ANALYZING BIM TOPICS AND CLUSTERS THROUGH TEN YEARS OF SCIENTIFIC PUBLICATIONS

Clément Lemaire
ETS Montréal, Canada

Louis Rivest, Professor,
ETS Montréal, Canada
E-mail: louis.rivest@etsmtl.ca

Conrad Boton
ETS Montréal, Canada
E-mail: conrad.boton@etsmtl.ca

Christophe Danjou
Polytechnique Montréal, Canada
E-mail: christophe.danjou@polymtl.ca

Christian Braesch
Université Savoie Mont-Blanc – SYMME, Annecy, France
E-mail : christian.braesch@univ-smb.fr

Felix Nyffenegger
HSR, Rapperswil, Switzerland
E-mail: felix.nyffenegger@hsr.ch

SUMMARY: There has been considerable interest in Building Information Modeling (BIM) research and development during the last decade. BIM has established itself as a field of research in (and beyond) scientific communities interested in information technologies in construction. Interestingly, the contours of BIM as a scientific field are still not clearly identified. Several studies have recently tried to analyze different aspects of this issue, without providing a systematic and comprehensive methodological approach to accurately define the major themes and clusters of the BIM domain. This paper uses a systematic literature review approach to map the BIM research themes and clusters over ten years of scientific publications. 1244 articles published in peer-reviewed journals between 2007 and 2016 were selected and the associated metadata analyzed in order to highlight co-occurrences in author's keywords. It appears that a few "big players" dominate the keywords, while most of the keywords used by authors are much less cited. Seven core clusters are identified using modularity optimization techniques: industry foundation classes, information technology, facility management, building, collaboration, computer aided design, and laser scanning.

1. INTRODUCTION

Research in Building Information Modeling (BIM) is currently at a crossroads, as researchers and the industry have established enough perspective to challenge themselves about the real value of the approach and about how the dedicated research works are aligned with the real needs of the Architecture, Engineering and Construction (AEC) industry. Indeed, while considerable effort has been dedicated to developing different aspects of BIM, it is still difficult to state how Building Information Modeling can be defined as a research domain.

Many literature review works have been proposed recently, exploring the technological review of issues encountered with BIM tools and standards (Cerovsek, 2011), interrogating the usefulness of BIM for construction safety (Zhou, Whyte and Sacks, 2012), evaluating the maintenance and deconstruction processes of existing buildings (Volk, Stengel and Schultmann, 2014), assessing the existing techniques for data acquisition in "as-built BIM" (Hichri et al., 2013), for facility management (Yalcinkaya and Singh, 2014) and for sustainable design (Dowsett and Harty, 2014). While these research works have the traditional narrative approach of literature review, other studies have used Systematic Literature Review (SLR) and mapping study approaches in order to be more comprehensive and to better reflect the complexity and the interdisciplinary nature of BIM research. Recent BIM-related SLR and mapping study works examine the key managerial issues in BIM adoption (He et al., 2017), the main components of BIM-based Construction Networks (Oraee et al, 2017a), BIM-related collaboration (Oraee et al., 2017b), and more generally the knowledge domains of BIM (Li et al., 2017; Santos, Costa and Grilo, 2017).

These varied works illustrate how SLR and mapping study approaches can be helpful in understanding BIM as a research domain. Moreover, they can be seen as important milestones towards a good understanding of BIM as a research field. The work presented in this paper complements the above-cited works and aims at identifying the main BIM themes and clusters, without focusing on a specific pre-determined aspect. It uses a mapping method to clearly identify the main topics and clusters in BIM over the last ten years. The main objective of the research is to provide a good understanding of the main areas addressed to date in BIM literature. Unlike most of the previous works, it only analyzes the author's keywords in order to characterize the domain, as it assumes that the authors are best qualified to summarize their work with keywords.

The paper is organized into four main parts. The next section presents a state-of-the-art of literature review and bibliometric mapping as research approaches. It also presents the recent BIM-related SLR and mapping studies. Section 3 introduces the methodology employed to address our research aims, including a straightforward presentation of its steps. The findings of our study are presented in section 4, along with the keywords' analysis, the co-occurrence network analysis, and the clusters analysis. These results are discussed in section 5, both in terms of the recent related works and in their implication for further studies.

2. RELATED WORKS

2.1 Literature review as a research approach

Luciano defines a research literature review as a "selection of available documents […] on the topic […] written from a particular standpoint to fulfill certain aims or express certain views on the nature of the topic and how it is to be investigated, and the effective evaluation of these documents in relation to the research being proposed" (Luciano, 2011). In fact, this definition discusses the traditional vision of literature reviews that mostly refers to narrative reviews. This type of review presents certain limitations, as its preparation depends on "implicit, idiosyncratic methods of data collection and interpretation" (Mulrow, Cook and Davidoff, 1997). Mulrow demonstrates that the lack of a transparent and understandable method can cause biases in an author's conclusions (Mulrow, 1994).

Another interesting definition can be found in Fink's work, as he defines research literature review as a "systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars, and practitioners" (Fink, 2014). In a sense, this definition suggests that all literature reviews should be systematic (Booth, Papaioannou and Sutton, 2012). The main differences between several reviews will come from their degree of systematicity (Booth, Papaioannou and Sutton, 2012). According to Rousseau et al. (2008), "systematic means comprehensive accumulation, transparent

analysis, and reflective interpretation of all empirical studies pertinent to a specific question. Reliance upon any sampling or subset of the literature risks misrepresenting its diversity in findings, outcomes, methods, and frames of reference". Rather than searching for a part of the truth, reviewing in a systematic way can reveal the whole truth on a research question, according to Mulrow (1994). A quite complete definition of systematic literature review can be found in Carney and Geddes (2002): "syntheses of primary research studies that use (and describe) specific, explicit and therefore reproducible methodological strategies to identify, assemble, critically appraise and synthesize all relevant issues on a specific topic". As a research methodology, Systematic Literature Review (SLR) helps researchers ensure an "unbiased, rigorous and auditable" review (Kitchenham, Budgen and Brereton, 2010).

It is important to distinguish mapping studies from SLRs. Indeed, mapping studies have the same methodological basis as SLRs, but focus on more global topics, while SLRs address specific research questions (Petticrew and Roberts, 2006). Also, mapping studies do not directly discuss findings but are based on the concept that published articles indirectly represent activity related to the findings. To sum up, a mapping study is a review that does not attempt to find results, but instead endeavors to find linkages (Cooper, 2016).

Mapping studies are often based on the use of bibliometrics, and it is primordial not to confound them with systematic reviews. In 1987, Broadus reviewed the evolution of the definition of the term bibliometrics (Broadus, 1987). Formerly called statistical bibliography, he highlighted the definition given by the ALA Glossary of Library and Information Science (Young, 1983): "statistical methods in the analysis of a body of literature to reveal the historical development of subject fields and patterns of authorship, publication, and use." Indeed, science mapping is also defined as a spatial representation of how disciplines, fields, specialties, and individual documents or authors are related to one another (Small, 1999). Several techniques have been implemented over time. Cobo et al. (2011) referenced the major bibliometric techniques and their units of analysis, and also specified the type of relationship between them (Table 1) (Cobo et al., 2011). The research work presented in this paper belongs to the 'co-word' bibliometric technique, as it focuses on keyword co-occurrences.

Table 1: Bibliometric techniques taxonomy (from Cobo et al., 2011)

Bibliometric technique		Unit of analysis used	Kind of relation
Bibliographic coupling	Author	Author's oeuvres	Common references among author's oeuvres
	Document	Document	Common references among documents
	Journal	Journal's oeuvres	Common references among journal's oeuvres
Co-author	Author	Author's name	Authors' co-occurrence
	Country	Country from affiliation	Countries' co-occurrence
	Institution	Institution from affiliation	Institutions' co-occurrence
Co-citation	Author	Author's reference	Co-cited author
	Document	Reference	Co-cited documents
	Journal	Journal's reference	Co-cited journal
Co-word		Keyword, or term extracted from title, abstract or document's body	Terms' co-occurrence

2.2 Literature review works in the field of Building Information Modeling

Various literature review works have been published recently in the field of BIM. Some have used the narrative approach, while some others are based on Systematic Literature Review and Mapping Studies.

2.2.1 Narrative review of the BIM literature

BIM studies are often characterized as multidisciplinary, involving different fields and disciplines. They address, through different angles, diverse and specific aspects of BIM. Cerovsek (2011) provided a more technological review of issues encountered with BIM tools and standards and proposed a methodological framework to improve them. A year later, Zhou et al. (2012) explored relationships between construction safety and digital design, showing how BIM can be useful to improve safety throughout design and construction stages (4D CAD, Virtual Reality, Sensors…). In 2014, Volk et al. (2014) concentrated their review on BIM for existing buildings,

related maintenance and deconstruction processes, revealing major challenges in automating data capture and BIM creation (without pre-existing BIM). Hichri et al. (2013) focused on a review of "As-Built BIM" approaches by identifying existing techniques of acquisition and their limitations, especially in semantically enriching point clouds. Yalcinkaya and Singh (2014) analyzed BIM literature in the field of Facility Management (FM) and showed the potential added values of BIM for FM. Dowsett and Harty (2014) studied the literature review of methodologies to evaluate the benefits of BIM for sustainable design. Wong and Zhou (2015) focused on green BIM literature by categorizing its shortcomings and outlining some directions for future research. In 2015, Cheng and Lu (2015) reviewed the efforts made worldwide by the public sector for BIM adoption, showing how the USA and the UK are the most mature countries in terms of BIM adoption.

2.2.2 BIM-related recent Systematic Literature Review and Mapping Studies

Unlike previous traditional literature reviews published on the subject, some papers have contributed to a more global analysis of BIM and not only to some of its aspects. He et al. (2017) presented their literature review in a systematic way, analyzing the authors' keyword co-occurrence network, identifying the 'bursting' keywords and the main clusters. However, this study was limited to the managerial issues in BIM adoption. Li et al. (2017) conducted a similar analysis but focused on the knowledge domains of BIM. They highlighted the document co-citations network and the references with strong citation bursts. In 2017, Oraee et al. used a mixed method, both quantitative and qualitative, to review available studies on BIM-based construction networks. After focusing on determining a co-occurrence network, they manually reviewed papers focused on collaboration, and categorized them using a "Collaboration Pentagon" (Oraee et al., 2017a). They also categorized papers by proposing and using a "Collaborative Working Model" (Oraee et al., 2017b). A manual categorization process was used by Santos et al. (2017), which addressed a bibliometric analysis of BIM-related articles published between 2005 and 2015. They limited their analysis sample to the 100 most-cited articles in the BIM field and those articles published in journals with impact factors higher than 1.0. They only analyzed 381 papers, but their study showed an exponential increase of BIM research in the last 5 years. Moreover, their research highlighted the most-cited subjects (BIM adoption, laser scanning, and literature reviews), the new trends (BIM tool development, semantic BIM and ontology, laser scanning) and the gaps in the literature (BIM training and education, BIM-GIS and quantity take-off). More recently, Abdal Noor and Yi (2018) presented a "meta-analysis" of BIM literature in construction and transportation in order to map the use of BIM and highlight the research gaps. Their research examined more than 3200 publications and focused on railway stations' construction phase. The findings suggest that the main research subjects over the years include education, scheduling, BIM integration and adoption, and interoperability.

Table 2 presents a summary of the recent BIM-related Systematic Literature Review and Mapping Studies mentioned above, facilitating their comparison according to several criteria.

We shall analyze more precisely the three publications that explore BIM-specific literature based on authors' keyword co-occurrence (highlighted in italics in Table 2). First, the data collection method differs among these three papers. Oraee et al. concentrated on the terms from BIM literature with the highest co-occurrences (based on the papers' titles and abstracts) and created a keywords co-occurrence network (based on the authors' keywords) for BIM studies associated with 'collaboration' (Oraee et al., 2017a). These studies represent only 271 papers and have a strong collaboration aspect that may not reveal a global view of BIM. Likewise, He et al. (2017) restricted their dataset to 308 publications which corresponds to a query related to managerial areas. Li et al. (2017) is the only paper to date that analyzes an authors' keywords co-occurrence network of BIM-related publications without a specific focus, hence with a global perspective.

Keyword normalization is a key aspect in the study of keyword co-occurrence, and publications can be compared on this basis. Indeed, results based on the same database can differ depending on the normalization rules. He et al. (2017) and Li et al. (2017) both specify the normalization process by grouping keywords by themes, and by normalizing different keywords to one meaning. While this is interesting and potentially very helpful, it creates the risk of losing the original meaning intended by the authors and can considerably affect the results (e.g. 'Interoperability' and 'Data exchange' are merged, just as '3D' and 'nD modeling' are merged). It may also create a bias based on how the authors regroup the keywords.

Table 2: Comparison of BIM-related recent SLR and Mapping Studies

Publication	Data collection						Data analysis	
	Source	Type of papers retained	Number of papers retained (sample size)	Field of study	Keywords' enquiry	Period	Analyzed elements	Type of treatment, Software, Normalization
Mapping the knowledge domains of Building Information Modeling (BIM): A bibliometric approach (Li et al. 2017)	*Web of Science*	*- Articles* *- Reviews* *- Proceeding papers*	*- 938 (core data set)* *- 1874 (expanded data set)*	*Knowledge domains of BIM*	*"BIM"* *"Building Information Model*"*	*2004 - 2015*	*- Authors' keywords cooccurence* *- Documents' cocitation* *- Citations' burst detection* *- Clusters' identification*	*Automatic, CiteSpace, Yes*
Bibliometric analysis and review of Building Information Modelling literature published between 2005 and 2015 (Santos et al. 2017)	- Web of Science - Selected journals	- Articles published in journals with impact factor > 1.0 - Top 100 most	- 381 (360 without BIM reviews)	Bibliometrics and literature review of BIM	"BIM" OR "Building Information model" OR "Building Information	2004 - 2016	- Paper's categorization based on authors' category proposition	Manual, N/A, N/A
Collaboration in BIM-based construction networks: A bibliometric-qualitative literature review (Oraee et al. 2017a)	*Scopus*	*- Articles published in journals* *- Review studies published in journals*	*- 1031 (Stage 1)* *- 271 (Stage 2)* *- 62 (Stage 3)*	*Collaboration in BIM-based Construction Networks (BbCNs)*	*"Building Information Modeling" OR "Building Information Modelling"*	*2006 - 2016*	*- Title/Abstract keywords cooccurence (Stage 1).* *- Authors' keywords cooccurence (Stage 2).* *- Paper's categorization based on collaboration pentagon (Stage 3).* *- Documents' citation.*	*Automatic and manual, Gephi and VOSviewer, NOT PRECISED*
Mapping the managerial areas of Building Information Modeling (BIM) using scientometric analysis (He et al. 2017)	*- Web of Science* *- Scopus*	*Articles published in peer-reviewed journals*	*- 308 (Stage 1)* *- 126 (Stage 2)*	*Managerial areas of BIM (MA-BIM)*	*("BIM" OR "building information modeling" OR "building information*	*2007 - 2015*	*- Authors' keywords cooccurence* *- Authors' keyword's burst detection* *- Clusters' identification*	*Automatic, CiteSpace, Yes*
Where the Gaps Lie: Ten Years of Research into Collaboration on BIM-Enabled Construction Projects (Oraee et al. 2017b)	Selected international ICT-oriented journals in the construction industry	Articles published in journals	- 1034 (Stage 1 : BIM-related) - 208 (Stage 2 : Collaboration) - 62 (Stage 3 : Collaboration	Collaboration on BIM-Enabled Construction Projects	"BIM" (Stage 1) "collaboration" OR "integration" OR "IPD" OR "collaborative'	2006 - 2016	- Paper's categorization based on Collaborative Working Model (CWM)	Manual, N/A, N/A
Eleven years of ARCOM : Bibliometric mapping of studies published between 2005 and 2016 (Hosseini et al. 2017)	Scopus	Only ARCOM publications	- 1505	Mapping study of ARCOM publications	N/A	2005 - 2016	- Authors' keywords & Title/Abstract keywords cooccurence - Sources' cocitation	Automatic, Gephi and VOSviewer, NOT PRECISED

In addition, the papers cited above mainly analyze the frequency of keywords, but do not focus on the links between them. However, examining keyword co-occurrence can highlight clusters, or communities of keywords, as well as the keywords that play the role of 'gateways' between communities. Oraee et al. provide a good example illustrating the connections between keywords (Oraee et al., 2017a). Their results show, amongst other things, that 'communication', while not a top keyword in frequency, is linked to many other keywords, such as 'Building information modeling' and 'Information technology'. However, that work is limited to collaboration aspects of BIM. Hence, work remains to be done in order to reveal all the BIM communities of keywords (beyond collaboration-related research) by filtering with nodes' degrees or edges' weights.

In this study, we chose to focus on authors' keywords. It is assumed that the authors are best qualified to summarize their work with their selection of keywords. Globally, keywords can be defined in different ways, such as by the authors, by a journal's prescribed categories, or by automatic tags generated at the proof stage (Hartley and Kostoff, 2003). Some keywords can also be generated by users, which is the case in some collaborative platforms such as CiteULike (Kipp, 2005). In such a context, the author keywords tend to achieve higher recall than collaborative tags (Lu and Kipp, 2014). There is no consensus on the best procedure for creating keywords. Indeed, it has been said that "authors are not very good at generating descriptors for their research" (Hartley and Kostoff, 2003). Moreover, "different authors can disagree over their choice of keywords for a particular paper" (Hartley and Kostoff, 2003). For these reasons and to avoid an out-of-hand increase of author-generated keywords, some 'standard lists' of 'closed sets' of keywords are created and used by journal editors. While very practical to avoid "letting everyone come up with their own" (Hartley and Kostoff, 2003), such a practice tends to decrease authors' ability to summarize their intent. To overcome this limitation, it seems appropriate to rely on the authors' keywords when conducting a keywords-based systematic literature review, since one can advocate that: "who but the author could possibly do a better job of picking the keywords for an article?", as concluded an editor cited by Hartley and Kostoff (2003).

3. METHODOLOGY

To ensure a sound and effective review of BIM-related scientific publications over the past ten years, a systematic review methodology has been used. It consists of three main complementary phases: data collection, data normalization and data analysis.

3.1 Data collection

We used Scopus to identify all the relevant publications on the subject. All searches were realized by October 2017 and do not consider subsequent updates of the Scopus database. No limitation in the language of the publications retained was applied. Indeed, as the study focuses on author's keywords, articles in foreign languages were only retained if their keywords are in English and correspond to the topic.

The refinement steps that were required in order to focus exclusively on the relevant publications are presented in this section.

3.1.1 Step 1: Type and date of retained publications

Thanks to its numerous search filters, Scopus allows to find the type of publications corresponding to the defined criteria. The research is focused on peer-reviewed journal articles published between 01.01.2007 and 31.12.2016, representing a decade of publications. 2017 has not been considered because the study began that year and not all of the 2017 publications had been published.

All publications with keywords "BIM" or "Building Information Model*" in the title, abstract and keywords (T/A/K) were first identified. The use of the "*" wildcard allowed us to consider all possible terminations of the word "Model", which is fundamental in the field of BIM, whose definitions can be multiple (Building Information Model, Building Information Modeling or Building Information Modelling). Then, we decided to add the term "Building Information Management" to our request so as to maximize our chances of finding articles related to our study. Using the Boolean "OR" in our query allowed us to capture a very large number of documents (4875) composed of

either "BIM" OR "Building Information Model *" OR "Building Information Management". A second filtration step was required to remove irrelevant documents.

3.1.2 Step 2: Second filtration step

The second filtration step was used to filter out all publications that were not related to the construction domain. The Scopus' "Subject Areas" filter was used, limiting the documents to those in domains related to our study, as indicated by the screen capture in Figure 1.

Figure 1: Subject areas maintained in our query

Areas such as medicine or chemistry were removed. On the other hand, some other areas potentially relating to construction-related BIM were kept, such as social and environmental sciences.

At the end of this stage, the number of remaining publications was 1976.

3.2 Data verification and normalization

After these filtration steps, the metadata (Title, Authors, Publication Date, Author Keywords etc.) was exported from Scopus to Zotero via the RIS file format. Next, all the publications were manually checked to improve data consistency. Even with the application of the subject areas filter, many publications that are not related to BIM in construction still appeared in the database. In fact, BIM is an acronym often used in other fields (e.g. BIM is the name of a protein). We manually removed all publications that were not relevant. It also appeared that some of the publications retained had no keywords. In order to include only authors' keywords in the database, we compared the original article version to the keywords from Scopus. We thus manually added or deleted keywords in order to keep only the authors' keywords.

After this verification step, the number of publications decreased to 1244, the final number of publications retained for this study. Afterwards, all the keywords were extracted from Zotero to Excel using an SQL query. To obtain consistent results and avoid unnecessary duplications, it was necessary to normalize the keywords. Some normalization rules were defined and applied to the entire group of keywords. Lower case letters and singulars were retained for almost every keyword[1]. Also, American English was opted for instead of Britain English[2] and special characters (such as "-", "&", "/", "_") were replaced by "and" or a space, depending on the sense. We did not

---

[1] Brand names, country name and unknown acronyms remain in capital letters.
[2] E.g. "Modelling" becomes "modeling" and "Visualisation" becomes "visualization"

regroup synonyms or terms with a strong similarity so as to avoid betraying the authors' intent. We normalized to their long form acronym which had only one definition (e.g. "RFID" became "radio frequency identification"). For those keywords combining a long form and an acronym, only the long form given by the authors was kept (e.g. "Building Information Model (BIM)" would be "building information model"). For those keywords providing an acronym in the short form only but with a potentially different meaning, only the acronym was kept. Therefore, "BIM" stayed "BIM" rather than becoming "building information model", "building information modeling" or "building information management". Moreover, acronyms were also preserved in the short form when they were followed by other terms specifying their meaning (e.g. "BIM Adoption" would be listed as "BIM adoption").

3.3 Data analysis

In Graph theory, graphs are defined as sets of vertices (also called nodes) connected by edges (also called links or arcs) (Newman, 2003). In our study, nodes correspond to normalized keywords. Edges are the links between those keywords, and hence express the co-occurrence of authors' keywords in the same publication. A weight is associated to each edge. Each edge's weight represents the number of times a pair of keywords appears in a publication (e.g. 'industry foundation classes' and 'interoperability' are linked with an edge which has a weight of 16, meaning that those keywords co-occur in 16 different publications). The degree of a node characterizes the number of edges linked to it (Opsahl, Agneessens and Skvoretz, 2010). If we want to consider an edge's weight in a node's degree, we discuss the weighted node's degree, which is the addition of all of the edges' weights linked to a node (Kapoor, Sharma and Srivastava, 2013).

This study uses an open source network exploration and manipulation software named Gephi. To visualize a network of keywords in Gephi, several steps are necessary (Bastian, Heymann & Jacomy, 2009). First, the normalized list of keywords is reintroduced into the SQL database to replace the original keywords. Each keyword is identified by an ID number, so we can track the publication from which it stems. Thanks to this step, an SQL query allows us to produce two Excel files that we can import into Gephi. One is for the nodes and regroups the normalized forms of the keywords as well as the number of times they appear (keyword count), as shown in Table 3. The other is for the edges that link (through their IDs) the keywords from the same publication, while counting the number of times the same pair of keywords appear. This is how each edge's weight is obtained.

After importing those files into Gephi, we use the layout algorithm called ForceAtlas2 proposed in the software. It is defined as a "force directed layout" that "simulates a physical system in order to spatialize a network" (Jacomy et al., 2014). By activating ForceAtlas2, nodes and edges respectively repulse and attract each other until converging to a balanced state. The configuration thus obtained allows a better visualization of the network and helps to interpret the data.

Modularity is also used to detect the communities of nodes (or clusters) in the network structure. Communities are then revealed using different colors. A community can be described as a group of densely connected nodes that has many edges between the members of a group, but only a few edges between the different groups. In our study, we therefore identify the major themes of BIM based on research publications and identify natural communities in this research domain.

Newman discusses different approaches to analyzing the topology of a network graph, such as graph partitioning and community structure methods (Newman, 2006). Community structure methods assume that the network of interest divides naturally into subgroups. If it is the experimenter's job to find those groups he or she could suggest the method of optimal modularity, referring to Guimerà and Amaral (2005) and Danon et al. (2005). The concept behind this approach is to quantify the assumption that the network of interest has a higher level of structure than a random network. The corresponding measure is called the modularity. The modularity is the number of edges falling into groups minus the expected number in an equivalent network with edges placed at random. The network can now be divided into groups such that the modularity is optimized. The software used to analyze the network is an implementation of optimal modularity attributed to Blondel et al. (2008).

Table 3: Most-used keywords in the collected publications

Item	Rank	Normalized keywords	Count
1	1	building information modeling	515
2	2	BIM	374
3	3	industry foundation classes	94
4	4	building information model	92
5	5	interoperability	49
6	6	information technology	48
7	7	facility management	47
8	8	construction	45
9	9	collaboration	31
10	10	construction management	30
11	11a	geographic information system	28
12	11b	ontology	28
13	13a	building	27
14	13b	computer aided design	27
15	15	sustainability	24
16	16a	laser scanning	22
17	16b	visualization	22
18	18	BIM technology	21
19	19a	case study	20
20	19b	construction industry	20
21	19c	integrated project delivery	20
22	22a	design	18
23	22b	energy efficiency	18
24	22c	lifecycle	18
25	22d	parametric design	18
26	22e	point cloud	18
27	22f	project management	18
28	28a	cloud computing	17
29	28b	simulation	17
30	30	information management	16

The final step is to modify the visual aspects settings: the nodes and edges' display sizes depends on their degree or their weight.

4. STUDY RESULTS

4.1 Keywords analysis

The 1244 papers retained for our study use a total of 3123 different authors' keywords. It is interesting to note that 2463 keywords are used only once, representing 78.9 % of all keywords. The total count of all the keywords is 6414, and the top 100 most-cited keywords account for 2415 of that number. Hence, 3.2 % (100/3123) of all keywords are responsible for 37.7 % (2415/6414) of the counts of keywords, an indication of the importance played by "major players". This is illustrated by the graph (Figure 2) of the count distribution. Its L-shape indicates that a few

keywords are much more common, while the majority only occur in a single paper, as indicated in another study (Nyffenegger, Rivest and Braesch, 2016). These results show that the global pool of keywords is mostly composed of "minor players" that have no impact in defining BIM as a research domain. Hence, keywords that are used only once can be set aside.

Table 3 lists (and ranks) the top 30 most-used keywords. 'building information modeling' ranks first and is used 515 times, while 'BIM' as an acronym ranks second. 'building information model' is used 92 times and hence ranks fourth. Not all authors used 'BIM' or one of its definitions in their keywords. Moreover, the edge between 'BIM' and 'building information modeling' has the greatest weight with 49, which means that both keywords were used in 49 different publications. In Table 3, each keyword has a distinct item number, but some keywords share the same rank. As an example, 'geographic information system' and 'ontology' share the 11th rank with a count of 28.

The third keyword in the list is 'industry foundation classes' (or IFC), while the fifth is 'interoperability', with counts of 94 and 49, respectively. There is a logical connection between these two keywords, since an IFC file format can be considered as the technological aspect of interoperability; both are clearly major subjects in the realm of BIM. Furthermore, it is noticeable that 'computer aided design', which also appears high in the ranking (15th with 27), relies directly on the IFC file format. 'interoperability' also refers to 'collaboration' (10th with a count of 31), which is a strong pillar of BIM. 'information technology' (6th with 48) and 'construction' (8th with 45) are two main words that define BIM itself. The keyword 'construction management' ranks 10th with 30. Interestingly, 'facility management' is 7th with a count of 47; it is therefore an important research focus even if it is only one of the referenced BIM uses. 'geographic information system' is in the 11th position, revealing the interest of the research community in the integration of geographic data in BIM models. The keywords 'laser scanning' (22) and 'point cloud' (18) together reach a count of 40, which indicates an important research subject. The keywords 'energy efficiency', 'sustainability' and 'lifecycle' also are keywords that could have environmental management in common and that reach high in the ranking. Another interesting finding is the position of 'case study' (19th with 20), as it suggests the importance of this type of research approach in the BIM field.
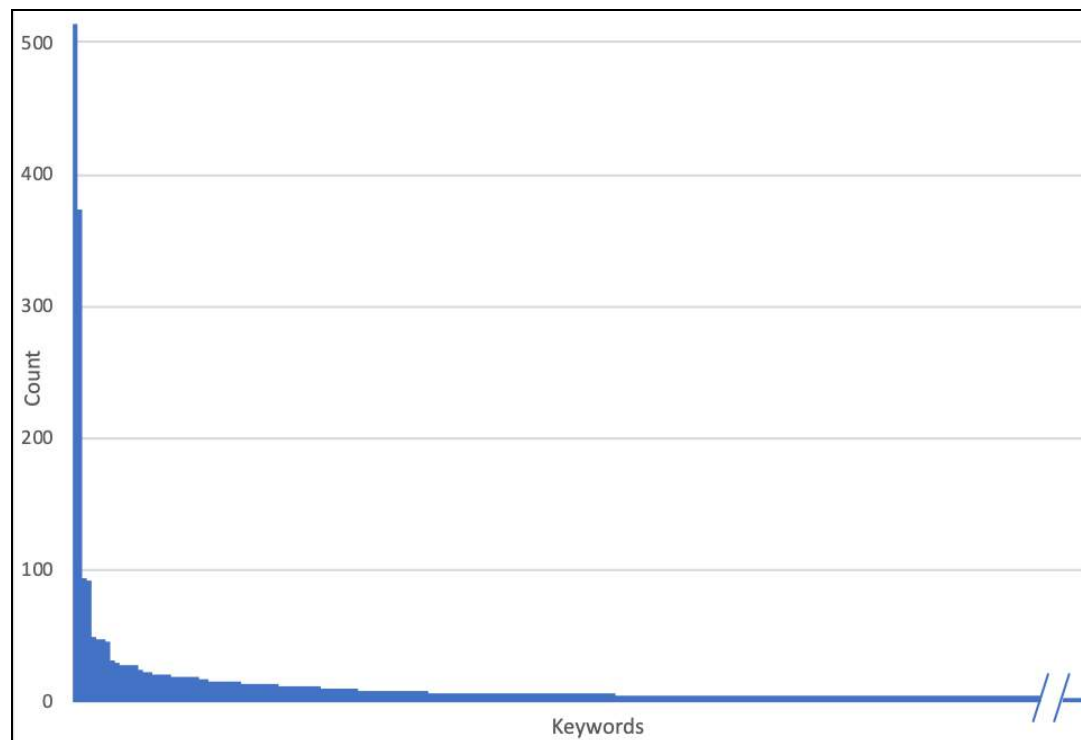

Figure 2: Counts' distribution of normalized keywords in the collected publications

The analysis of top-ranking keywords is significant, but it depicts BIM in a limited way. Analyzing the links between keywords may be even more revealing of BIM than a simple list of top-ranking keywords. This is explored in the following sections.

4.2 Keyword co-occurrence network analysis

Our complete network is composed of 3123 nodes and 13096 edges. On average, each node has a degree of 8.4. Another interesting parameter to focus on is the relation between nodes. 93.6 % of the edges have a weight of one, and the average weight is 1.12 per edge.

A first representation of the network is shown in Figure 3. Major players (BIM and building information modeling) can be characterized as "trees that hide the forest". The high number of nodes with a count of one also makes it difficult to interpret the network in its entirety.

Figure 3: Co-occurrence network of all normalized keywords in the collected publications

In order to lighten the graph and better visualize the data, we chose not to show the highest-ranking keywords such as 'BIM', and 'building information modeling'. We also removed 'building information model' and 'building information management'. Lastly, we removed all the keywords that are used only once. The result of this first step of filtering is shown in Figure 4. This new network is now composed of 656 nodes and 2989 edges. The visualization of nodes and edges depends on their characteristics. For nodes, the size of the font is directly linked with the degree. For edges, the thickness represents the weight.
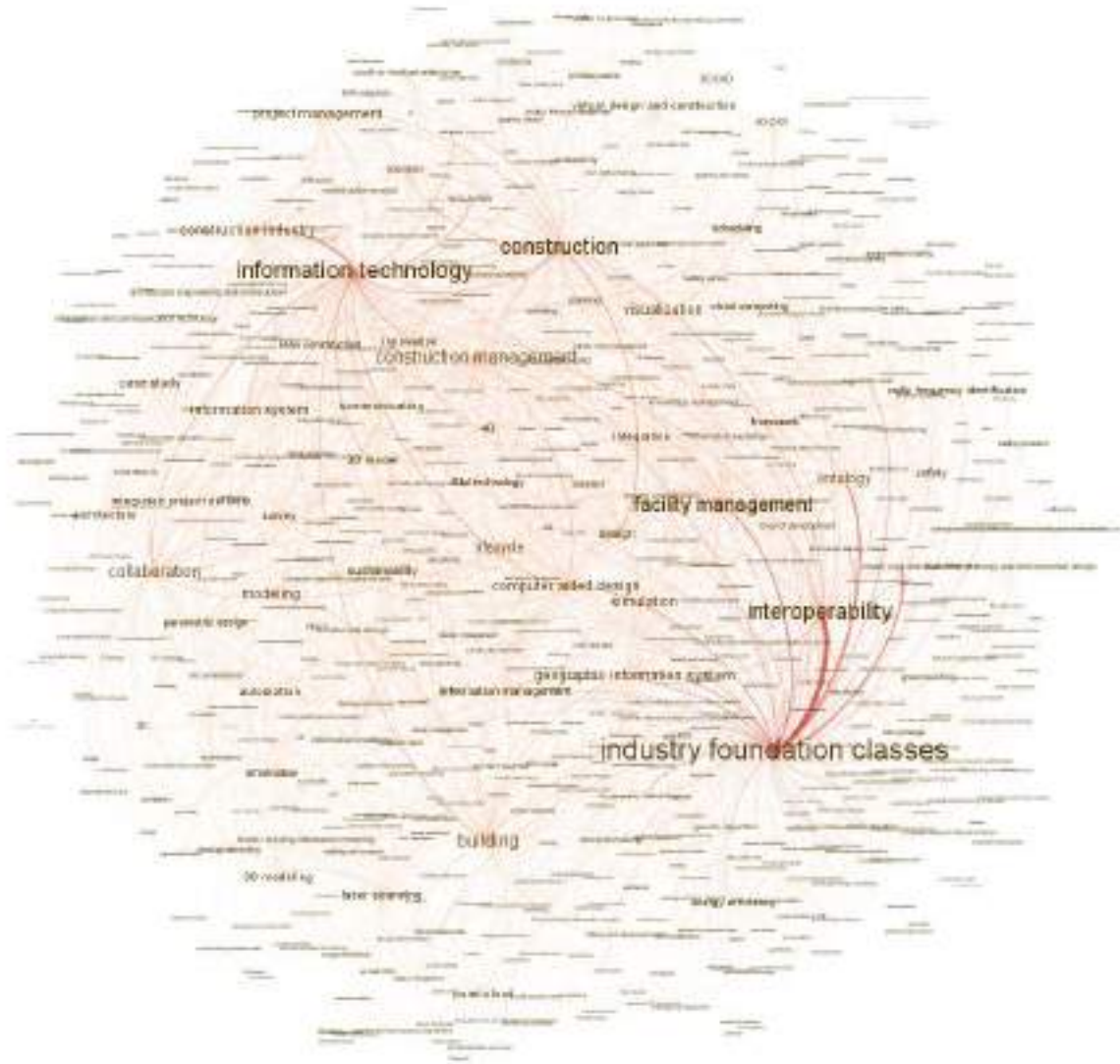

Figure 4: Keyword co-occurrence network after the first filtering step

A first interesting (but not surprising) finding is that 'industry foundation classes' and 'interoperability' are strongly linked. At 16, it is by far the highest weight of all the remaining edges. The next edge in the list is between 'information technology' and 'construction industry', with a weight of 8. Table 3 lists all the edges with a weight of 4 and more, characterized by the names of the nodes (keywords) that they link. There are 23 of them. The 56 edges with a weight of 3 are not listed for the sake of space.

Table 3: Pairs of keywords with highest co-occurrences (edge weights)

Item	Node 1	Node 2	Edge's weight
1	industry foundation classes	interoperability	16
2	information technology	construction industry	8
3	industry foundation classes	ontology	7
4	industry foundation classes	model view definition	7
5	industry foundation classes	facility management	5
6	information technology	MALAYSIA	5
7	construction	design	5
8	facility management	operation and maintenance	5
9	industry foundation classes	geographic information system	4
10	industry foundation classes	city geography markup language	4
11	industry foundation classes	computer aided design	4
12	information technology	computer aided design	4
13	information technology	construction management	4
14	construction management	computer aided design	4
15	construction	education	4
16	education	curriculum	4
17	industry foundation classes	information delivery manual	4
18	interoperability	model view definition	4
19	genetic algorithm	optimization	4
20	laser scanning	3D modeling	4
21	information technology	3D model	4
22	green building	leadership in energy and environmental design	4
23	process modeling	product modeling	4

Table 5 shows the 30 top-ranking keywords with the highest count and with the highest weighted degree. The count is the number of times each normalized keyword appears in our database. The degree of each normalized keyword is the number of keywords associated to it (i.e. the number of keywords that are also used in publications where it has been found). The weighted degree of a node considers the weights of the edges linked to it. This means that if a pair of keywords are used in five publications, the edge between them has a weight of five and it affects the weighted node degree accordingly. The difference between the degree and the weighted degree of a node is that the degree does not take into account an edge's weight, so that if a pair of keywords co-occurs in several publications, it will only count a single time in the degree. In Table 5, the keyword 'industry foundation classes' has the highest count, with 94, the highest degree, with 124, and the highest weighted degree, with 194. It naturally appears as the dominant node in Figure 4. The keyword 'information technology' has a degree of 93, and also appears as a major keyword in Figure 4.

Table 5: Top 30 keywords with the highest counts vs Top 30 keywords with the highest weighted degrees

Item	Rank	Normalized keywords	Count	Rank	Normalized keywords	Count	Degree	Weighted degree
1	1	industry foundation classes	94	1	industry foundation classes	94	124	194
2	2	interoperability	49	2	information technology	48	93	141
3	3	information technology	48	3	interoperability	49	83	122
4	4	facility management	47	4	construction	45	84	113
5	5	construction	45	5	facility management	47	71	95
6	6	collaboration	31	6	building	27	62	81
7	7	construction management	30	7	collaboration	31	59	80
8	8a	geographic information system	28	8	construction management	30	60	77
9	8b	ontology	28	9	ontology	28	46	68
10	10a	building	27	10	construction industry	20	48	66
11	10b	computer aided design	27	11	computer aided design	27	45	65
12	12	sustainability	24	12	geographic information system	28	48	63
13	13a	laser scanning	22	13	visualization	22	46	56
14	13b	visualization	22	14	simulation	17	47	50
15	15	BIM technology	21	15	modeling	14	44	49
16	16a	case study	20	16	laser scanning	22	36	48
17	16b	construction industry	20	17a	lifecycle	18	44	47
18	16c	integrated project delivery	20	17b	project management	18	40	47
19	19a	design	18	19	sustainability	24	38	46
20	19b	energy efficiency	18	20	information system	14	40	45
21	19c	lifecycle	18	21a	design	18	37	43
22	19d	parametric design	18	21b	information management	16	35	43
23	19e	point cloud	18	23	3D model	15	30	41
24	19f	project management	18	24	innovation	16	34	40
25	25a	cloud computing	17	25	point cloud	18	30	39
26	25b	simulation	17	26	integration	15	30	38
27	27a	information management	16	27a	3D modeling	14	33	37
28	27b	innovation	16	27b	case study	20	36	37
29	29a	3D model	15	29	communication	11	30	35
30	29b	integration	15	30	MALAYSIA	13	24	34

One could ask, what is the most significant indicator for a keyword amongst count, degree and weighted degree? There is a high count when a keyword is used often in publications, meaning it is important in the BIM research field. A high degree means that many other terms are associated with this keyword, which may indicate that it plays a central role in a domain. A high weighted degree is quite the same as a high degree, but it may also signify that the keyword is often used with some specific keywords (with a highly-weighted link).

In general, keywords with a high degree also have a high-weighted degree, as both are related. Also, the higher the count, the higher the degree and the weighted degree, as the keyword is more represented.

4.3 Keywords Clusters

In the previous section, the data was analyzed mainly based on keywords' counts and degrees. Another way of analyzing the data is to filter the graph with edges' weights. This helps to reveal meaningful communities (or clusters) of keywords. Depending on the filtering conditions, we can distinguish the **core cluster of BIM** and the **global cluster of BIM**. The core cluster is obtained by considering edges with a weight of 3 or more. The global cluster is obtained by considering edges with a weight of 2 or more.

4.3.1 BIM Core Cluster

The BIM core cluster is shown in Figure 5. The edges are filtered in order to keep only those with a weight of 3 or more, so that related keywords appear in pairs in at least three publications. The result of this filtering reveals the BIM core cluster composed of 43 nodes and 59 edges. Figure 5 shows the **core cluster of BIM as a research domain**. Some other small communities, isolated from this BIM core cluster, are also revealed by this filtering level but will be covered in section 4.3.2.

This graph size is a convenient starting point for analysis. We observe that this BIM core cluster is composed of six communities identified using modularity analysis tools from Gephi. These clusters are organized around their dominant keywords: 'industry foundation classes' in purple, 'information technology' in green, 'facility management' in blue, 'collaboration' in orange, 'computer aided design' in pink, and 'building' in dark grey (Figure 5). These keywords that dominate their communities are not exactly the same as the list of top keywords by count. They do, however, give a clear indication about how research work is organized within BIM as a research domain.
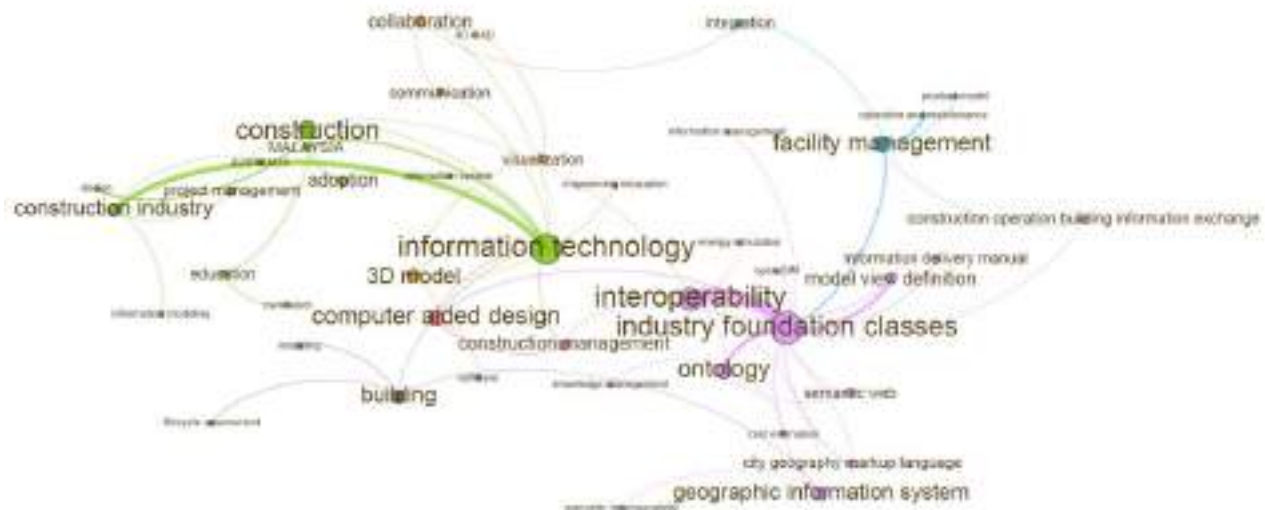

Figure 5: BIM core cluster obtained by keeping edges with a weight of 3 or more and colored by communities.

The **'industry foundation classes' community**, shown in the purple cluster in Figure 5, can be considered the dominant cluster since: 1) it groups the two keywords with the highest count 'industry foundation classes' (IFC) and 'interoperability' (Table 5); and 2) it includes the edge with the highest weight (Table 3). We could reasonably consider that these two keywords are very important to BIM as a research domain. It can also be observed that this cluster includes 'ontology', which is linked to 'knowledge management', 'semantic web' and 'cost estimation'. This community also includes 'model view definition' (MVD) and 'information delivery manual' (IDM). The association of these three concepts is not surprising since MVDs can be defined as subset of the IFC schema that documents the data structures adapted to a particular discipline, and a clear definition of the IDM is a necessary step for the creation

of new MVDs (East, Nisbet and Liebich, 2013). One can also observe the presence of the keywords 'geographic information system' and 'city geography markup language'.

The **'information technology' community**, shown in green, is structured around three main nodes: 'information technology', 'construction' and 'construction industry'. These last two keywords remain distinct because of the normalization rules that were used, but also are somehow similar. One can note that the keyword 'adoption' acts as an intermediate node between 'construction' and 'information technology'. The keyword 'education', which is linked with 'curriculum', also appears in this community. The keyword 'MALAYSIA' has also been used in a few publications and appears in this community.

In the top right of the graph, the **'facility management' community** is shown in blue. Its central node, 'facility management', is linked with 'operation and maintenance' and with 'product model'. This indicates that those two keywords are amongst those most associated with 'facility management' in the literature. This community is also connected with two others, which are 'collaboration' and 'industry foundation classes'. Interestingly, the 'facility management' community is linked to the 'collaboration' cluster via 'integration'.

On the top left of the graph, the **'collaboration' community** is shown in orange. Collaboration, the 6th keyword by count in Table 5, is an essential area in BIM. An important node of this community is '3D model'. Other meaningful nodes are 'communication', 'visualization', and '4D CAD'. This community is connected, via '3D model', to the communities 'information technology' and 'computer aided design'.

The **'computer aided design' community**, in pink, is a small one with only 3 nodes: 'computer aided design', 'construction management' and 'software' keywords. This community connects, via 'computer aided design', to the 'industry foundation classes' and 'information technology' communities. It is interesting to note that this community is linked with all of the other communities except 'facility management' and 'building'.

At the bottom left of the graph, the **'building' community** is shown in dark grey. It is a small community in the BIM core cluster with only 3 nodes: 'building', 'lifecycle assessment' and 'modeling'. This community is linked, via 'building', to two other communities: 'industry foundation classes' and 'information technology'.

4.3.2 BIM Global Cluster

The graph shown in Figure 5 above was designated the 'BIM core cluster', since it shows the core content of BIM as a research domain. The graph shown in Figure 6 corresponds to the **BIM Global cluster as a research domain**. It is consistent with the core cluster shown in Figure 5, but includes more details. It was obtained by filtering the edges with a weight of 2 or more. This filtering leaves a graph composed of 271 nodes and 79 edges; the BIM global cluster corresponds to the most important portion of this graph with 174 nodes and 295 edges. As mentioned earlier, some other isolated small clusters, or communities, are also revealed. We consider that the BIM global cluster is more complete. While describing the BIM core cluster helped to organize the major ideas, describing the BIM global cluster in a second step provides a more comprehensive understanding of BIM as a research domain.

This BIM global cluster brings out several distinct clusters, also identified with the modularity tools in Gephi. These communities are organized similarly to the BIM core cluster, around dominant keywords: 'industry foundation classes' in purple, 'information technology' in green, 'facility management' in blue, 'collaboration' in orange, 'computer aided design' in pink, 'building' in dark grey and 'laser scanning' in turquoise blue. These are the same communities as for the BIM core cluster, but have more nodes, except for the 'laser scanning' group. This 'laser scanning' community regroups two previously isolated communities, 'laser scanning' and 'point cloud', that are now part of the 'BIM global cluster' as one community.
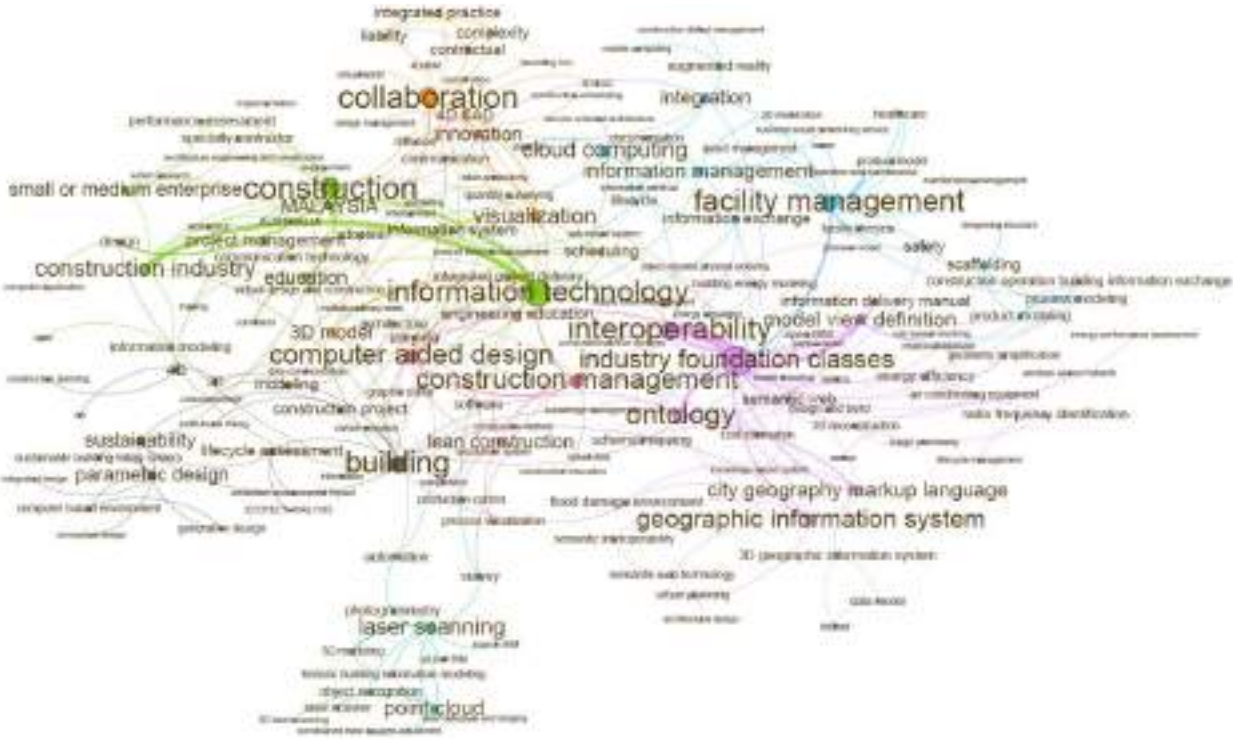
Figure 6: BIM global cluster obtained by keeping edges with a weight of 2 or more, colored by communities.

The **'industry foundation classes' community**, in purple (Figure 7), includes dominant keywords such as 'industry foundation classes' (IFC), 'interoperability', 'ontology' and 'geographic information system'. Amongst the information added at the global cluster level that was not visible from the BIM core cluster, we observe two keywords linked to 'ontology': 'schema mapping' and 'linked data'. The keyword 'geographic information system' is now linked to 'ontology', and to the newly appearing keywords 'urban planning', 'architectural design' and 'semantic web technology'. The keyword 'knowledge based system' now appears and is linked to 'semantic web'. The keywords 'image processing' and '3D reconstruction' now appear. The latter is linked to 'industry foundation classes'. A few energy-related keywords appear: 'energy efficiency' and 'energy performance assessment', 'air conditioning equipment', 'energy simulation' and 'building energy modeling'. Other keywords also appear, such as 'open BIM', 'rule-based checking', 'relational database', 'model checking' and 'WEB GL'. Interestingly, 'supply chain management' and 'process model' now appear, linked to 'interoperability'.

The **'information technology' community**, shown in Figure 8, is structured around three main nodes: 'information technology', 'construction' and 'construction industry'. In the BIM global cluster, new ideas now appear: 'small or medium enterprise', 'performance assessment' and 'specialty contractor'.

Interestingly, this community has a node 'supply chain', somewhat similar to the 'supply chain management' mentioned above for the IFC community. The keyword 'product lifecycle management' appears, connected to 'information technology', 'estimating' and 'quantity surveying'. 'scheduling' and 'planning' appear here, as well as 'virtual design and construction'.

Figure 7: Zooming in on the 'industry foundation classes' community from the BIM global cluster

Figure 8: Zooming in on the 'information technology' community from the BIM global cluster

The **'facility management' community** is shown in Figure 9. While this community counted just four nodes in the BIM core cluster, in the BIM global cluster this community is much more significant. The keyword 'information management', which was in the 'industry foundation classes' community, now belongs to this community, along with 'information retrieval', 'information exchange', 'documentation' and 'cloud computing'. The keyword 'augmented reality' now appears, linked to 'integration', 'mobile computing' and 'construction defect management'. Interestingly, we now find 'asset management' and 'maintenance management' along with 'operation and maintenance'. In addition, 'owner' and 'health care' now appear, linked to 'facility management'. We also find 'product model', 'product modeling' and 'process modeling'. Not surprisingly, we find the keywords 'safety', 'scaffolding', 'facility lifecycle' and 'temporary structure' in this community.

Figure 9: Zooming in on the 'facility management' community from the BIM global cluster

The **'building' community**, shown in Figure 10, was a very small community of the BIM core cluster with only three nodes: 'building', 'lifecycle assessment' and 'modeling'. Now, in the BIM global cluster, the 'building' community has gained importance. Keywords '3D', '4D', 'construction planning' and '5D' are connected to 'modeling'. 'sustainability' also is important with the presence of the keywords 'lifecycle assessment', 'sustainability', 'sustainable building rating systems', 'carbon emission', 'embodied environmental impact', and 'ECOTECT ANALYSIS' (a sustainable design analysis software). Interestingly, we find an important trio of keywords: 'building', 'information' and 'modeling', as in BIM. The keywords 'construction project', 'cultural heritage' and 'institutional theory' are also found here. And lastly, this cluster also includes a 'technology' theme with the keywords 'parametric design', 'computer based environment', 'generative design', 'integrated design' and 'conceptual design'.
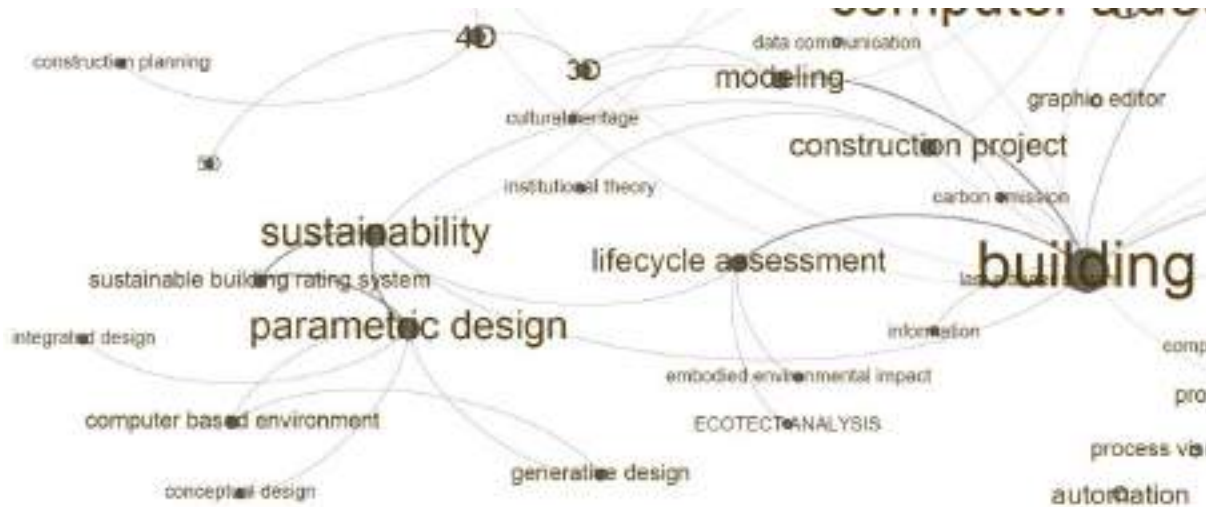
Figure 10: Zooming in on the 'building' community from the BIM global cluster

The **'collaboration' community** is shown in Figure 11. The keywords 'collaboration', 'communication' and 'visualization' were already visible in the BIM core model. At the BIM global cluster level, we observe a few keywords related to collaboration: 'contractual', 'liability', 'integrated practice' and 'complexity'. This community also regroups '3D model', '3D CAD', '4D CAD', '4D BIM' and 'construction scheduling'. The keyword 'coordination' also appears in this community.
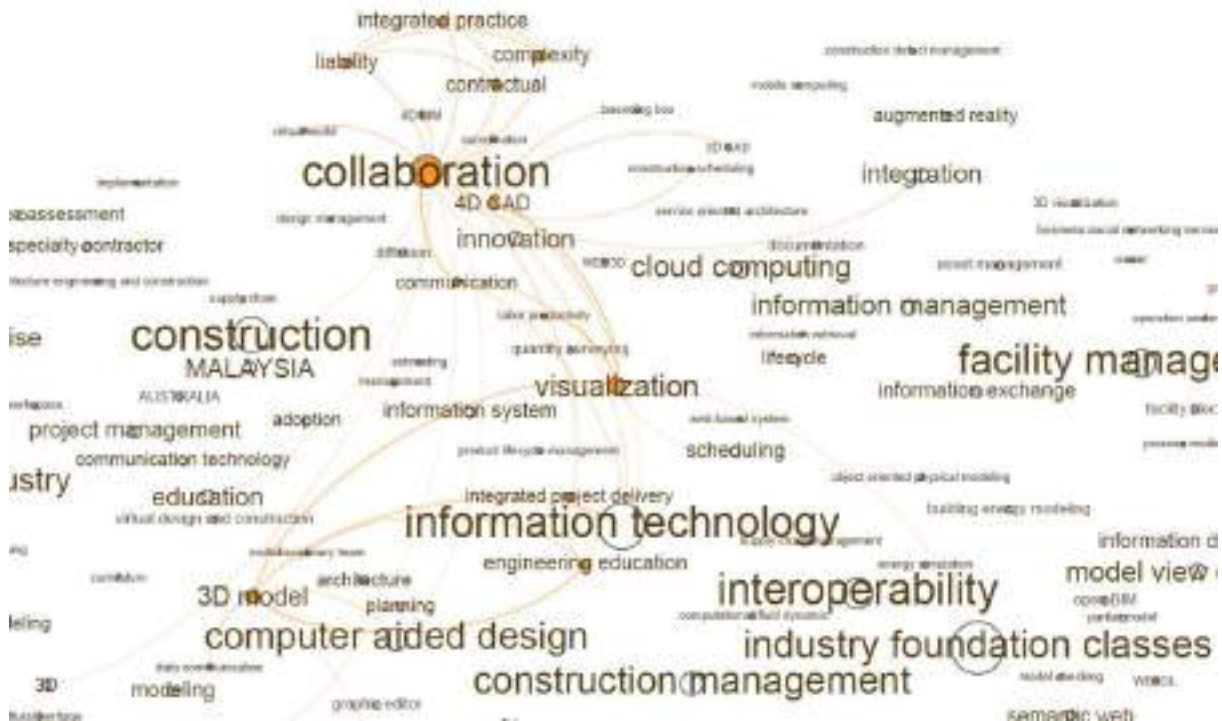

Figure 11: Zooming in on the 'collaboration' community from the BIM global cluster

The **'computer aided design' community**, Figure 12, had three nodes in the BIM core cluster: 'computer aided design', 'construction management' and 'software'. In the BIM global cluster, this cluster now has 'innovation' linked to 'design management' and 'diffusion'. The keyword 'lean construction' also appears, linked to 'construction management' and to 'computer aided design'. Attached to 'lean construction', we find 'production control' and 'process visualization'. On the software side, we find 'graphic editor' and 'comparative'.
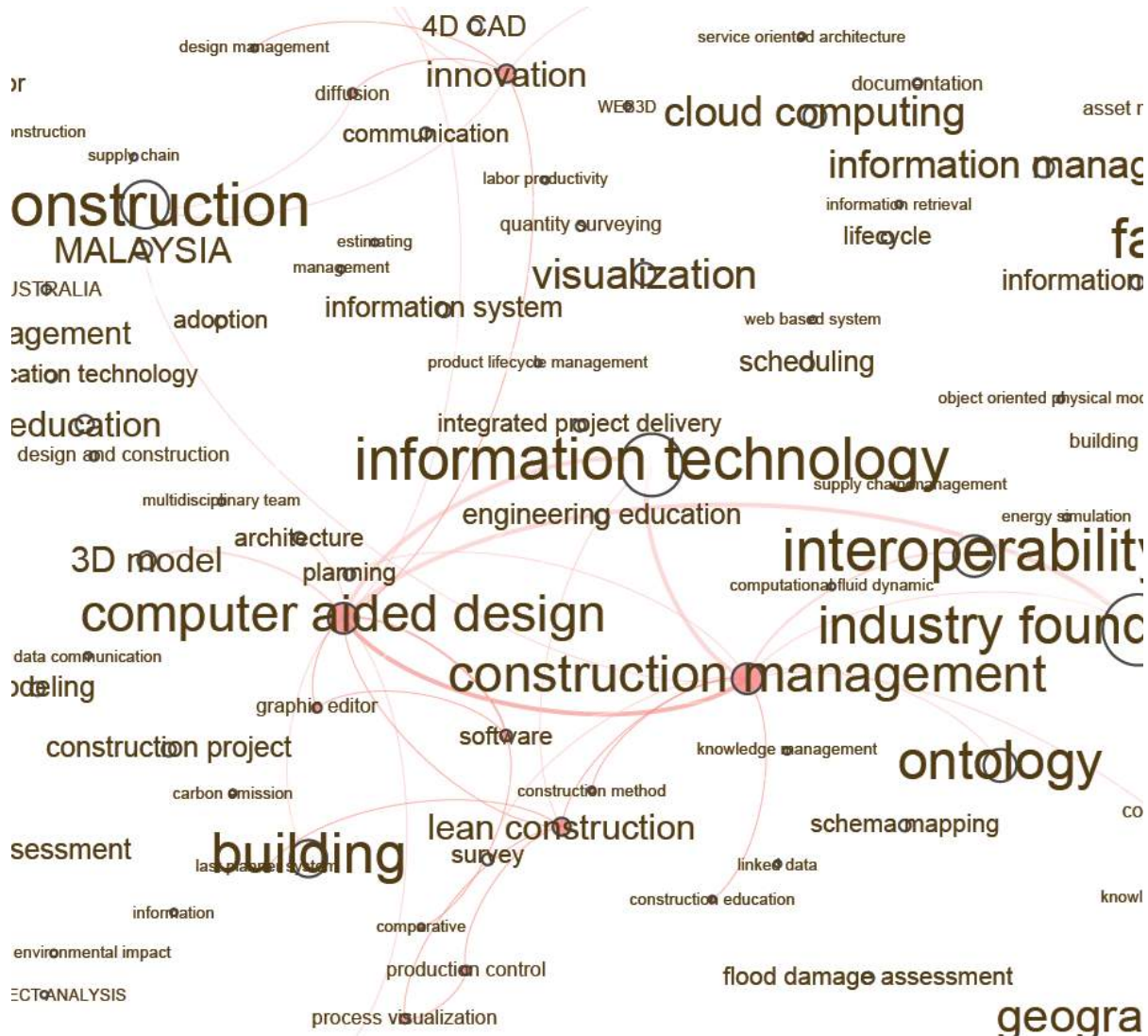

Figure 12: Zooming in on the 'computer aided design' community from the BIM global cluster

The **'laser scanning' community** is a new cluster that now appears in the BIM global cluster (Figure 13). Its dominant keywords are 'laser scanning' and 'point cloud'. A different technique appears: 'photogrammetry'. We observe a group of logically connected keywords: '3D modeling', 'as built BIM', 'scan to BIM' and 'object recognition'. On the technical side, we find 'laser scanner', '3D laser scanning', 'laser detection and ranging' and 'constrained least square adjustment'. On the application side, the graph shows 'historic building information modeling'. This community is linked to other clusters via only two of its keywords: 'automation' and 'survey'.

Figure 13: Zooming in on the 'laser scanning' community from the BIM global cluster

5. DISCUSSION

Many of this study's findings confirm the outcomes from recent research, especially regarding the most-used keywords. Table 4 compares the top keywords appearing in our study with those found by Li et al. (2017). However, the grouping done by Li et al. (2017) during normalization makes the comparison somewhat difficult. Overall, the main themes identified are similar in both studies. Moreover, the count of the main keywords is comparable. The research methodology (retained dates of publications, sources, datasets and standardization rules) certainly has an impact on keyword ranking and counts.

Figure 14 highlights 19 keywords amongst the 30 top keywords from Table 6; these 19 are common to or very similar to those both found by Li et al. (2017) and identified in our study. The keywords that are not common may differ because the study from Li et al. uses both keyword co-occurrence and the co-citation network, while our study does not use co-citation.

Table 4 : Comparison of the top 30 keywords from (Li et al., 2017) with our study

Top keywords from (Li et al., 2017)	Frequency	Our study's top keyworks	Count
BIM/BIMS	732	building information modeling	515
Construction/Industry/AEC	321	BIM	374
System/Information system	148	industry foundation classes	94
3D/nD Modeling Application	112	building information model	92
Design (e.g. Parametric/Rule-based)	100	interoperability	49
Software	68	information technology	48
Green Building/Sustainability/Energy	65	facility management	47
Industry Foundation Classes/IFC	55	construction	45
Interoperability/Data Exchange	51	collaboration	31
Simulation	44	construction management	30
Laser Scanning/Point Cloud	42	geographic information system	28
Visualization	39	ontology	28
Geographic Information	34	building	27
Implementation/Adoption	33	computer aided design	27
Cost control	33	sustainability	24
Augmented Reality/Virtual Reality	33	laser scanning	22
Facility Management	30	visualization	22
Performance	27	BIM technology	21
Life Cycle Management	27	case study	20
Ontology	27	construction industry	20
Knowledge Management	25	integrated project delivery	20
Collaboration	25	design	18
Engineering Education	23	energy efficiency	18
Algorithm	22	lifecycle	18
Scheduling	19	parametric design	18
Operation/Operator	17	point cloud	18
Integration/Segmentation	17	project management	18
Safety Management	16	cloud computing	17
Standard	15	simulation	17
Spatial Analysis	15	information management	16

Our results can also be compared with the top keywords presented by He et al. (2017). Even though their dataset focuses on managerial areas of BIM, keywords such as 'information technology' and 'construction management' appear high in both their ranking and ours.
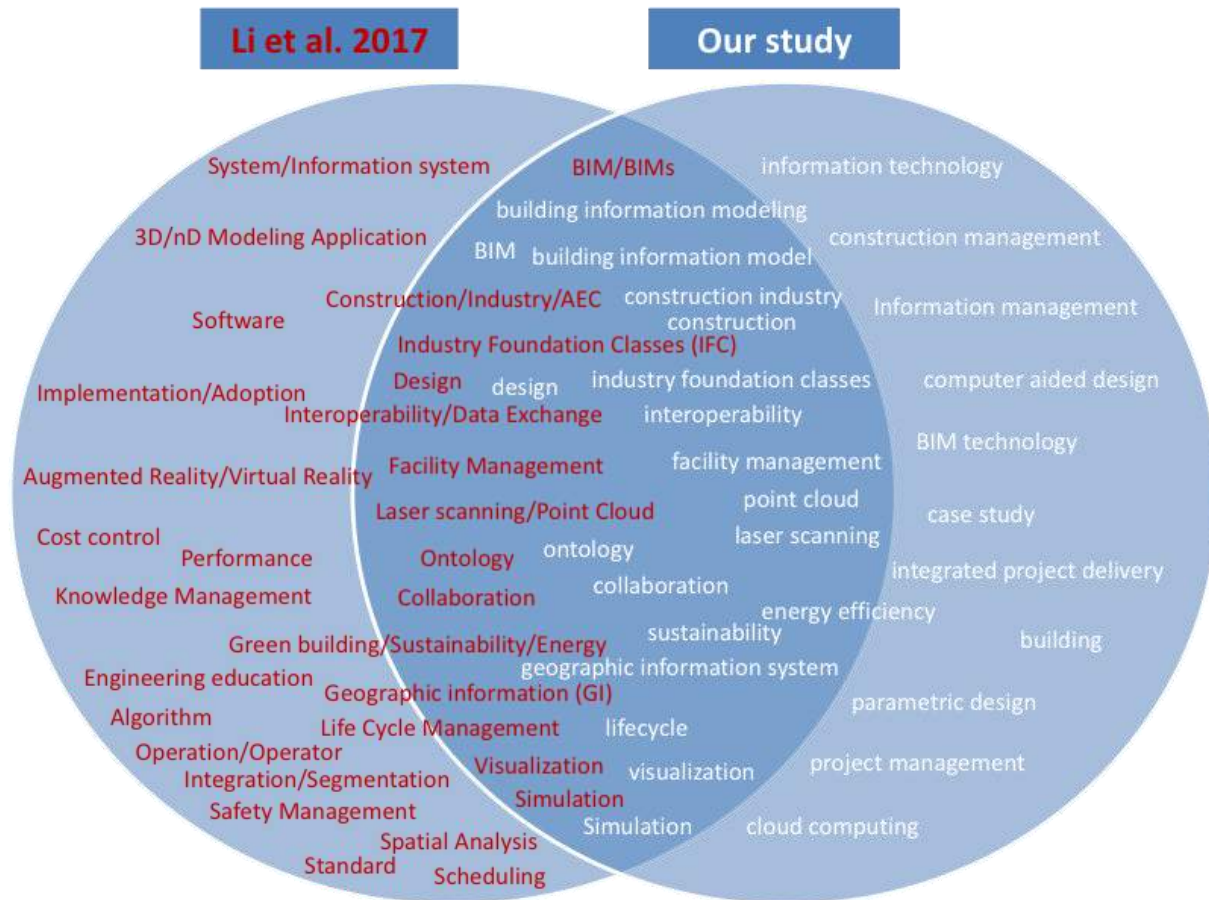
Figure 14: Similar keywords amongst the 30 top keywords from (Li et al., 2017) and our study

On the co-occurrence network side, the graph proposed by Oraee et al. (Oraee et al., 2017a) comes close to our study, as they propose a co-occurrence network of authors' keywords. However, Oraee et al. (2017a) concentrated on BIM publications that have a collaboration connotation, while our paper addresses BIM in a more neutral way. We can certainly observe that our BIM core cluster, described above, has many common nodes with Oraee et al.'s results, but in their result 'Building information modeling' plays a central role that has a major impact on the network, which is the very reason why we decided to filter it from our graph.

A significant contribution of our study is that we have been able to use modularity to identify, within the network of normalized keywords, the communities - or clusters - of keywords that reveal the components of the clusters, and not only the clusters' names. None of the previous studies has provided such a deep graph analysis as our filtering using the edges' weights.

One remark is needed about the method used in this study, which leads both to a strong point and to a possible weakness. When normalizing the authors' keywords, we chose not to modify authors' intent by grouping keywords into categories. Hence, the resulting data includes distinct keywords that are somewhat similar, such as 'education' and 'engineering education'; both appear in distinct communities: 'education' appears in the 'information technology' community, while 'engineering education' appears in the 'collaboration community'. Another example of such a situation where two different, but related, keywords are 'construction scheduling' found in the 'collaboration' community and 'construction planning', found in the 'building' community. Hence, grouping keywords, or not, in the normalization process remains a delicate issue.

6. CONCLUSION AND FUTURE WORKS

This article introduces a method to analyze BIM literature through 10 years of scientific publications. The authors retained 1244 publications that fit a list of criteria and suggested a process based on keywords selected by the authors of these publications. After extracting these keywords from the database, a normalization step was necessary to avoid variations and to obtain consistent results. Next, the authors' keywords were analyzed in terms of their frequency of appearance in publications, showing that "big players" have an important role and that most of the terms are only cited once. Furthermore, keywords were used to build a co-occurrence network. Several steps of filtering and analysis of the graphs were conducted to better understand and visualize the relations between keywords. Thanks to a modularity calculation, seven distinct and major clusters were identified.

The contribution of this paper to the corpus of knowledge is to propose an unbiased and reproducible method of reviewing BIM literature. This can be used in opposition or in addition to traditional narrative literature reviews. Entirely based on bibliometrics, this approach can be generalized and used for mapping knowledge in other domains. Therefore, it is recommended that similar studies should be conducted in other fields, as well as studies in comparing graphs from different domains. These may require some mathematical knowledge, including exploiting works in graph isomorphism and graph matching areas. In the case of BIM, analogies and interactions with PLM or Lean Construction have already been identified as potential further work.

REFERENCES

Abdal Noor, B. and Yi, S. (2018) 'Review of BIM literature in construction industry and transportation: meta-analysis', *Construction Innovation*, 18(4), pp.433-452. doi: 10.1108/CI-05-2017-0040

Bastian, M., Heymann, S., and Jacomy, M. (2009) 'Gephi: an open source software for exploring and manipulating networks', *Icwsm*, *8*(2009), 361-362.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E. (2008) 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10). doi: 10.1088/1742-5468/2008/10/P10008.

Booth, A., Papaioannou, D. and Sutton, A. (2012) 'Systematic approaches to a successful literature review', *Educational Psychology in Practice*. SAGE Publications. doi: 10.1080/02667363.2014.900913.

Broadus, R. N. (1987) 'Toward a definition of "bibliometrics"', *Scientometrics*, 12(5–6), pp. 373–379. doi: 10.1007/BF02016680.

Carney, S. and Geddes, J. (2002) 'Systematic reviews and meta-analyses', in *Evidence in mental health care*. London: Routledge, Taylor & Francis Group, pp. 73–80.

Cerovsek, T. (2011) 'A review and outlook for a "Building Information Model" (BIM): A multi-standpoint framework for technological development', *Advanced Engineering Informatics*, 25(2), pp. 224–244. doi: 10.1016/j.aei.2010.06.003.

Cheng, J. C. P. and Lu, Q. (2015) 'A review of the efforts and roles of the public sector for BIM adoption worldwide', *Journal of Information Technology in Construction*, 20(July), pp. 442–478.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., and Herrera, F. (2011) 'Science mapping software tools: Review, analysis, and cooperative study among tools', *Journal of the American Society for Information Science and Technology*, 62(7), pp. 1382–1402. doi: 10.1002/asi.21525.

Cooper, I. D. (2016) 'What is a "mapping study" ?', *Journal of the Medical Library Association : JMLA*, 104(1), pp. 76–8. doi: 10.3163/1536-5050.104.1.013.

Danon, L., Diaz-Guilera, A., Duch, J., and Arenas, A. (2005) 'Comparing community structure identification', *Journal of Statistical Mechanics: Theory and Experiment*, (9), pp. 219–228. doi: 10.1088/1742-5468/2005/09/P09008.

Dowsett, R. M. and Harty, C. F. (2014) 'Evaluating the benefits of BIM for sustainable design - A review', *Proceedings 29th Annual Association of Researchers in Construction Management Conference, ARCOM 2013*, (September), pp. 13–23.

East, E. W., Nisbet, N. and Liebich, T. (2013) 'The Facility Management Handover Model View', *Journal of Computing in Civil Engineering*, 27(February), pp. 61–67. doi: 10.1061/(ASCE)CP.1943-5487.0000196.

Fink, A. (2014) '*Conducting research literature reviews: From the Internet to paper',* Fourth edition, Sage Publications.

Guimerà, R. and Amaral, L. A. N. (2005) 'Functional cartography of complex metabolic networks', *Nature*, 433(7028), pp. 895–900. doi: 10.1038/nature03288.

Hartley, J., and Kostoff, R. N. (2003) 'How Useful are `Key Words' in Scientific Journals?' *Journal of Information Science*, 29(5), 433–438. https://doi.org/10.1177/01655515030295008.

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*. Elsevier Ltd, APM and IPMA, 35(4), pp. 670–685. doi: 10.1016/j.ijproman.2016.08.001.

Hichri, N., Stefani, C., De Luca, L., and Veron, P. (2013) 'Review of the « As-Built BIM » Approaches', *3D-ARCH 2013 - 3D Virtual Reconstruction and Visualization of Complex Architectures*, XL-5/W1(February), pp. 107–112. doi: 10.5194/isprsarchives-XL-5-W1-107-2013.

Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014) 'ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software', *PLoS ONE*, 9(6), pp. 1–12. doi: 10.1371/journal.pone.0098679.

Kapoor, K., Sharma, D. and Srivastava, J. (2013) 'Weighted node degree centrality for hypergraphs', in *Proceedings of the 2013 IEEE 2nd International Network Science Workshop, NSW 2013*, pp. 152–155. doi: 10.1109/NSW.2013.6609212.

Kitchenham, B., Budgen, D. and Brereton, O. P. (2010) 'The value of mapping studies – A participant-observer case study', *Proceedings of the international conference on Evaluation and Assessment in Software Engineering (EASE'10)*, (June), pp. 25–33.

Kipp, M.E.I. (2005). 'Complementary or discrete contexts in on-line indexing: A comparison of user, creator and intermediary keywords'. *Canadian Journal of Information and Library Science*, 29(4), 419–436.

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*. Elsevier, 84(October), pp. 195–206. doi: 10.1016/j.autcon.2017.09.011.

Luciano, T. (2011) 'Doing a literature review: releasing the social science research imagination', *Evaluation & Research in Education*. Taylor & Francis, 24(4), pp. 303–304. doi: 10.1080/09500790.2011.588012.

Lu, K., and Kipp, M. E. I. (2014). Understanding the Retrieval Effectiveness of Collaborative Tags and Author Keywords in Different Retrieval Environments: An Experimental Study on Medical Collections. Journal of the Association for Information Science and Technology, 65(3), 483–500. https://doi.org/https://doi.org/10.1002/asi.22985.

Mulrow, C., Cook, D. and Davidoff, F. (1997) 'Systematic reviews: Critical links in the great chain of evidence', *Annals of Internal Medicine*, 126(5), pp. 389–391.

Mulrow, C. D. (1994) 'Systematic Reviews: Rationale for systematic reviews', *Bmj*, 309(6954), pp. 597–599. doi: 10.1136/bmj.309.6954.597.

Newman, M. E. J. (2003) 'Structure and function of complex networks.', *SIAM Reviews*, pp. 167–256. doi: 10.1137/S003614450342480.

Newman, M. E. (2006) 'Modularity and community structure in networks', *Proceedings of the national academy of sciences*, *103*(23), 8577-8582.

Nyffenegger, F., Rivest, L. and Braesch, C. (2016) 'Identifying PLM themes, trends and clusters through ten years of scientific publications', in *IFIP Advances in Information and Communication Technology*, pp. 579–591. doi: 10.1007/978-3-319-54660-5_52.

Opsahl, T., Agneessens, F. and Skvoretz, J. (2010) 'Node centrality in weighted networks: Generalizing degree and shortest paths', *Social Networks*. Elsevier B.V., 32(3), pp. 245–251. doi: 10.1016/j.socnet.2010.03.006.

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

Oraee Namzadi, M., Hosseini, M. R., Banihashemi Namini, S., and Merschbrock, C. (2017b) 'Where the gaps lie: ten years of research into collaboration on BIM-enabled construction projects', *Construction economics and building*, *17*(1), 121-139.

Petticrew, M. and Roberts, H. (2006) 'Systematic Reviews in the Social Sciences: A Practical Guide', *Evaluation Journal of Australasia*, 9(1), p. 352. doi: 10.1027/1016-9040.11.3.244.

Rousseau, D. M., Manning, J. and Denyer, D. (2008) 'Evidence in management and organizational science: Assembling the field's full weight of scientific knowledge through syntheses. The Academy of Management Annals, 2 (1): 475-515, 2008', *Marketing Letters*, 21, p. 6581. doi: 10.2139/ssrn.1309606.

Santos, R., Costa, A. A. and Grilo, A. (2017) 'Bibliometric analysis and review of Building Information Modelling literature published between 2005 and 2015', *Automation in Construction*. Elsevier B.V., 80, pp. 118–136. doi: 10.1016/j.autcon.2017.03.005.

Small, H. (1999) 'Visualizing science by citation mapping', *Journal of the American Society for Information Science*, 50(9), pp. 799–813. doi: 10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G.

Volk, R., Stengel, J. and Schultmann, F. (2014) 'Building Information Modeling (BIM) for existing buildings - Literature review and future needs', *Automation in Construction*. Elsevier B.V., 38, pp. 109–127. doi: 10.1016/j.autcon.2013.10.023.

Wong, J. K. W. and Zhou, J. (2015) 'Enhancing environmental sustainability over building life cycles through green BIM: A review', *Automation in Construction*. Elsevier B.V., 57, pp. 156–165. doi: 10.1016/j.autcon.2015.06.003.

Yalcinkaya, M. and Singh, V. (2014) 'Building information modeling (BIM) for facilities management-literature review and future needs', *IFIP Advances in Information and Communication Technology*, pp. 1–10. doi: 10.1007/978-3-662-45937-9_1.

Young H. (1983) '*The ALA Glossary of Library and Information Science*', American Library Association. Chicago.

Zhou, W., Whyte, J. and Sacks, R. (2012) 'Construction safety and digital design: A review', *Automation in Construction*. Elsevier B.V., 22, pp. 102–111. doi: 10.1016/j.autcon.2011.07.005.