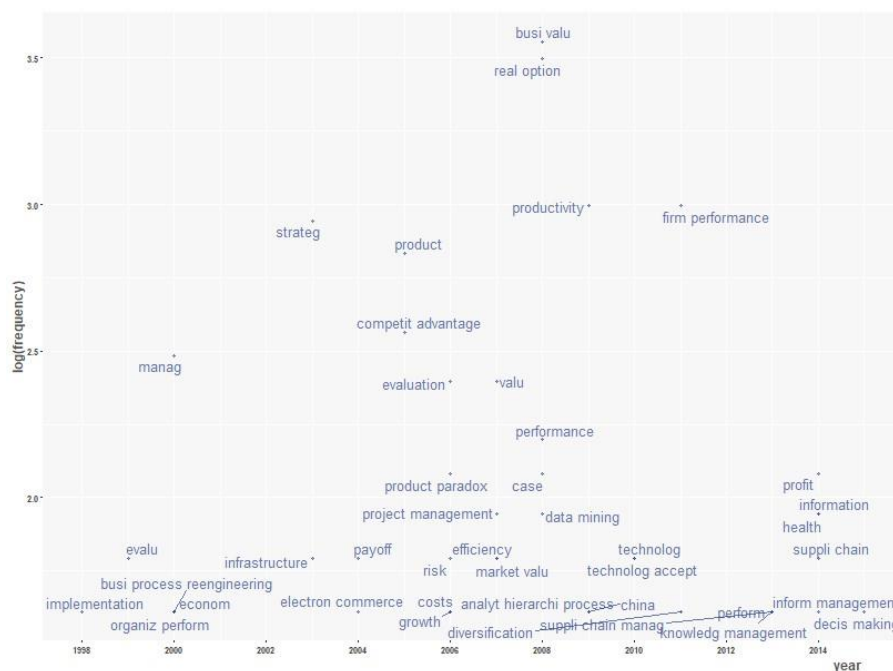
Another interesting plot that can be drawn using the top ten author's keyword is the trend of the keywords growth over time, as illustrated in Figure 7. It shows that the trend was almost steady for some of the keywords. However specifically for "competit advantage", "busi valu", "real option", "manag", and "firm performance", a rising pattern can be observed, until 2005-2012 and then a drop afterwards.

Figure 7. Top ten author's keywords trend



The trend of the keywords can also be drawn, based on the most frequent keywords per year. Figure 8 plots the logarithm of the author's keywords frequency versus years. As can be observed from the figure, "econom", "evalu", or "implementation" are the commonly used keywords in the early years. In contrast, "inform management", "supply chain", "health information" keywords appeared most recently. It shows that the concerns of scholars have shifted from IT/IS economic evaluation to business value creation from the investment. It also shows that recently we had more tendency to industry-specific applications or decision making in supply chain and healthcare. According to the figure, we have "busi valu" and "real option" as the most frequent keywords in 2008 and total.

Figure 8. The trend of author's keywords



Using the information of the author's keywords, authors' names, and countries, we can draw a Three-Fields Plot as illustrated in Figure 9. This plot shows the most relevant keywords, authors, and countries in a single view. The size of boxes refers to the number of frequency. Using this plot, it is possible to trace the contribution of each author on the most frequent keywords and his/her respected country. For instance, it can be observed that Irani_Z has appeared 12 times, mainly worked on "evaluation" and affiliated with the UK.

Figure 10 shows the co-occurrence network of the author's keywords. There are four main clusters of keywords specified in different colors. The violet and green clusters seem to cover almost the same set of topics, but with different words. For instance, we have organization performance and productivity, in the green cluster, which technically has the same concept as business value and firm performance in the violet cluster. Nevertheless, R is not able to deal with synonyms, and it could be done only manually for each term which is time consuming with large datasets. The blue cluster addresses IT/IS evaluation and risk management with a focus on real options theory. Myers (1977) coined the term *real options* and envisioned bringing the theory of financial options to the realm of strategic decision making. Real options were seen as opportunities to purchase real assets on possibly favorable terms (Trigeorgis & Reuer, 2017). The red cluster also concerns the supply chain and value creation through IT/IS investments.

Figure 9. Three-Fields Plot

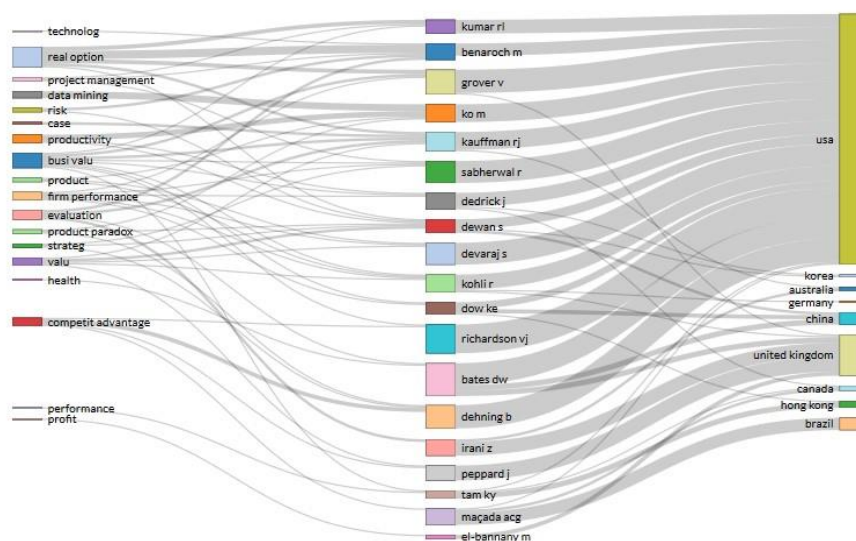


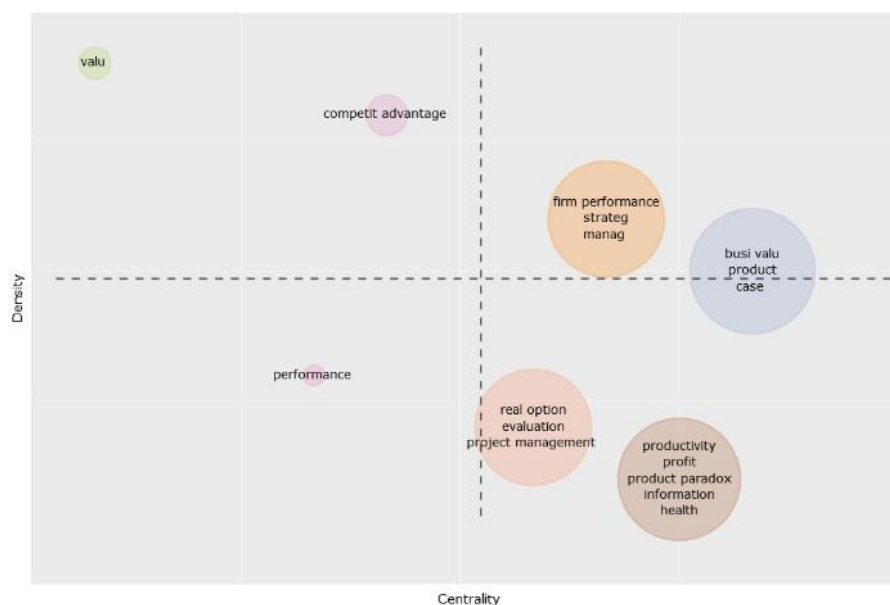
Figure 10. Co-occurrence network of author's keywords



We can further categorize the identified themes into a strategic diagram to analyze the importance and development of the research theme (Aria & Cuccurullo, 2017). Figure 11 represents the thematic map based on density (y-axis) and centrality (x-axis). The centrality measures the importance of the selected theme, and density measures the development of the chosen theme. The graph is divided into four parts. Themes that appear on the lower left part are emerging or declining themes (performance, as a declining theme in our case). These are new themes that can emerge to be better or drop from the research area. Themes that come under the lower right part of the thematic map are the primary or transversal themes (categories

with productivity or real option). These themes represent low density but high centrality. Much research has been done on these themes. The Upper left part represents high density but lower centrality, these themes are highly developed but isolated (competitive advantage, and value). The upper right part represents high density and high centrality. The themes in this part are highly developed and essential (firm performance and business value). However, they tend to move to basic or transversal themes. The size of the thematic map is to the factors that come under the theme (*Biblioshiny_Bibliometrix for No Coders.*, 2020). We have used the top 250 author's keywords for visualization, but items shown in the clusters are set to the minimum frequency of 10.

Figure 11. Thematic Map



The last figure is the social structures that show how authors, institutions, or countries relate to others in the research domain. In this plot, a group of regular authors, institutions, or countries, influent ones, and hidden communities among them can be discovered. Figure 12 visualizes the institutional collaboration network.

Figure 12. Institutional collaboration network



4. Conclusion and Limitations

This study provided a bibliometric analysis of IT/IS investment literature published from 1983 to 2020. We used several R packages and Biblioshiny library for R to conduct the impact and content analysis of 766 Scopus indexed journal papers. The results of the analysis show the applicability of the R and available libraries for bibliometric analysis purposes. Besides, the results present very quickly and, in some cases, detailed view of the trend and status of the IT/IS investment literature, in terms of authors, institutions, countries, and author's keywords. However, this research suffers from some limitations which we are going to cover in our future research. First, this research only used the author's keywords for data analysis for the sake of simplicity. However, index (Scopus) keywords and abstracts can also be utilized. Second, this study did not include synonyms in content analysis. For instance, we observed "firm performance" and "organization performance" which refer to the same concept. However, no readily available R library could manage this issue (with merging the synonyms under one term). Third, we also did not include the proximity concept in our content analysis. For instance, enterprise information systems, ERP, SCM, and CRM could be considered as a type of information system. Nevertheless, in this research, we did treat them just as distinct terms. We are going to work on the limitations mentioned above in the future extensions of the current text.

Acknowledgments

We acknowledge the financial support of the Grant Agency of the Czech Republic, GAČR (grant no. GA20-12081S).

References

- [1] Ale Ebrahim, N., Salehi, H., Embi, M. A., Danaee, M., Mohammadjafari, M., Zavvari, A., Shakiba, M., & Shahbazi-Moghadam, M. (2014). Equality of Google Scholar with Web of Science citations: Case of Malaysian engineering highly cited papers. *Modern Applied Science*, 8(5), 63–69.
- [2] Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.
- [3] Bernroider, E. W. N., & Schmöllerl, P. (2013). A technological, organisational, and environmental analysis of decision making methodologies and satisfaction in the context of IT induced business transformations. *European Journal of Operational Research*, 224(1), 141–153.
- [4] Biblioshiny_Bibliometrix for no Coders. (2020, August 10). <https://bibliometrix.org/biblioshiny/>
- [5] Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, 105(3), 1809–1831.
- [6] Fardi, A., Kodonas, K., Gogos, C., & Economides, N. (2011). Top-cited articles in endodontic journals. *Journal of Endodontics*, 37(9), 1183–1190.
- [7] Gomez-Jauregui, V., Gomez-Jauregui, C., Manchado, C., & Otero, C. (2014). Information management and improvement of citation indices. *International Journal of Information Management*, 34(2), 257–271.
- [8] Kalantari, A., Kamsin, A., Kamaruddin, H. S., Ebrahim, N. A., Gani, A., Ebrahimi, A., & Shamshirband, S. (2017). A bibliometric approach to tracking big data research trends. *Journal of Big Data*, 4(1), 30.
- [9] Lu, Y., & Ramamurthy, K. (Ram). (2011). Understanding the link between information technology capability and organizational agility: An empirical examination. *MIS Quarterly*, 931–954.
- [10] Miao, S., & Yang, J.-M. (2018). Bibliometrics-based evaluation of the Blockchain research trend: 2008–March 2017. *Technology Analysis & Strategic Management*, 30(9), 1029–1045.
- [11] Mortenson, M. J., & Vidgen, R. (2016). A computational literature review of the technology acceptance model. *International Journal of Information Management*, 36(6), 1248–1259.

- [12] Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, 5(2), 147–175.
- [13] Repanovici, A. (2011). Measuring the visibility of the university's scientific production through scientometric methods. *Performance Measurement and Metrics*.
- [14] Rouzbahani, H. M., Karimipour, H., Dehghantanha, A., & Parizi, R. M. (2020). Blockchain applications in power systems: A bibliometric analysis. In *Blockchain Cybersecurity, Trust and Privacy* (pp. 129–145). Springer.
- [15] Royston, P. (1992). Lowess smoothing. *Stata Technical Bulletin*, 1(3).
- [16] Trigeorgis, L., & Reuer, J. J. (2017). Real options theory in strategic management. *Strategic Management Journal*, 38(1), 42–63.
- [17] Wilson, V. (2013). Formalized curiosity: Reflecting on the librarian practitionerresearcher. *Evidence Based Library and Information Practice*, 8(1), 111–117.