This is the Pre-Published Version.
This chapter has been accepted for the forthcoming book titled "Secondary Research Methods in the Built Environment" to be published by Routledge: Taylor & Francis Group. Submitted on 30 September 2019; Accepted on 19 May 2020.

CHAPTER (X)

APPLYING SCIENCE MAPPING IN BUILT ENVIRONMENT RESEARCH

Amos Darko and Albert P.C. Chan

Department of Building and Real Estate, The Hong Kong Polytechnic University

(amos.darko@connect.polyu.hk; albert.chan@polyu.edu.hk)

SUMMARY

Science mapping (SM) is an effective and useful methodology for studying and understanding the structural and dynamic features of a scientific domain through constructing, analysing and visualising bibliometric networks. Hence, using the method has acquired substantial attention in many fields, including the built environment (BE). Research in BE involves many diverse domains, subjects or topics. Objective mapping and understanding of the knowledge in these domains warrants SM. This chapter provides an introduction to the topic of applying SM in BE research. Essentially, it contributes to addressing the broader question of 'how to use SM in BE'. A tutorial is given that demonstrates in a step-by-step manner how three software, VOSviewer, CiteSpace and Gephi, can be applied together to conduct robust/deep SM-based research. This chapter could help researchers and other interested stakeholders undertake quality research using SM.

Keywords: Bibliometrics; Science mapping; Built environment; Built environment research; VOSviewer; CiteSpace; Gephi.

Introduction

BE research

Research in the built environment (BE) encompasses behavioural, affective and cognitive components, and a systematic process of investigation that increases knowledge by answering an unanswered question or by solving an unsolved problem (Amaratunga et al., 2002). It is

This is an Accepted Manuscript of a book chapter published by Routledge in Secondary Research Methods in the Built Environment on March 11, 2021, available online: http://www.routledge.com/9780367429874.

basically based on two inquiry paradigms. Positivism uses quantitative/experimental methods to test hypotheses. Interpretivism uses qualitative/naturalistic methods to produce hypotheses. While Amaratunga et al. (2002) discuss research in BE involving the two inquiry paradigms in detail, these paradigms have led to two key research types in BE, quantitative research and qualitative research. The former focuses upon numbers representing concepts or opinions, although the latter focuses upon observations and words for expressing reality. BE embodies various domains, subjects or topics, each of which is advanced via building a unique body of knowledge through quantitative and qualitative research. Mapping and understanding these large bodies of knowledge in an *objective* manner needs science mapping (SM) methodology.

Therefore, in recent years, SM has gained wide attention in BE, where researchers have used it to analyse various bibliometric networks (e.g., keyword co-occurrence networks). The main benefit of SM of a BE research domain is in achieving good understanding of how the domain is structured and of how it dynamically evolves. Specifically, SM is useful for detecting the major areas of the research in the domain, intellectual milestones in developing core specialties, evolutionary stages of key specialties involved and the dynamics of transitions from one specialty to another (Chen, 2017). All this knowledge could inform research, policy and practice decisions.

What is SM?

Bibliometrics provides techniques for quantitatively assessing research outputs (Cobo et al., 2011a). It has two main pillars: performance analysis and SM (Noyons et al., 1999). Whereas performance analysis assesses the activity, and the impact of this activity, of scientific *actors* (e.g., researchers) based upon bibliographic data (Noyons et al., 1999), SM displays the dynamic and structural features of scientific research (Börner et al., 2003). This chapter deals with SM. SM is an effective methodology for finding "representations of the intellectual connections within the dynamically changing system of scientific knowledge" (Small, 1997, p. 275) by creating, analysing and visualising bibliometric networks. These networks show how scientific specialties, fields, disciplines and authors or articles are conceptually, socially and intellectually structured and related to one another as illustrated by their relative locations and physical proximity (Small 1999; Cobo et al., 2011b). This is like how geographic maps display the relations of physical or political features of the Earth (Small, 1999). SM focuses on monitoring and delimiting research areas of scientific realms to ascertain their (cognitive) structures, their evolutions as well as the core actors within (Noyons et al., 1999).

Focus of this chapter

This chapter provides an introduction to the topic of applying SM in BE research. Essentially, it contributes to addressing the broader question of 'how to use SM in BE', by addressing the following key questions: (1) what is SM? (2) why is it useful for BE research? (3) how can it be done effectively? (4) which bibliographic databases can be used for SM in BE? (5) what is the acceptable sample size? (6) which SM analysis types and units can be used? (7) what are the commonly used software and what combinations of them can be implemented to perform quality SM-based research? The chapter gives a brief tutorial that shows how three software, VOSviewer, CiteSpace and Gephi, can be applied together to conduct robust/deep SM-based research.

Having described SM and its usefulness for BE research earlier, a survey of attempts to apply SM in BE field, with focus on the research domains addressed, is presented next, followed by discussions on issues in using SM in this field as regards data sources, sample size, analysis types and units and software. The tutorial upon the cooperative use of VOSviewer, CiteSpace and Gephi is then given, followed by conclusions of this chapter.

Historical survey

In this section, a brief overview of some of the attempts to apply SM in BE research is given. It is identified that applying SM in BE research is a relatively recent development. One of the first attempts was by Jide et al. (2015), who presented a SM analysis of construction worker's occupational mobility literature. SM has been applied to analyse construction and demolition (C&D) waste research (Liu et al., 2017; Chen et al., 2018; Jin et al., 2019a). There have been several SM-based studies in BE field focusing on the building information modelling (BIM) research area (Zhao, 2017; Li et al., 2017; He et al., 2017; Oraee et al., 2017; Chen and Man, 2018; Hosseini et al., 2018; Jin et al., 2019b; Chihib et al., 2019; Saka and Chan, 2019a, b). While these studies analysed the research on application of BIM to BE in general, others employed SM to analyse research on the role of BIM in specific BE domains. For example, Yin et al. (2019) undertook a SM analysis of research on BIM for offsite construction. In fact, using SM in BE so far has focused predominantly on the BIM literature. However, limited attention has been paid to integrating BIM with other digital technologies in BE issues. One of the few attempts is by Wang et al. (2019), who presented a bibliometric analysis of the research on integrating BIM and GIS in sustainable BE. Other BE research domains that have seen application of SM include green building (Darko et al., 2019), embodied energy of buildings (Zeng and Chini, 2017), value management (Ekanayake et al., 2019), mental health (Nwaogu et al., 2019) and lean construction (He and Wang, 2015). We find that there exists scope for further use of SM in these domains.

SM data sources for BE

A SM analysis begins with relevant bibliographic data retrieval. Today, there are many online bibliographic databases wherein scientific documents together with their citations are stored. These bibliographic data sources enable the searching and retrieval of information concerning most scientific fields (Cobo et al., 2011b). Nonetheless, choosing an appropriate source for data retrieval (one that contains data that could offer answers to the questions to be explored) is important (Börner et al., 2003). Chen (2017) and Cobo et al. (2011b) advocate that the most commonly used bibliographic databases include Web of Science (WoS), Scopus, Google Scholar and PubMed. Results of analysis of 20 selected SM applications in BE (Table 1) reinforce this where WoS and Scopus have dominated current SM-based research program in BE, with major reasons being:

- WoS is the most authoritative database for studying literature in many fields because it contains the most prestigious and important journals of influence in the world.
- Scopus has a wider scientific publications coverage and more recent publications than other databases, such as WoS.

It is logical that PubMed is not popular within this field because it is principally a biomedical database. Although Google Scholar can be used in this field, "downloading large datasets from Google Scholar is difficult and a dump of the entire dataset is not available" (Cobo et al., 2011b, p. 1383). This may be a reason for BE researchers to avoid using Google Scholar for SM. Another database that can be employed is CNKI (Chen and Man, 2018).

Funding or grants data and patent data can also be used for SM, but that is beyond the scope of this chapter, thus databases for these are not considered.

Table 1. Selected SM applications in BE.

Study	Research domain	Data sources	Sample size[a]	Timespan	Document types	Analysis types	Analysis units	Software
Jide et al. (2015)	Construction worker's occupational mobility	WoS	190	1986 to 2014	Articles	Co-citation; co-occurrence; citation burst; clustering	Document co-citation (also known as cited references); author co-citation (also known as cited authors);	CiteSpace

							keywords	
Liu et al. (2017)	C&D waste	WoS	857	2000 to 2016	Articles; reviews; proceedings papers; editorials	Co-authorship; co-citation; clustering	Institutions; journal co-citation (also known as cited sources); document co-citation	CiteSpace
Chen et al. (2018)	C&D waste	WoS	261	2006 to 2018	Articles	Co-citation; co-occurrence; citation burst; clustering	Document co-citation; keywords	CiteSpace
Jin et al. (2019a)	C&D waste	Scopus	370	2009 to 2018	Articles	Direct citation; co-occurrence; co-authorship	Journals; keywords; authors; documents; countries	VOSviewer
Zhao (2017)	BIM	WoS	614	2005 to 2016	Articles	Co-citation; direct citation; co-occurrence; co-authorship	Countries; institutions; subject categories; keywords; journal co-citation; author co-citation; document co-citation; clustering	CiteSpace
Li et al. (2017)	BIM	WoS	1,874	2004 to 2015	Articles; reviews; proceedings papers	Co-occurrence; co-citation; clustering; citation burst	Keywords; document co-citation; documents	CiteSpace
He et al. (2017)	BIM	WoS; Scopus	126	2007 to 2015	Articles	Co-occurrence; citation burst; clustering	Keywords	CiteSpace
Oraee et al. (2017)	BIM	Scopus	1,031	2006 to 2016	Articles; reviews	Direct citation; co-occurrence	Documents; keywords; journals	VOSviewer; Gephi
Chen and Man (2018)	BIM	WoS; CNKI	8,897	2003 to 2017	Articles	Co-authorship; co-occurrence	Authors; keywords	VOSviewer
Jin et al. (2019b)	BIM	Scopus	276	2008 to 2018	Articles	Direct citation; co-occurrence	Journals; documents; keywords	VOSviewer
Saka and Chan (2019a)	BIM	Scopus	93	2010 to 2018	All	Co-authorship; co-occurrence; co-citation	Keywords; authors; author co-citation; document co-citation	VOSviewer
Saka and Chan (2019b)	BIM	WoS	914	2006 to 2017	Articles	Co-authorship; co-citation; co-occurrence	Keywords; authors; author co-citation; document co-citation	VOSviewer; CiteSpace
Chihib et al. (2019)	BIM	Scopus	4,307	2003 to 2018	All	Co-authorship; co-occurrence; clustering	Countries; keywords	VOSviewer
Wang et al. (2019)	BIM-GIS integration in sustainable BE	WoS	76	2008 to 2018	Articles; proceedings papers	Co-occurrence; co-authorship	Keywords; authors	VOSviewer

Wuni et al. (2019)	Green building	Scopus	1,147	1992 to 2018	Articles	Direct citation; co-occurrence; bibliographic coupling; co-authorship	Journals; keywords; documents; countries; authors	VOSviewer
Nwaogu et al. (2019)	Mental health	WoS; Scopus	145	1974 to 2018	Articles	Co-citation; co-authorship; co-occurrence; clustering; bibliographic coupling; citation burst	Author co-citation; document co-citation; keywords; authors; institutions; countries; documents	CiteSpace
He and Wang (2015)	Lean construction	WoS; Scopus	621	1995 to 2014	Articles	Co-authorship; co-occurrence	Countries; keywords	CiteSpace
Ekanayake et al. (2019)	Value management	WoS	1,139	1990 to 2017	Articles; reviews; proceedings papers	Co-citation; clustering; co-occurrence; citation burst	Document co-citation; keywords; documents	CiteSpace
Cristino et al. (2018)	Energy efficiency in buildings	Scopus	513	1980 to 2016	Articles	Clustering	Keywords	VOSviewer
Zeng and Chini (2017)	Embodied energy of buildings	WoS	398	1996 to 2015	Articles; reviews	Co-occurrence; clustering; citation burst	Keywords	CiteSpace; Gephi

Note: [a] Average sample size = summation of sample sizes (23,849) divided by 20 = 1,192.45.

Sample size

A key essence of SM is to be *comprehensive* in the number of documents to be analysed. It is widely known that SM is a quantitative method proposed to overcome certain limitations of manual literature analysis, one of which relates to the number of documents that can be analysed (Yalcinkaya and Singh, 2015). It is therefore interesting that, as shown in Table 1, some SM applications in BE still involved sample sizes of, for example, less than 200 documents, which could still be analysed manually. This could be due to the lack of standards guiding the determination of SM sample size adequacy in the field. Based on the average sample size in Table 1, we recommend that a sample of 1,000 or more documents may be considered acceptable or adequate for SM; while below 1,000 may require the researcher to gather more data (some ways to do this are discussed later in this section). The robustness of the literature sample is directly related to the robustness of the results.

SM can be employed to analyse "huge amounts of data" (Börner et al., 2003, p. 209) that may not be possible to analyse manually because SM data analysis is computer aided. Though there is no specific definition of what "huge amounts of data" constitutes, looking at data of most SM examples in bibliometrics field, from where SM originates, (e.g., 36,000 documents in Small (1999); and 25,242 in van Eck and Waltman (2014a)), it is sound to assume that this refers to datasets in the thousands of documents. It would be useful for BE researchers to follow this practice in conducting SM-based research. However, if datasets in the hundreds of documents have to be used, then we advise that such datasets should consist of 500 to 999 documents. Anything below 500 documents can be deemed a weak or unacceptable sample size for SM.

While we appreciate that the sample size will depend on the amount of publications available in the particular research domain, we advise that this should not be used as a reason to compromise or undermine the comprehensiveness and robustness SM can offer in secondary research. There are several ways to improve the sample size:

- Using a comprehensive list of keywords for the literature search. An example of this will be demonstrated later in the "Cooperative use of VOSviewer, CiteSpace and Gephi: a brief tutorial" section.
- Employing the citation expansion method to expand the original dataset by merging it with dataset of documents within its citation record (Li et al., 2017).
- Using a comprehensive timespan (e.g., all years to present) for the literature search.
- Including multiple document types (e.g., journal and conference articles).
- Employing multiple bibliographic databases (e.g., WoS and Scopus).

In any case, data pre-processing is necessary for fixing errors, such as irrelevant documents, and thus improving data quality, which will impact the SM results quality. Cobo et al. (2011b) discuss SM data pre-processing in detail.

If all possible ways are employed and yet an adequate sample size, 1,000 or more documents, or 500 to 999 documents, cannot be reached, then the question is: why conduct SM when the body of literature in the particular research domain is still young, knowing that having a huge amount of dataset is critical for robust SM?

SM analysis types and units

There are several SM analysis types for identifying the dynamic evolution and knowledge structure of a domain through constructing, analysing and visualising bibliometric networks. In VOSviewer software, for instance, these include co-authorship, citation, co-occurrence, co-citation, bibliographic coupling and clustering analyses which along with their application are presented in Table 2. Extra ones like citation burst analysis can be found in software such as CiteSpace.

Table 2. VOSviewer-based SM analysis types and their application.

Analysis types	Application
Co-authorship	Identifying influential researchers, institutions or countries based on collaborations
Co-occurrence	Identifying major research interests, areas or topics based on keywords
Citation	Identifying influential journals, researchers, institutions, countries or articles based on direct citations
Bibliographic coupling	Identifying influential journals, researchers, institutions, countries or articles based on shared references
Co-citation	Identifying influential journals, researchers or articles based on co-citations
Clustering	Identifying groups of related topics, researchers, institutions, countries, journals or articles

In BE, the most commonly used or most important analysis types include co-occurrence, co-authorship, clustering, co-citation, citation burst and citation analyses – with the most commonly used analysis units including keywords, journals, authors, documents, institutions, document co-citation or cited references and author co-citation or cited authors (Table 1). While bibliographic coupling can also be used, it has yet to attract the level of attention the other analysis types have attracted. Additionally, journal co-citation or cited sources as an analysis unit has yet to receive much attention. Because of the word limitation, we refer to van Eck and Waltman (2014a) and Börner et al. (2003) for discussions of SM analysis types and units.

SM software tools

Various software have been developed to conduct SM analysis, including:

- CiteSpace (http://cluster.cis.drexel.edu/~cchen/citespace/; Chen, 2014).
- VOSviewer (https://www.vosviewer.com; van Eck and Waltman, 2019).
- Sci2 (https://sci2.cns.iu.edu/user/index.php; Sci2 Team, 2009).
- BibExcel (https://homepage.univie.ac.at/juan.gorraiz/bibexcel/; Persson et al., 2009).
- Ucinet (https://sites.google.com/site/ucinetsoftware/home, Borgatti et al., 2002).
- SciMAT (https://sci2s.ugr.es/scimat/; Cobo et al., 2012).

We aim not to give an overview of available SM software tools, but rather to introduce/direct the reader (especially those who are new to SM) to some of the tools. Cobo et al. (2011b) and van Eck and Waltman (2014a) have already provided good overviews of software tools for SM. Their overviews involve other tools not mentioned herein.

Some tools (e.g., BibExcel and Ucinet) are developed specifically for bibliographic data processing, after which the results must be imported into visualisation software like Gephi (https://gephi.org; Bastian et al., 2009) and Pajek (http://pajek.imfm.si; Batagelj and Mrvar, 1998) before the network can be visualised and analysed. Note that because Gephi and Pajek are specifically developed for network visualisation and analysis, they have no functionality for data processing. However, tools such as CiteSpace, VOSviewer, Sci2 and SciMAT can perform both data processing and network visualisation and analysis. Tools possess different capabilities and strengths and present complementary features. Therefore, Cobo et al. (2011b, p. 1383) recommended "to take their synergies to perform a complete SM analysis." This has not been sufficiently addressed in BE. While CiteSpace, VOSviewer and Gephi have been the most popular or most important tools in BE domain, only limited attempts have been made toward using multiple tools (e.g., VOSviewer+CiteSpace (Saka and Chan, 2019b), CiteSpace+Gephi (Zeng and Chini, 2017) and VOSviewer+Gephi (Oraee et al., 2017)) in one study (Table 1).

Much of the popularity gained by CiteSpace, VOSviewer and Gephi within this field could be accredited to some of their individual strengths. CiteSpace can build and dynamically visualise bibliometric networks to show how the studied domain has evolved over time and provide users with a wider range of visualisation and analysis options than other software, such as VOSviewer. A key strength of VOSviewer lies in it being easy to use with special attention paid to the graphical presentation of the bibliometric networks; it can also handle large networks. Unlike other software (e.g., CitNetExplorer (https://www.citnetexplorer.nl; van Eck and Waltman, 2014b)) that can visualise and analyse only one bibliometric network type, i.e. citation networks of publications, Gephi can visualise and analyse all network types.

In the next section, we present a tutorial that shows how VOSviewer, CiteSpace and Gephi can be combined for robust and deep SM analysis. Although other possible synergies among the different software exist, this (involving three software) is one of the highest synergies implemented in BE field so far. In fact, a complete, thorough and deep SM analysis of any domain requires cooperative use of different software (Cobo et al., 2011b). Such an approach affords the extraction of all the useful knowledge and diverse perspectives about the domain hidden behind the dataset.

Cooperative use of VOSviewer, CiteSpace and Gephi: a brief tutorial

This tutorial aims to facilitate and promote the idea of using different software in a combined manner to perform complete, thorough and deep SM in BE. To this end, we demonstrate an example (Darko et al., 2020) of how VOSviewer, CiteSpace and Gephi can be used cooperatively through the following steps:

1. Determine the research domain you wish to study/analyse. For example, in this tutorial we study the 'AI in the AEC industry' research domain.
2. Establish a list of keywords for the literature search. This step is important because the comprehensiveness and robustness of the search keywords list affect those of the whole SM work from sample size to final results. As an example, results (sample sizes) obtained from searches in Scopus with three different possible keywords lists for the research domain considered in this tutorial are shown in Table 3. All the searches were run at the same time on 26 January 2020. Figures 1 and 2 provide examples of how the final results/networks might also be affected in terms of comprehensiveness and robustness. Note that the two figures were created through the same process. For instance, they were both created using VOSviewer wherein "author keywords" was the analysis unit, fractional counting method was used and the minimum number of occurrences a keyword should have in order to be included in the network was set to five for both figures. The only difference is the dataset.

Table 3. Keywords and literature search results.

Query string	Number of keywords	Search results (sample size)
"Artificial intelligence" AND "Architecture, Engineering and Construction industry" AND (LIMIT-TO (SUBJAREA , "ENGI")) AND (LIMIT-TO (DOCTYPE, "ar"))	2	58 documents
"Artificial intelligence" OR "Machine intelligence" OR "Machine learning" OR "Expert systems" OR "Genetic algorithms" OR "Artificial neural networks" OR "Artificial general intelligence" OR "Case-based reasoning" AND "Architecture, Engineering and Construction industry" AND (LIMIT-TO (SUBJAREA , "ENGI")) AND (LIMIT-TO (DOCTYPE , "ar"))	9	102 documents

"Artificial intelligence" OR "Machine intelligence" OR "Machine learning" OR "Expert systems" OR "Genetic algorithms" OR "Neural networks" OR "Case-based reasoning" OR "Data mining" OR "fuzzy logic" OR "Fuzzy sets" OR "Expert systems" OR "Robotics" OR "Knowledge-based systems" OR "Support vector machines" OR "Deep learning" OR "Artificial general intelligence" OR "Computational intelligence" AND "Construction industry" OR "Civil engineering" OR "Structural engineering" OR "Architectural engineering" OR "Construction engineering" OR "Construction management" OR "Construction engineering and management" AND (LIMIT-TO (SUBJAREA, "ENGI")) AND (LIMIT-TO (DOCTYPE, "ar"))	24	53,924 documents

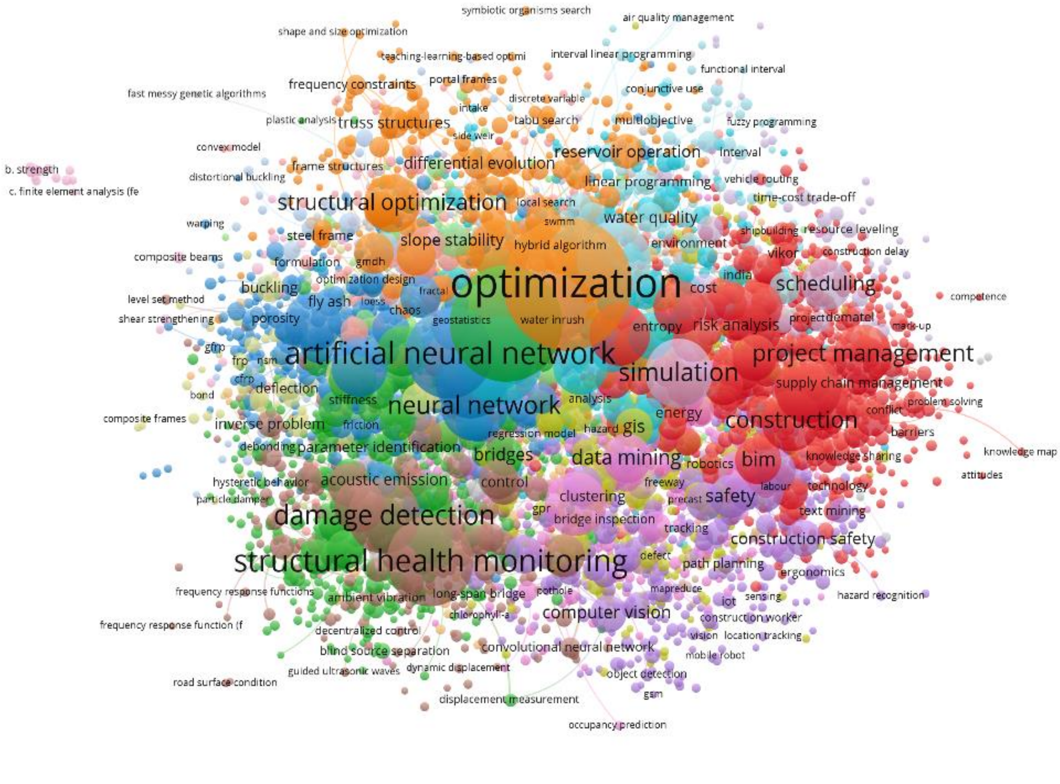

Figure 1. VOSviewer visualisation of a 4,297 author keywords co-occurrence network based on the dataset of 53,924 documents.

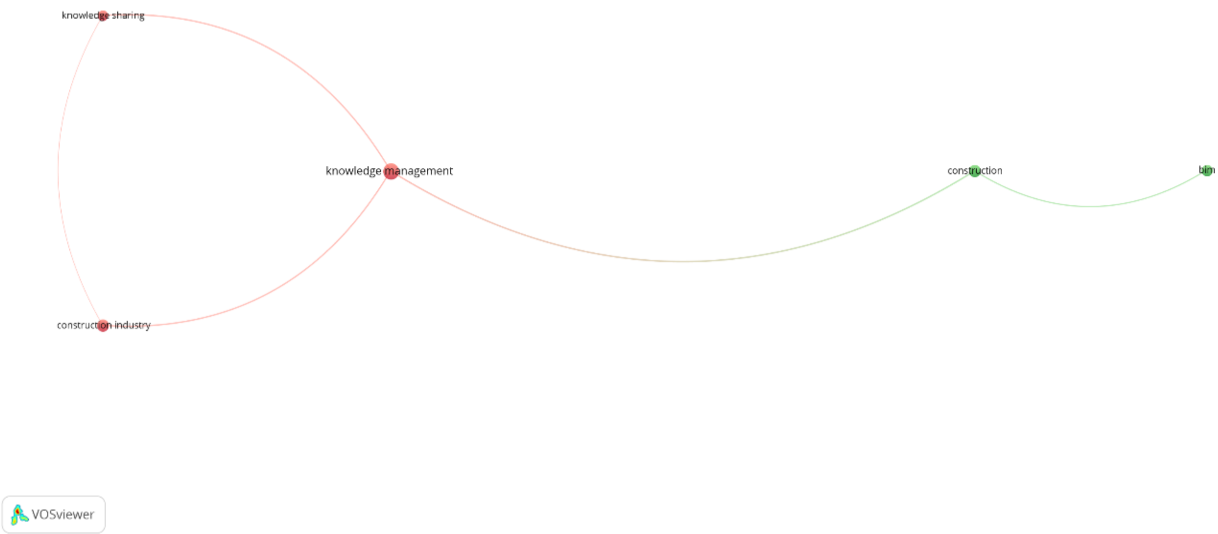

Figure 2. VOSviewer visualisation of a five author keywords co-occurrence network based on the dataset of 102 documents.

3. Choose the bibliographic database(s) to be used and collect/download the bibliographic data from there using the keywords list established in Step 2. For example, in this tutorial, we use Scopus and download the data in comma-separated values (CSV) format. In Scopus, bibliographic data for only 2,000 or less documents can be downloaded at a time. So, if you have more than 2,000 documents, then downloading can occur in batches based upon, for example, journals.
4. Once the dataset is ready, download and launch VOSviewer, CiteSpace and Gephi. They are all freely downloadable (a reason for their wide use) from https://www.vosviewer.com; http://cluster.cis.drexel.edu/~cchen/citespace/; and https://gephi.org, respectively.
5. Use VOSviewer to build a keyword co-occurrence network (figures 1 and 2 are examples). However, a drawback of VOSviewer is that it does not have data pre-processing modules so there may be duplicate items (keywords in this case). Though at a point, it offers the possibility to manually remove items from the network, we chose not to remove any from

figures 1 and 2, to emphasise the importance of data pre-processing. Even if you choose to remove items manually, this can be a tedious task especially with numerous keywords and missing any item before moving to the next step requires you to restart the whole analysis process. Of course, there are duplicate items such as BIM, building information modeling and building information modelling in figure 1. Duplicate items can be merged using a thesaurus file, but this is still a manual, time-consuming process. Before the analysis in VOSviewer, de-duplication could be done using CoPalRed software (Cobo et al., 2011b).

6. Next, use CiteSpace to perform cluster analysis with document co-citation as the analysis unit (other analysis units can also be used). Because of the space limitation, we are unable to show more figures and tables herein; more can be found in Darko et al. (2020). While VOSviewer can also be used for cluster analysis (for instance, the different colours in figure 1 represent different clusters of the keywords), "CiteSpace provides more precise ways to identify groupings, or clusters, using the clustering function" (Chen, 2014, p. 14).
7. Use CiteSpace again to perform citation burst analysis to detect how the most active areas or features of the research domain have evolved over time. Of course, a key strength of CiteSpace is its ability to detect time-based evolutions of features of the domain by automatically slicing the dataset into different time periods for the analysis. Citation burst analysis can be performed over many analysis units, including keywords, documents, authors and institutions. In each case, active areas and/or emerging trends can be revealed. An example of citation burst analysis over keywords is provided in Darko et al. (2020).
8. Now, close CiteSpace and go back to VOSviewer. Utilize VOSviewer to conduct a direct citation analysis of sources (or journals). Save the generated network in graph modeling language (GML) format and then submit this to Gephi for further analysis. The benefit of this cooperative approach is that it leads to higher quality visualisation of the network, as Gephi affords further options and capabilities. For example, it allows us to choose between

directed, undirected and mixed network types. Moreover, sometimes, VOSviewer does not show the names of some nodes in the network. This is evident in figure 1, for example. We know this is a large network, but sometimes even with relatively small networks, this problem still occurs and can be fixed with Gephi. Besides, VOSviewer does not always present the full names of nodes (Jin et al., 2019b), another problem Gephi can solve. VOSviewer also could only show lower-case letters where "IT", for example, may appear as "it" (Jin et al., 2019c). Gephi can solve this issue by allowing us to edit the names of nodes. It also allows further manipulation of the network to achieve the desired shape and appearance. Direct citation analysis can be performed for documents, authors, institutions or countries as well.

9. Employ VOSviewer to perform co-authorship analysis of institutions and of countries, enabling to explore collaborations among research institutions and among countries. Save the generated networks in GML format and then submit them to Gephi for further analysis. Co-authorship analysis can be done for authors too.

10. After completing the above steps, note that the software assist us to produce results and networks based on the research dataset but cannot interpret and draw conclusions from the results. Thus, the final step is for you to interpret and draw conclusions from the results. The quality and depth of the interpretation and conclusions depend on your knowledge and experience. The key point is to extract useful knowledge that could well inform further research, policy and practice decisions.

For SM-based studies in BE where this joint use of VOSviewer, CiteSpace and Gephi was employed, we refer to Darko et al. (2020), Darko et al. (2019) and Hosseini et al. (2018). This tutorial gives an example on generic steps to follow to conduct SM through combined use of these three software but does not show how to use the software. This can be found in manuals for VOSviewer (van Eck and Waltman, 2019), CiteSpace (Chen, 2014) and Gephi (Bastian et al., 2009). While we focused only on some synergies between VOSviewer, CiteSpace and Gephi, other synergies among these and/or other software (Cobo et al., 2011a, b) could also be explored in BE. Hence, we propose a generic six-steps procedure (Figure 3) that could be adopted/adapted for any SM-based research (especially those conducted through joint use of different software).

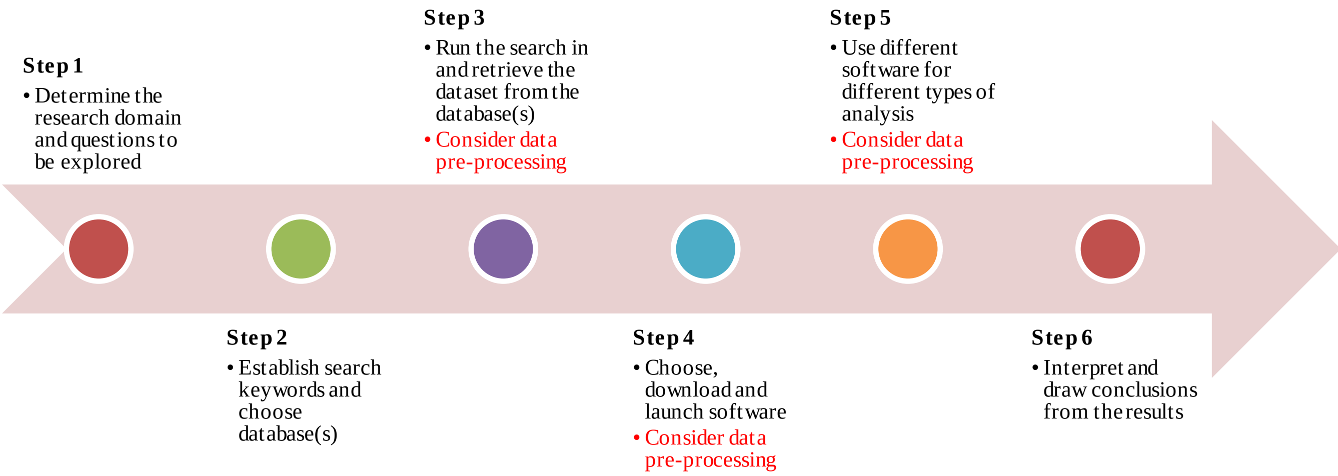

Figure 3. Generic six-steps procedure for SM-based research.

Conclusions

BE domains, subjects or topics are advanced by building unique bodies of knowledge through quantitative and qualitative research. Mapping and understanding these large bodies of knowledge in an *objective* manner requires SM. This chapter provided an introduction to the topic of applying SM in BE research. Essentially, it contributed to addressing the question of 'how to use SM in BE'. Our two major recommendations are to: (1) use *large* sample size (1,000 or more documents) for SM in BE because SM quality and robustness depends on sample size robustness; and (2) widely adopt the combined use of various software to conduct deep, thorough and complete SM works in BE. Such an approach aids to not only extract all the useful knowledge from the dataset, but also to cover for each software's weaknesses by combining their strengths. We hope that this chapter will help promote widespread use of this integrated approach and help researchers and other interested stakeholders conduct quality research using SM.

References

Amaratunga, D., Baldry, D., Sarshar, M., and Newton, R. (2002). Quantitative and qualitative research in the built environment: application of "mixed" research approach. *Work Study, 51(1), 17-31*.

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Third International AAAI Conference on Weblogs and Social Media*.

Batagelj, V., and Mrvar, A. (1998). Pajek-program for large network analysis. *Connections*, 21(2), 47-57.

Borgatti, S.P., Everett, M.G., and Freeman, L.C. (2002). Ucinet for windows: Software for social network analysis. Harvard, MA: Analytic Technologies.

Börner, K., Chen, C., and Boyack, K.W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37(1), 179-255.

Chen, C. (2014). CiteSpace Manual. http://cluster.ischool.drexel.edu/~cchen/citespace/CiteSpaceManual.pdf.

Chen, C. (2017). Science mapping: a systematic review of the literature. *Journal of Data and Information Science*, 2(2), 1-40.

Chen, J., and Man, Q. (2018). A critical review on BIM research process and future trends at home and abroad. *ICCREM 2018: Innovative Technology and Intelligent Construction* (pp. 104-112).

Chen, J., Su, Y., Si, H., and Chen, J. (2018). Managerial areas of construction and demolition waste: A scientometric review. *International Journal of Environmental Research and Public Health*, 15(11), 2350.

Chihib, M., Salmerón-Manzano, E., Novas, N., and Manzano-Agugliaro, F. (2019). Bibliometric maps of BIM and BIM in universities: A comparative analysis. *Sustainability*, 11(16), 4398.

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., and Herrera, F. (2011a). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, 5(1), 146-166.

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., and Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology*, 63(8), 1609-1630.

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., and Herrera, F. (2011b). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382-1402.

Cristino, T.M., Neto, A.F., and Costa, A.F.B. (2018). Energy efficiency in buildings: analysis of scientific literature and identification of data analysis techniques from a bibliometric study. *Scientometrics*, 114(3), 1275-1326.

Darko, A., Chan, A.P.C., Adabre, M.A., Edwards, D.J., Hosseini, M.R., and Ameyaw, E.E. (2020). Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. *Automation in Construction*, 112, 103081.

Darko, A., Chan, A.P.C., Huo, X., and Owusu-Manu, D.G. (2019). A scientometric analysis and visualization of global green building research. *Building and Environment*, 149, 501-511.

Ekanayake, E.M.A.C., Shen, G., and Kumaraswamy, M.M. (2019). Mapping the knowledge domains of value management: a bibliometric approach. *Engineering, Construction and Architectural Management*, 26(3), 99-514.

He, Q. H., and Wang, G. (2015). Hotspots evolution and frontier analysis of lean construction research—integrated scientometric analysis using the Web of Science and Scopus databases. *Frontiers of Engineering Management*, 2(2), 141-147.

He, Q., Wang, G., Luo, L., Shi, Q., Xie, J., and Meng, X. (2017). Mapping the managerial areas of Building Information Modeling (BIM) using scientometric analysis. *International Journal of Project Management*, 35(4), 670-685.

Hosseini, M.R., Maghrebi, M., Akbarnezhad, A., Martek, I., and Arashpour, M. (2018). Analysis of citation networks in building information modeling research. *Journal of Construction Engineering and Management*, 144(8), 04018064.

Jide, S., Qi, N., and Liangfa, S. (2015). Analysis on current situation and development trend of construction worker's occupational mobility. *Open Construction and Building Technology Journal*, 9, 303-310.

Jin, R., Yuan, H., and Chen, Q. (2019a). Science mapping approach to assisting the review of construction and demolition waste management research published between 2009 and 2018. *Resources, Conservation and Recycling*, 140, 175-188.

Jin, R., Zou, P.X., Piroozfar, P., Wood, H., Yang, Y., Yan, L., and Han, Y. (2019c). A science mapping approach based review of construction safety research. *Safety Science*, 113, 285-297.

Jin, R., Zou, Y., Gidado, K., Ashton, P., and Painting, N. (2019b). Scientometric analysis of BIM-based research in construction engineering and management. *Engineering, Construction and Architectural Management*, 26(8), 1750-1776.

Li, X., Wu, P., Shen, G.Q., Wang, X., and Teng, Y. (2017). Mapping the knowledge domains of Building Information Modeling (BIM): A bibliometric approach. *Automation in Construction*, 84, 195-206.

Liu, Y., Sun, T., and Yang, L. (2017). Evaluating the performance and intellectual structure of construction and demolition waste research during 2000–2016. *Environmental Science and Pollution Research*, 24(23), 19259-19266.

Noyons, E., Moed, H., and Van Raan, A. (1999). Integrating research performance analysis and science mapping. *Scientometrics*, 46(3), 591-604.

Nwaogu, J.M., Chan, A.P.C., Hon, C.K., and Darko, A. (2019). Review of global mental health research in the construction industry. *Engineering, Construction and Architectural Management*, https://doi.org/10.1108/ECAM-02-2019-0114.

Oraee, M., Hosseini, M.R., Papadonikolaki, E., Palliyaguru, R., and Arashpour, M. (2017). Collaboration in BIM-based construction networks: A bibliometric-qualitative literature review. *International Journal of Project Management*, 35(7), 1288-1301.

Persson, O., Danell, R., and Schneider, J.W. (2009). How to use Bibexcel for various types of bibliometric analysis. *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday*, 5, 9-24.

Saka, A.B., and Chan, D.W.M. (2019a). A scientometric review and metasynthesis of building information modelling (BIM) research in Africa. *Buildings*, 9(4), 85.

Saka, A.B., and Chan, D.W.M. (2019b). A global taxonomic review and analysis of the development of BIM research between 2006 and 2017. *Construction Innovation*, 19(3), 465-490.

Sci2 Team. (2009). Science of Science (Sci2) Tool. https://sci2.cns.iu.edu/user/index.php.

Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics*, 38(2), 275-293.

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American society for Information Science*, 50(9), 799-813.

van Eck, N.J., and Waltman, L. (2014a). Visualizing bibliometric networks. *Measuring scholarly impact* (pp. 285-320). Springer, Cham.

van Eck, N.J., and Waltman, L. (2014b). CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics*, 8(4), 802-823.

van Eck, N.J., and Waltman, L. (2019). VOSviewer Manual. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.11.pdf.

Wang, H., Pan, Y., and Luo, X. (2019). Integration of BIM and GIS in sustainable built environment: A review and bibliometric analysis. *Automation in Construction*, 103, 41-52.

Wuni, I.Y., Shen, G.Q., and Osei-Kyei, R. (2019). Scientometric review of global research trends on green buildings in construction journals from 1992 to 2018. *Energy and Buildings*, 190, 69-85.

Yalcinkaya, M., and Singh, V. (2015). Patterns and trends in building information modeling (BIM) research: A latent semantic analysis. *Automation in Construction*, 59, 68-80.

Yin, X., Liu, H., Chen, Y., and Al-Hussein, M. (2019). Building information modelling for off-site construction: Review and future directions. *Automation in Construction*, 101, 72-91.

Zeng, R., and Chini, A. (2017). A review of research on embodied energy of buildings using bibliometric analysis. *Energy and Buildings*, 155, 172-184.

Zhao, X. (2017). A scientometric review of global BIM research: Analysis and visualization. *Automation in Construction*, 80, 37-47.