VILNIUS TECH
Vilnius Gediminas
Technical University

TEXT MINING-BASED PATENT ANALYSIS OF BIM APPLICATION IN CONSTRUCTION

Xing PAN, Botao ZHONG*, Xiaobo WANG, Ran XIANG

Hubei Engineering Research Center for Virtual, Safe and Automated Construction,
Huazhong University of Science & Technology, Wuhan, China
School of Civil and Hydraulic Engineering, Huazhong University of Science & Technology, Wuhan, China

Abstract. As a data tool applicable to the full life-cycle of construction engineering and management, Building Information Modeling (BIM) has great potential for significantly increasing project productivity and performance. Awareness of BIM application hotspots and forecasting its trends can drive innovations in construction field. Using patents as data resources, this study develops an effective framework integrating the citation network analysis and the topic clustering technology to identify BIM application information and forecast its trends. This framework comprises three-step analysis: (1) quantitative characteristic analysis of patent outputs; (2) Social Network Analysis (SNA)-based co-occurrence network analysis; and (3) identification of BIM topics using a Latent Dirichlet Allocation (LDA). Finally, the case demonstrates the effectiveness of this framework contributing to promote technological development and innovation of BIM. The contributions of this study are threefold: (1) an innovative text mining-based framework for BIM patent analysis in construction is developed; (2) patents that have focused on identifying the application hotspots and development trend of BIM in accordance with our developed framework are reviewed; and (3) a signpost for technological development and innovation of BIM is provided.

Keywords: Building Information Modeling (BIM), patent-driven analysis, text mining, Social Network Analysis (SNA), Latent Dirichlet Allocation (LDA), application hotspots and forecasting.

Introduction

The *Engineering Fronts* research project has been jointly launched by the Chinese Academy of Engineering, Clarivate Analytics, and Higher Education Press since 2017 (Cai et al., 2018). The global experts are gathered to evaluate current frontiers in engineering research, and provide strategic responses to the issues of global challenges and sustainable development. In the field of construction engineering and management, the frontiers focus on demand-driven technology. The development front, referring to the development of technology, can be screened out from patent data aiming to solve practical problems in technology. It is forward-looking, pioneering and exploratory, which has a major impact and indispensable role in the future of engineering technology. The top development fronts include construction, transportation, mechanical, energy, medicine, electronics, and others. Especially, the Building Information Modeling-based (BIM-based) construction management system is one of the top 10 development fronts (Cai et al., 2018).

BIM is often interpreted as an overarching tool for describing a variety of activities in object-oriented Computer Aided Design (CAD), using modern modeling techniques, such as Revit and Tekla. The content generated by architects, designers and engineers has evolved from traditional 2D-drawings to object-oriented 3D-models embedded with detailed information (Johansson et al., 2015). As a digital replica, BIM is used as an information repository to support a variety of applications in the design and construction stages (Miettinen & Paavola, 2014). Moreover, BIM can further promote the real-time visualizations since all data can be represented with a 3D fashion. This particular feature can be used to communicate ideas and share information when engaging with multiple stakeholders in a project (Johansson et al., 2015). As seen from the aforementioned studies, as a data tool applicable to engineering design, construction, and management, Building Information Modeling (BIM) has brought forth great changes to precisely digitalize the virtual model of

*Corresponding author. E-mail: dadizhong@hust.edu.cn

building in the full life-cycle of construction engineering and management (Azhar, 2011). For example, BIM can be applied to cost management, quality and safety management, technical management, and field management, which improves greatly the communication efficiency of project participators and realizes the control over cost and progress to a larger degree (Volk et al., 2014; Park et al., 2013; Lu et al., 2017).

Nevertheless, many experts have different views on the development trend of BIM. Mohd and Ahmad (2013) explored BIM research in construction planning and found that BIM can be used to detect clash analysis during design phase before construction, as well as to optimize the project schedule, cost, quality, and enhance communication between constructors. Jia et al. (2017) investigated the research of BIM on residential buildings' design stage under the generally used Design-Bid-Build (DBB) mode. However, quite little has been done about the analysis of the BIM' application hotspots and development trend, which is harmful to the development and innovation of BIM technology.

Patent analysis may help to understand the innovation and development of a technological field in the course of time (Griliches, 1990; Mcaleer & Slottje, 2005). For instance, patents can be used to optimize investments by avoiding ill investment decisions in unnecessary technology areas and design research strategy to secure core patents (Kim et al., 2008; Ju & Sohn, 2015). The patent is one of the most reliable sources of information technology, and hence, they are used as a data tool for analyzing technological developments (Pilkington et al., 2009; Zhang, 2011). Patent citation information is often applied to investigate the development of patented technologies in different fields. For instance, Daim et al. (2020) discussed the relative locations of a company in the technological network based on patent citations of the IoT cybersecurity and blockchain. However, there is still a need to apply patent data analysis in the BIM field to explore the BIM' development trend and application hotspots.

This study aims to identify the application hotspots of BIM and forecast its trends in an effective framework using patents as data resources and integrating two text mining technologies, the citation network analysis and the topic clustering technology. This framework comprises three-step analysis: (1) quantitative characteristic analysis of patents outputs to provide a visual overview of BIM patents; (2) the production of a Social Network Analysis (SNA)-based co-occurrence network analysis to determine the citation interrelations between BIM patents; and (3) identification of BIM patent topics using a Latent Dirichlet Allocation (LDA) to further analysis the topic characters of critical patents. Besides, a good clustering approach of patent map demonstrates the effectiveness of the developed LDA model. In this study, the actual case demonstrates the effectiveness of this framework contributing to promote the knowledge diffusion within the BIM field for further research and explore critical patents to promote technological development and innovation.

The contributions of this study are threefold: (1) an innovative text mining-based framework for BIM patent analysis in construction is developed; (2) patents that have focused on identifying the application hotspots and development trend of BIM in accordance with our developed framework are reviewed; and (3) a signpost for technological development and innovation of BIM is provided.

1. Related research

1.1. Patents' analysis

Patents are often regarded as an output of Research and Development, an indicator of innovation and the timely measure of rapid technological change (Kang, 2014; Markatou & Vetsikas, 2015). Patent analysis usually refers to the analysis, processing and combination of a large number of fragmented patent information in patent specifications and patent publications. Using statistical methods, patent information can be transformed into intellectual understanding with overall and predictive capabilities.

With the widely used Information Technology (IT), patent analysis has been extensive used as a powerful tool in different fields (Kim & Bae, 2017; Boshnakoska et al., 2013). For instance, in the hydrogen energy and fuel cell field, Li et al. (2020) used a patent citation analysis and text mining to monitor and forecast the development trends of nanogenerator technology. In the coherent light generator technology field, You et al. (2017) applied patent analysis technology to predict the development trends of coherent light generator technology and revealed the focal technology areas in Research and Development. In the Biosensor Technologies field, N. J. Sheikh and O. Sheikh (2017) used patent analysis to identify the trends for the three types of biosensors, including blood, saliva, and breath. In the medical field, Zhang et al. (2016) used patent analysis to reveal the hot research fields of Chinese medicine patents and the future development trend of traditional Chinese medicine. In the building field, Altwies and Nemet (2013) made an assessment of patent citations in building energy control technology to improve the innovation in the U.S. building sector.

All in all, patent analysis helps the researchers understand the trends of related technologies and research hotspots in the targeted field. Analysis of patent data allows subsequent research to focus on core patents. Besides, patent analysis is able to identify new development directions and explore new and competitive patented technologies, providing technical strategies for the development of construction industry.

1.2. Social network analysis of patents

Social Network Analysis (SNA), as a useful tool in analyzing relationship in a social network, has some benefits of powerful management of examining their characteristics such as relationships, information exchange, physical attributes, etc. (Hatala & Lutta, 2009). It is particularly useful for the quantitative and visual analysis of the interpretation of human associations.

Researchers have been investigating the applicability of the social networks since the early 20th century in diverse fields, including sociology, biology, economics, and several others (Al Hattab & Hamzeh, 2015; Han et al., 2017; Ardito et al., 2017). Yoon et al. (2011) proposed a novel method called Invention Property-Function Network Analysis (IPFN) to extract properties and functions from patents related to silicon-based thin film solar cells, and to identify the technological implications of properties and functions using SNA. Perng and Huang (2016) revealed the current status of patent activities, major assignees, assignee countries, and core technology fields of shading devices by using SNA.

As seen from the aforementioned studies, SNA is able to help researchers understand the network relationship visually, as well as convey the results of the analysis. Most importantly, SNA can reveal hidden patterns that might not be captured by conventional qualitative measures and assist professionals to identify the emerging development trends of new technologies.

1.3. Text clustering of patents

Text clustering can be utilized to extract data distribution information from textual unstructured data where there is no prior knowledge (Sohrabi et al., 2018). Given a sample set of texts, text clustering relies on a distance-based approach to divide text data into multiple categories by using common methods, such as agglomeration and partitioning algorithms. Among them, agglomeration-based algorithms are designed to merge documents into different clusters according to the identified relationships among documents, and the generated clusters can then be organized into a dendrogram or a cluster hierarchy (Sohrabi et al., 2018).

Many attempts can be found in some literatures. Yang et al. (2009) used text clustering to forecast the future possible tendency of Machine Learning. Li and Wu (2010) used text clustering to detect and forecast online forums research hotspots. Han et al. (2017) used patent information to obtain an overview of different technology topics in a given field and find core patents. Sohrabi et al. (2018) applied text clustering algorithms on the frontpage information and acquired trends in information systems journal articles in the field of Human Resources Management. Li et al. (2019a) used text clustering to identify technology evolution paths and forecast technology development trends of perovskite solar cell technology within the short term.

From the investigation of those previous literatures, it is clear that, although text clustering method and SNA technology have been widely used, quite little has been done about the analysis of the BIM patent for identifying the application hotspots of BIM and forecasting its trends.

2. Methods

This study develops an effective framework using patents as data resources and integrating two text mining technologies, the citation network analysis and the topic clustering technology, to identify the application hotspots of BIM and forecast its trends. Figure 1 shows the outline of research design, detailly discussed in Sections 2.1 to 2.2.

2.1. Data collection and processing

In this study, the patent data is obtained from the Derwent Innovations Index (DII) database entering from 2004 to 2020. Patent retrieval queries consists of keywords and publication dates. The patent data is initially screened by the
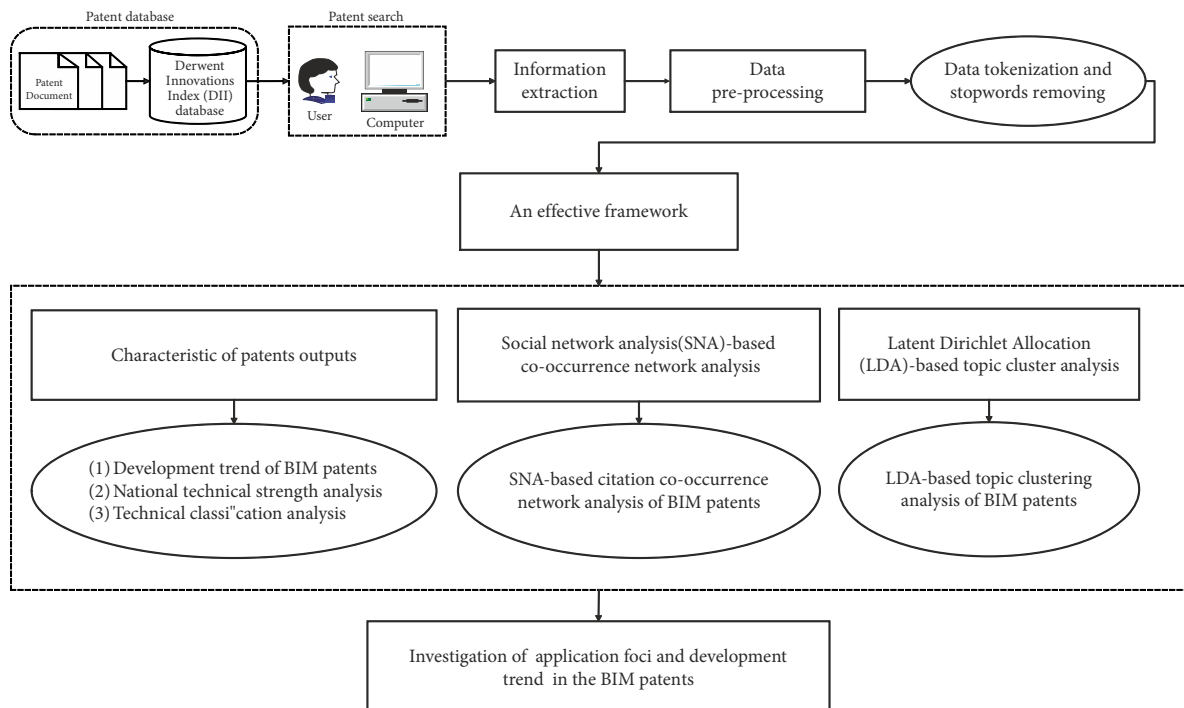

Figure 1. Outline of research design

search formula TI = (BIM) AND TI = (building information modeling) AND AB = (BIM) AND AB = (building information modeling) AND DP >= (20040101) AND DP <= (20201231). The entire data collection process mainly follows two steps. Firstly, considering the BIM is an integrated concept, the keyword expansion and searched both title and abstract are preformed to capture a wider picture. In this study, the "BIM" and "Building Information Modeling" are as keywords to search for the title and abstract of the patents in the Derwent Innovations Index database with confined publication time from 2004 to 2020. Secondly, a resultant 1510 patents are exported and manually screened out. Data dictionary backup is also established based on the those manually screened patents.

After the data collection, the preprocessing for BIM patent is conducted using tokenization and stopwords removing. The data preprocessing is performed by stopwords which removed and used the dictionary to merge the keywords in the original text to strictly conform to the dictionary content. Developing tool of models in Python 3.6 of the Anaconda environment main packages (such as: genism, nltk, openpyxl) are used for implementing the Latent Dirichlet Allocation (LDA) model. Gephi 0.9.2 is then used to perform the visualization of network analysis.

2.2. Text mining

2.2.1. SNA-based citation network analysis

Social Network Analysis (SNA) is a scientific and systematic approach in which a node is a role and links are used to show relationships between nodes (Perng & Huang, 2016). Regarding the attributes of the links, there are two types of social networks: directed networks and undirected networks. Links in undirected networks are merely used for representing a general relationship between nodes, while arrows are used to indicate a directional relationship between nodes in directed networks. By using the network measures such as density, degree centrality, intermediateness, and tightness centrality, the critical nodes and prominent network features can be revealed accordingly. The following network measures are applied in this study:

(1) Density

For a given network or graph, the Density (D) is a measure that counts the number of links to other nodes in a directed or undirected network, as shown in Eqns (1) and (2), respectively (Zhong et al., 2019). This measure is a network-level measure for assessing the overall connectivity of a network with N nodes. It can be used to identify the completeness of the network as a network with a high density is highly connected.

$$D = \frac{L}{N(N-1)};$$ (1)

$$D = \frac{L}{N(N-1)/2},$$ (2)

where N is the total number of nodes and L is the total number of links in a given network.

(2) Degree centrality

Degree centrality (CD) is a nodal-level centrality measure that measures the number of directed or undirected links that a node has with other nodes in a given network (Zhong et al., 2019). This nodal measure can be calculated using Eqn (3) and can be used to indicate a node's number of connected neighbors in the network. In a social network, generally speaking, a node with a higher degree centrality indicates that it has more influence and importance than other nodes (Krebs, 2002).

$$C_D(P_k) = \sum_{i=1}^{N} a(P_i, P_k),$$ (3)

where P_k is a given node k; N is the total number of nodes; $a(P_i, P_k)$ = [1 if node i and node k are connected, 0 otherwise].

(3) Eigenvector centrality

Eigenvector centrality is another important nodal measure, which measures the influence of a node in a given network by assigning relative scores to all nodes based on the concept: connections to high-scoring nodes contribute more to the score of the node than equal connections to low-scoring nodes. Let $G(E, V)$ be a graph, consisting of vertices V and edges E. Let A be the adjacency matrix for this graph; $a_{ij} = 1$ if vertices i and j are connected by an edge and $a_{ij} = 0$ if they are not. Because A is symmetric all its eigenvalues are real, its eigenvectors are orthogonal, and it is diagonalizable (Batool & Niazi, 2014). Eqn (4) describes eigenvector centrality x in two equivalent ways, as a matrix equation and as a sum. The centrality of a vertex is proportional to the sum of the centralities of the vertices to which it is connected. λ is the largest eigenvalue of A and n is the number of vertices:

$$Ax = \lambda x, \lambda x_i = \sum_{j=1}^{n} a_{ij} x_j, i = 1, \ldots, n.$$ (4)

2.2.2. LDA-based topic clustering

Latent Dirichlet Allocation (LDA), originally developed by Kim et al. (2018), is a generative probabilistic modeling approach for revealing hidden semantic structures in a collection of textual documents. This technique uses the "Bag of Words" approach, which sees each document as a word-frequency vector. It transforms the text data into a bundle of numerical inputs that is compatible with ample data mining algorithms (Wang & Xu, 2018). Each document can be viewed as a representation of a probability distribution of topics, and each topic can be viewed as a representation of a probability distribution of many words (Blei et al., 2017). Moreover, considering the non-correlation between the components of the random vector in the Dirichlet distribution, the candidate topics are independent of each other (Wang & Xu, 2018). To achieve the above objectives, the LDA uses a joint distribution to compute the conditional distribution of the hidden variable under a given observable variable. The observable variable is a set of words, and the latent variable is the topics.

The concept of LDA is presented in Figure 2. LDA has two main steps: topic distribution per document and word distribution per topic. First, based on the Dirichlet distribution with super parameter of α, the probability of the topic in the document is derived. Then, based on the Dirichlet distribution with super parameter of β, the probability of the word in the topic is derived. A common approach, Gibbs sampling, is implemented for approximate inference.

In Figure 2 α is the parameter of the Dirichlet prior on the per-document topic distributions; β is the parameter of the Dirichlet prior on the per-topic word distribution.

The perplexity (*P*) approach is used to determine the topic number *K*. An appropriate probability distribution has a relatively low perplexity. The perplexity (*P*) can be calculated by Eqn (5):

$$P = \exp\left(-\frac{\sum_d \sum_i^{N_d} \ln p(w_{d,i})}{\sum_d N_d}\right), \qquad (5)$$

where N_d eans word frequency in the d document. $W_{d,i}$ eans the nth word in the d document.

One challenge of this method, similar to other unsupervised learning method, is that the correct result is often unobtainable, which means the result needs to be manually checked. Inspired by the article of Kim et al. (2018), the empirical parameters of LDA are α = *K*/50 and β = 0.01, respectively.

3. Result analysis

In this section, the obtained results and corresponding analyses of the study are presented, including quantitative characteristics of BIM patents outputs, SNA-based citation network analysis of BIM patents, and LDA-based topic clustering of BIM patents. This section aims to draw a comprehensive understanding picture of the selected BIM patents through these three aspects of analysis and summarized insights.

3.1. Quantitative characteristics of BIM patent outputs

3.1.1. Development trend of BIM patents

After data collection and preprocessing, the trend analysis result is shown in Figure 3. This process provided the annual number of patents on BIM technology. As shown in Figure 3, the trend of granted BIM patents from 2010 to 2020 increases significantly. It indicates that BIM serves can be as a more and more important and useful tool in facilitating the construction for its benefits of powerful management of physical and functional digital presentations.

3.1.2. National technical strength analysis of BIM patents

Analyzing the countries of the BIM patent yields the following Figure 4. It shows that the share of China (CN)
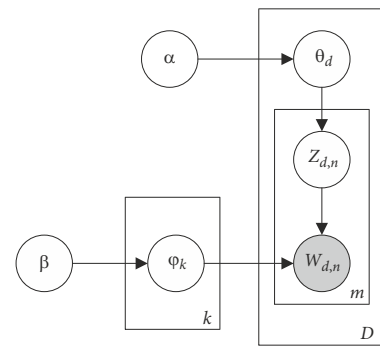

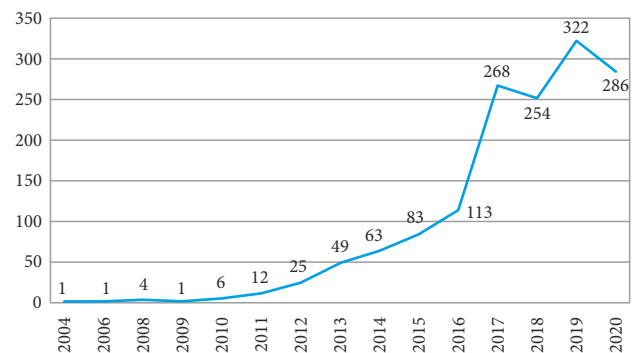Figure 2. Concept of LDA (Kim et al., 2018)

Figure 3. Number of BIM granted patents during the period of 2004–2020
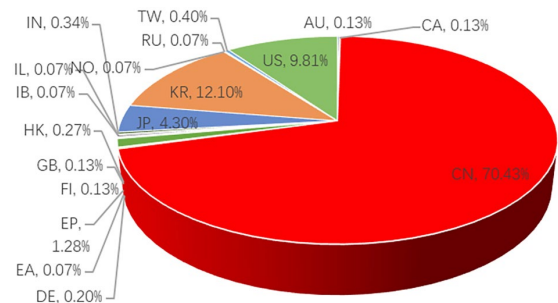

Figure 4. Percentage of BIM granted patents priority countries

reaches 70.43%, the share of Korea (KR) reaches 12.10% and the United States of America (US) reaches 9.81%, respectively, in the overall picture. A possible explanation of this result could be based on the fact that China has been known as a major construction country. Under the initiative of the "Belt and Road" national strategic planning, it is very likely that China occupies an important and a dominant role in the construction field.

3.1.3. Technical classification analysis of BIM patents

The statistical analysis of the International Patent Classification (IPC) (as shown in Figure 5) shows that the three main technical fields of BIM patents are G06F (electrical digital data processing), G06Q (data processing system or method specifically for administrative, commercial, financial, management, supervisory or predictive purposes) and G06T (general image data processing or generation).

Since each patent contained one or more IPC codes, the total number of IPC is greater than the number of patents. Shown in Figure 5, most of the authorized patents are concentrated on G06F, which counted for 48.49% of the total. The second dominant category is G06Q, followed by G06T. After analyzing the corresponding results above, it reveals that the three IPC of G06F, G06Q, and G06T, accounting for 84.78% of total, are recognized as important front direction of BIM technology. The front direction of BIM application can assist the managers make decision and improve the management of construction scheme design, field management and later operation. For the IPC of G06F, for example, the current application trend indicates that researchers are increasingly interested in applying data processing system or method for BIM information management. Accordingly, manager may need focus on the standard for BIM technology with complete, accurate, and timely data exchange and interoperability to improve the cooperation and communication between the various project participants.

3.2. SNA-based citation network analysis of BIM patents

SNA-based citation network analysis, as a typical technique for measuring the knowledge flow of information and innovation, contains valuable data and if analyzed well, can reveal the development trend and application hotspots of technology (Sharma & Tripathi, 2017). The constructed patent citation network is a directed graph that points from the cited patent to the referenced patent. This patent citation network can show the flow of information and knowledge, including the spread and the diffusion of knowledge.

Since the citations indicate the technological development context, individuals can develop citation networks through patent observation techniques. The more the patent is cited, the more important this patent is. The highly cited patents represent the hotspots of the technology and the guiding future technological innovations. Generally, the highly cited patents are at the core of the network, indicating that it is at a central place and conducive to technology diffusion.

To analyze these influential technologies, a core network of BIM technology reference relationships is built based on link weights (Figure 6). The patent citation network consisted of 1941 nodes with a graph density of 0.001 and an average degree of 1.116. The higher the degree of concatenation, the more frequently the patent is cited, and the more important the display is.

As shown in Figure 6 and Table 1, the six core technical fields of BIM patents are US20150248503A1, US20150057981A1, KR2010020060A, US5228038A, CN103093061A, and CN102609417A, respectively. Analysis of core patents finds that Figure 6 shows a core network of "US20150248503A1" consisting of a set of major nodes and several citations. In Figure 6, the "US20150248503A1", the patent code of "A method and system for creating 3D models from 2D data for building
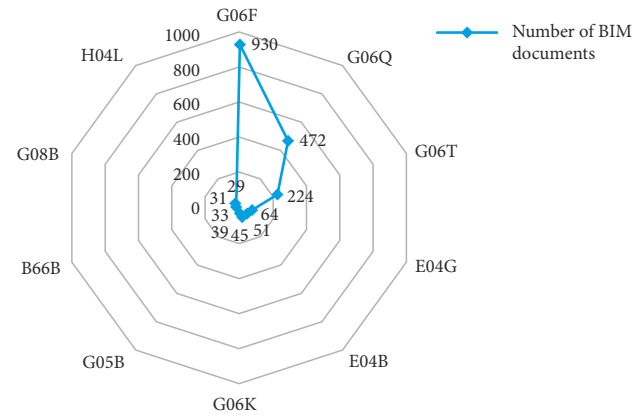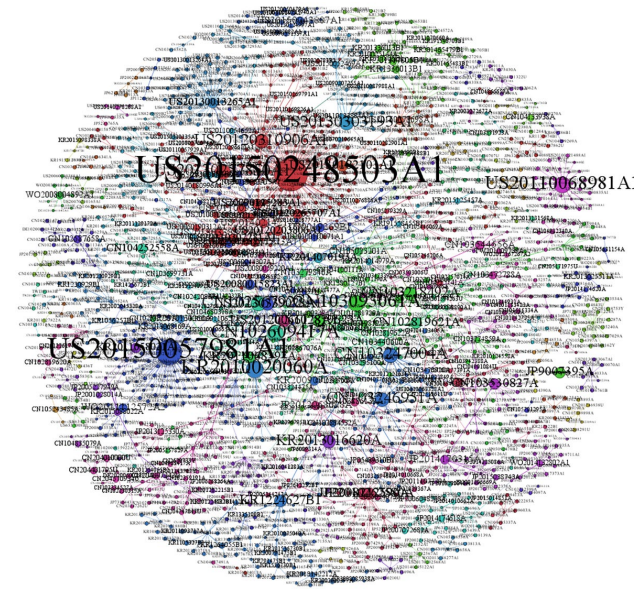

Figure 5. Number of BIM documents

Figure 6. Citation network of BIM Patent code

information modeling (BIM)", appears as the larger sized node. It indicates the "US20150248503A1" patent is highly cited and represents the application hotspot of BIM. Besides, shown in Table 1, the three key performance indicators, namely connected degree, eigenvector centrality and page rank, garner further evaluation into the high value of "US20150248503A1" patent. The modeling of computer aided design drawings into a 3D model (such as AUTODESK REVIT, AUTOCAD, VECTORWORKS, MICROSTATION, ARCHICAD and other software) will reduce the time and labor costs involved in the on-site construction process. Besides, the co-occurrence citation interrelations of the important patents are identified. For the highly cited patent of "US20150248503A1", for example, the "US20120310906A1" (Building Information Tracking System and Method of Use) has interrelations with "US20150248503A1". The interrelations may indicate that the building information tracking system (e.g. the transferring unit for transferring captured information to data repository and versioning unit for versioning infor-

Table 1. Six core technical fields of BIM related patents

Rank	Patent code	IPC	Title	Connected degree	Date of application	Country	Eigenvector centrality	Page rank
1	US20150248503A1	G06F	Method and system for creating 3D models from 2D data for building information modeling (BIM).	29	2015/9/3	US	0.20822832	0.00800604
2	US20150057981A1	G06T	Method and apparatus for multi-mode tracking and display of personnel locations in a graphical model.	22	2015/2/26	US	0.14469843	0.00699503
3	KR2010020060A	G06Q	System for managing BIM information based on a session, particularly for reducing the time and cost for establishing a system.	14	2010/2/22	KR	0.17651657	0.00339755
4	US5228038A	G06Q	Communication system network that includes a BIM status update method.	13	1993/7/13	US	0.08550362	0.00427867
5	CN103093061A	G06F	A metro electromechanical engineering comprehensive pipeline collision optimizing method.	13	2013/5/8	CN	0.16420628	0.00220978
6	CN102609417A	G06F	Engine device and method for data integration and exchange of building information mode based on IFC (industry foundation classes) standards.	13	2012/7/25	CN	0.31446724	0.00297257

mation in repository) plays an important part for tracking content of 3D models by using mobile computing devices. Noteworthy the six core technical fields extracted from the above have close links with the results for the International Patent Classification (IPC) of BIM presented in Section 3.1.3. In Table 1, the "US20150248503A1", as a typical technology belonging to the IPC of G06F (electrical digital data processing), reveals that this technology of 3D model from two-dimension data plays an important part in the BIM-related data processing technologies. For example, as the development of 3D model from two-dimension data, managers can intuitively understand the potential relationship between variables that contribute to causing a collision, and then constructor can optimize the construction design by the navigation collision function of BIM 3D.

3.3. LDA-based topic clustering of BIM patents

As a typically topic clustering approaches, the LDA model has the benefits of considering the word frequency and semantic features of surrounding words to effectively extract the keywords (Zhong et al., 2020). Besides, this model can reduce intensive manual effort, which does not require labeled data. For these benefits, the LDA technique is often used to mine the main topics and identify their respective (corresponding) keywords (Bastani et al., 2018). In this study, the LDA-based topic clustering model is proposed to garner further insights into the application hotspots and development trend of BIM technology.

Here, the topic model LDA is used to investigate the main topics and identify their respective (responding) keywords. These keywords under a topic are treated as the features hidden in the topic, and then the keywords reflect some correlation in a topic.

The relatively appropriate number of topics of all pre-processing data are as shown in Figure 7. The appropriate numbers of topics of documents are 11 by calculating the perplexity. After analyzing all patents, 10 high-quality topics are then selected, as shown in Table 2. LDA is made for 10 topics to identify the latent topics and words distributions. Table 2 presents the top 10 words with the largest weights in each topic, where some impact factors of pre-topic can be extracted. According to the words distributions, the topics are: "Power supply module based on BIM technology", "Steel structure analyzing and detecting management system", "Energy production apparatus management system", "Mobile terminal of quantity calculation method management system based on BIM technology", "BIM model monitoring and memory system", "Plate object display processing management system", "Scene simulation analysis based on BIM technology", "Device management and control center", "Construction method of lift operation system based on BIM technology", and "Project control management system based on Banknote identification module technology". In Table 2, keywords in the per-topic distribution represent closely related factors. Besides, the number represents the probability of keyword in the per-topic.

Table 2. Distributions of the topic of all BIM patents

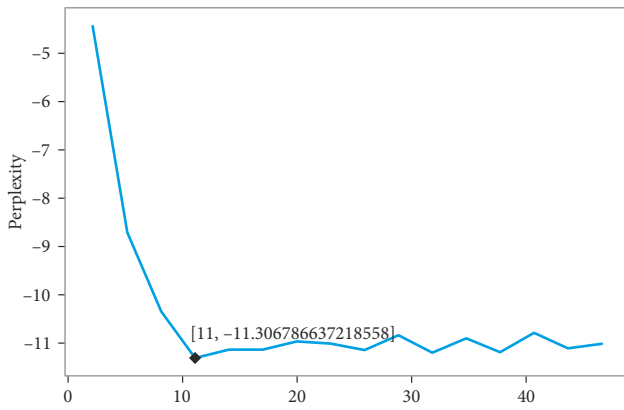| Topic	Keywords and probabilities									
	BIM model	BIM technology	Dimensional	Power supply module	Processor	Assembling	Space	Control system	Area	Software
Power supply module based on BIM technology	BIM model 0.028	BIM technology 0.014	Dimensional 0.014	Power supply module 0.014	Processor 0.010	Assembling 0.010	Space 0.010	Control system 0.010	Area 0.010	Software 0.010
Steel structure analyzing and detecting management system	Detecting 0.022	Camera 0.017	Engineering 0.015	Steel structure 0.013	Processor 0.013	Analyzing 0.013	Dimensional 0.012	Wireless 0.012	Management system 0.012	Pre 0.011
Energy production apparatus management system	Dimensional 0.019	Formed 0.014	Fixed 0.013	Apparatus 0.013	Part 0.012	Energy 0.012	Producing 0.012	Display 0.011	Management system 0.011	Performance 0.011
Mobile terminal of quantity calculation method management system based on BIM technology	Dimensional 0.033	Utilizing 0.017	Structure 0.017	Engineering 0.013	Mobile terminal 0.012	Updating 0.012	BIM technology 0.012	Form 0.011	Quantity 0.011	Management method 0.011
BIM model monitoring and memory system	Monitoring system 0.023	BIM mode 0.014	Fixed 0.014	Object 0.014	Dimensional 0.013	Forming 0.012	Technology 0.010	Supporting 0.010	Memory 0.010	Creating 0.010
Plate object display processing management system	Object 0.018	Dimensional 0.018	Plate 0.017	Multi 0.016	Extracted 0.015	Processing 0.013	Display 0.013	Managing 0.012	BIM model 0.012	Management system 0.011
Scene simulation analysis based on BIM technology	Dimensional 0.018	Technology 0.016	Device 0.016	BIM technology 0.012	Analyzing 0.012	Simulation 0.012	Scene 0.011	Base 0.011	Utilizing 0.010	Perform 0.010
Device management and control center	BIM model 0.017	Management 0.016	Calculating 0.016	Control center 0.014	Device 0.013	Platform 0.012	Network 0.012	Dimensional 0.011	Equipment 0.010	Fixed 0.010
Construction method of lift operation system based on BIM technology	Constructing method 0.020	Dimensional 0.019	Part 0.017	Lift 0.160	BIM technology 0.015	Calculating 0.012	BIM system 0.012	Operation 0.012	Three 0.012	Attribute information 0.011
Project control management system based on Banknote identification module technology	Project 0.018	File 0.016	Judging 0.015	Controlling 0.013	Base 0.012	Management system 0.012	Dimensional 0.011	BIM technology 0.011	Management 0.011	Banknote identification module 0.010

Figure 7. The perplexity of all BIM patents

After analyzing the corresponding results generated by the LDA model, the keywords under a topic are identified as the features hidden in the large-scale BIM-related technology corpus. For the topic "Steel structure analyzing and detecting management system", for example, it provides a more in-depth analysis of the hotspots technology of G06Q (data processing system or method specifically for administrative, commercial, financial, management, supervisory or predictive purposes). This topic would be further taken as the intellectual base and application hotspots of G06F for the future research. Furthermore, the keyword distribution ("Detecting", "Camera", "Engineering", "Steel structure", "Processor") may have a closely co-occurrence relationship with each other (e.g. the detecting camera may be a favorable tool to obtain the detection image for the position arrangement of steel structures). Similarly, other topics shown in Table 2, extracted using the LDA model, contain qualitative information, such as the keyword distribution of the per-topic.

All in all, it is clear to see that the ten topics reflect the application hotspots and development trend of BIM, and they may become important technologies leading the development of emerging industries in the future.

These right research directions are more than a necessity for scholars and practitioners in this industry to provide some references and help for scholars in the field of BIM research.

In order to compare and verify the results of the LDA topic mining, the Clarivate Analytica's Derwent Innovation, as the world's most authoritative and reliable patent data and powerful analytical capabilities, provides technical support for this research. Derwent Innovation includes patent maps which display data in topographical maps and capture technology topics intelligently. Through the patent map, it assists researchers to understand the status of technology trends, technology branches, and technology relations, which will help determine the direction of mining, inspire mining ideas, stimulate new ideas, avoid patent infringement, and improve the quality of research and development technology. In this way, it can reveal the new technical fields and means, find opportunities for technological development in relatively technology-intensive fields, and then promote innovation activities, promote technology research and development, and finally translate them into corresponding patent achievements.

The patent map is generated based on a co-word analysis. Co-word analysis is a relatively new method of bibliometrics. The main principle of co-word analysis is to count the occurrence of a group of words in the same literature in pairs, and then make clustering analysis on these words, so as to reflect the affinity between these words, and finally analyze the structure and change of the topics that these words represent.

Therefore, this study performs a BIM patent map to analyze the BIM technology hotspots and predict the promising BIM technology flows (Figure 8). Through the cluster analysis of the database and the analysis of quoted figures, this study displayed the hot topics in white and the gray topics in the secondary. In the Figure 8, these topics labeled in white (e.g. "Module connect", "Reinforce steel", "Attribute parameter" and "Scan cloud", etc.) and topics labeled in gray (e.g. "Sensor connect", "Power supply module
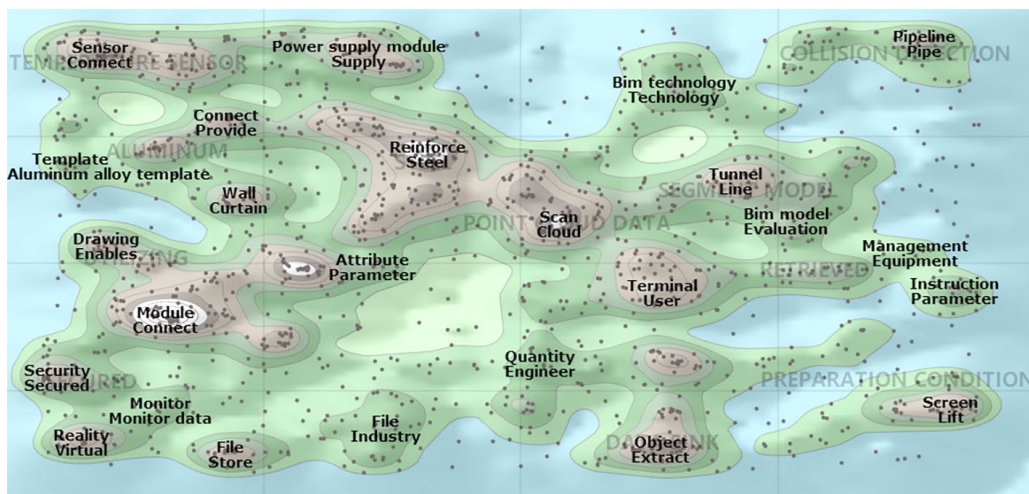

Figure 8. Co-word analysis-based BIM patent map

supply", "Terminal user", "Screen lift", etc.), reveal the research hotspots and evolutionary trends of BIM.

Through the results of LDA topic mining and co-word analysis above, it can be found that most topics are similar and only some topics are different. These topics shown in Figure 8 have close links with these topics shown in Tab. 2. For example, "Power supply module" can be matched with "Power supply module based on BIM technology", "Management equipment" can be matched with "Device management and control center", "Terminal user" can be matched with "Mobile terminal of quantity calculation method management system based on BIM technology", "Monitor data" can be matched with "BIM model monitoring and memory system", "Reinforce steel" can be matched with "Steel structure analyzing and detecting management system", and "Screen lift" can be matched with "Construction method of lift operation system based on BIM technology". Co-word analysis is essentially a co-occurrence analysis method, but LDA is a method of introducing probability distribution and Bayesian prior theory into topic analysis. Thence, there are a few different topics due to the different algorithms.

4. Discussion and limitations

Using patents as data resources, an effective framework integrating citation network analysis and LDA-based topic clustering technology is developed to identify the application hotspots of BIM and forecast its trends. Shown in the results above, it can be seen that the current BIM-based patents are mainly concentrated in different directions. Especially, BIM-related integration technology is increasingly interested by researchers. As is known to all, computer vision is used to detect and track site entities for the extraction of the safety context on a real-time basis and computer vision-based systems are able to automate the monitoring of sites and retrieving contextual information processes (Seo et al., 2015). With the development of computer vision technology, BIM technology will be integrated with it to promote the development of the construction industry. Ye et al. (2018) developed a cup of the theory to represent the integration of BIM, IoT (Internet of Things), and blockchain, which can be used to design an efficient building maintenance system. In the article of Liu et al. (2019), a framework combing blockchain with BIM was proposed for sustainable building design coordination and collaboration in multiple building life cycle stages and solve the challenge of BIM in intellectual property and legal responsibility. Therefore, BIM technology can be integrated with it and this will provide unique opportunities for the development of smart buildings and smart cities.

Furthermore, the GIS-based technologies can deal with a large-scale presentation and centralization of the data with geographical information using real world coordinates (Isikdag, 2015). Particularly, they are relatively competent in 2D geometry modeling and provide mechanisms of multi-representation, such as multiple layers of detailed information (Mignard & Nicolle, 2014). In contrast, BIM is more emphasized on object-oriented modeling of data with complete semantic and typically applied for modeling new buildings and structures and cover both physical and functional characteristics of the subject in 3D representation (Mignard & Nicolle, 2014). Therefore, there is an opportunity for BIM technology to integrate with GIS to manage urban facilities.

Last but not the least, shown in the results above, the significance of BIM technology with complete, accurate, and timely data has been increasingly prominent in recent years. As said by Turk and Klinc (2017), the blockchain technology can potentially prove the development of BIM for its benefits of ensuring the BIM data immutable, traceable, and transparent. As a relatively new concept in the construction industry, the blockchain technology provides a method for creating solutions to the long-standing issue of lack of accountability and fragmented information sources in the construction industry (Sheng et al., 2020). This method can well realize the decentralized information interaction and provides a secure and stable trust environment to improve the collaboration among participants (Li et al., 2019b). From our point of view, the combination of BIM and blockchain technology could offer huge potential and could be seen as one of the future directions of development. In fact, it has already been starting to reveal that the construction industry is increasingly moving toward an advanced, automated, and intelligent way of development in the overall picture.

The research presented in this study, however, is not without its limitations. Firstly, this study identifies application hotspots merely based on LDA topic mining, which might cause a limited exploration of more hidden patterns in the field. In addition, the accuracy of the LDA algorithm can be difficult to justify. In our study, a good clustering approach of patent map demonstrates the effectiveness of the developed LDA model. However, expert knowledge and professional assessments can be the other effective alternative to validate the LDA results. Secondly, there could be some emerging frontiers of BIM research in the patent pool. However, it is very likely that these emerging and cutting-edge patents only stand for a friction of the total population, which might potentially lead to misidentification at current state. Therefore, a continuous monitoring and updating work would be beneficial in the foreseeable future.

Conclusions

Building Information Modeling (BIM) has great potential for increase project productivity and performance. Awareness of BIM application hotspots and forecasting its trends can significantly drive innovations in the construction industry. In this study, patent information extracted from the Derwent Innovations Index (DII) database is a proactive source for analyzing the development trend and application hotspots in the BIM field. Such patent information can provide practitioners with important sources of information that can be used to retrospectively inform BIM services and promote technology innovation in projects.

This study aims to identify the application hotspots of BIM and forecast its trends in an effective framework integrating citation network analysis and LDA-based topic clustering. This framework comprises three-step analysis: (1) quantitative characteristic analysis of patent outputs to provide a visual overview of BIM patents; (2) the production of a Social Network Analysis (SNA)-based co-occurrence network analysis to determine the citation interrelations between BIM patents and (3) the identification of BIM patent topics using a Latent Dirichlet Allocation (LDA) to further analysis the topic characters of critical patents. For this study, the 1510 patents gathered over sixteen years (2004–2020) are analyzed.

In this study, the characteristics of BIM patents outputs show that there has been a significant increase in the number of patent applications since 2010 and that BIM serves can be as a more and more important and useful tool in facilitating the construction for its benefits of powerful management of physical and functional digital presentations. The patented current version of International Patent Classification (IPC) is mainly included in the G06F (electrical digital data processing) classification meaning the application hotspots of G06F is the core technology. The "US20150248503A1", the patent code of "a method for creating a three-dimensional (3D) model of two-dimensional (2D) data of building information modeling (BIM) and the system", has the highest degree of citation in the citation network, considered as the application hotspots in the BIM field. Besides, the co-occurrence citation interrelations of the important patents (such as the "US20120310906A1" and "US20150248503A1") are identified. The results indicate that the SNA-based network analysis can be effectively used to identify important patent codes relating to BIM (such as: "US20150248503A1" and "US20150057981A1"), as well as the co-occurrence citation interrelations of the important patents. For in-depth analyzing of BIM application hotspots, ten topics ("Power supply module based on BIM technology", "Steel structure analyzing and detecting management system", "Energy production apparatus management system" and others) are identified based on the LDA clustering model. These topics and correspondent keywords represent the current trends of the BIM field. For example, integrating BIM technology with other technologies will promote the development of the construction industry. Blockchain provides a method for creating solutions to the long-standing issue of lack of incomplete, inaccurate information sources when using BIM in the construction field. Besides, a good clustering approach of patent map demonstrates the effectiveness of the developed LDA model.

To the end, the contributions of this study are threefold: (1) an innovative text mining-based framework for BIM patent analysis in construction is developed; (2) patents that have focused on identifying the application hotspots and development trend of BIM in accordance with our developed framework are reviewed; and (3) a signpost for technological development and innovation of BIM is provided.

Conflict of interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted. We declare that the work presented is original research that has not been published previously, and is not under consideration for publication elsewhere, in whole or in part.

References

Al Hattab, M., & Hamzeh, F. (2015). Using social network theory and simulation to compare traditional versus BIM-lean practice for design error management. *Automation in Construction*, *52*, 59–69. https://doi.org/10.1016/j.autcon.2015.02.014

Altwies, J. E., & Nemet, G. F. (2013). Innovation in the U.S. building sector: An assessment of patent citations in building energy control technology. *Energy Policy*, *52*, 819–831. https://doi.org/10.1016/j.enpol.2012.10.050

Ardito, L., Adda, D., & Petruzzelli, A. M. (2017). Mapping innovation dynamics in the internet of things domain: Evidence from patent analysis. *Technological Forecasting and Social Change*, *136*, 317–330. https://doi.org/10.1016/j.techfore.2017.04.022

Azhar, S. (2011). Building information modeling (BIM): trends, benefits, risks, and challenges for the AEC industry. *Leadership & Management in Engineering*, *11*(3), 241–252. https://doi.org/10.1061/(ASCE)LM.1943-5630.0000127

Batool, K., & Niazi, M. A. (2014). Towards a methodology for validation of centrality measures in complex networks. *Plos One*, *9*(4), 90283. https://doi.org/10.1371/journal.pone.0090283

Bastani, K., Namavari, H., & Shaffer, J. (2018). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Systems with Applications*, *127*, 256–271. https://doi.org/10.1016/j.eswa.2019.03.001

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: a review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Boshnakoska, D., Chorbev, I., & Davcev, D. (2013). Ontology supported patent search architecture with natural language analysis and fuzzy rules. *Advances in Intelligent Systems and Computing*, *207*, 275–284. https://doi.org/10.1007/978-3-642-37169-1_27

Daim, T., Lai, K. K., Yalcin, H., Alsoubie, F., & Kumar, V. (2020). Forecasting technological positioning through technology knowledge redundancy: Patent citation analysis of IoT, cybersecurity, and Blockchain. *Technological Forecasting and Social Change*, *161*, 120329. https://doi.org/10.1016/j.techfore.2020.120329

Cai, F., Ji, J. M., Jiang, Z. Q., Mu, Z. R., Wu, X., Zheng, W. J., Zhou, W. X., Tu, S. T., & Qian, X. H. (2018). Engineering fronts in 2018. *Engineering*, *4*, 748–753. https://doi.org/10.1016/j.eng.2018.11.004

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, *28*(4), 1661–1707. https://doi.org/10.3386/w3301

Han, Q., Heimerl, F., Codinafilba, J., Lohmann, S., Wanner, L., & Ertl, T. (2017). Visual patent trend analysis for informed decision making in technology management. *World Patent Information*, *49*, 34–42. https://doi.org/10.1016/j.wpi.2017.04.003

Jia, J., Sun, J., Wang, Z., & Xu, T. (2017). The construction of BIM application value system for residential buildings' design stage in china based on traditional DBB mode. *Procedia Engineering*, *180*, 851–858. https://doi.org/10.1016/j.proeng.2017.04.246

Johansson, M., Roupé, M., & Bosch-Sijtsema, P. (2015). Real-time visualization of building information models (BIM). *Automation in Construction*, *54*, 69–82. https://doi.org/10.1016/j.autcon.2015.03.018

Hatala, J.-P., & Lutta, J. G. (2009). Managing information sharing within an organizational setting: A social network perspective. *Performance Improvement Quarterly*, *21*(4), 5–33. https://doi.org/10.1016/j.autcon.2015.03.018

Ju, Y., & Sohn, S. Y. (2015). Patent-based QFD framework development for identification of emerging technologies and related business models: A case of robot technology in Korea. *Technological Forecasting and Social Change*, *94*, 44–64. https://doi.org/10.1016/j.techfore.2014.04.015

Kang, B. (2014). The innovation process of Huawei and ZTE: Patent data analysis. *China Economic Review*, *36*, 378–393. https://doi.org/10.1016/j.chieco.2014.12.003

Kim, G., & Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change*, *117*, 228–237. https://doi.org/10.1016/j.techfore.2016.11.023

Kim, Y. G., Suh, J. H., & Park, S. C. (2008). Visualization of patent analysis for emerging technology. *Expert Systems with Applications*, *34*(3), 1804–1812. https://doi.org/10.1016/j.eswa.2007.01.033

Kim, B. S., Chang, S., & Suh, Y. (2018). Text analytics for classifying types of accident occurrence using accident report documents. *Journal of the Korean Society of Safety*, *33*(3), 58–64. https://doi.org/10.14346/JKOSOS.2018.33.3.58

Krebs, V. E. (2002). Mapping networks of terrorist cells. *Connections*, *24*(3), 43–52.

Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, *48*(2), 354–368. https://doi.org/10.1016/j.dss.2009.09.003

Li, X., Xie, Q., Daim, T., & Huang, L. (2019a). Forecasting technology trends using text mining of the gaps between science and technology: The case of perovskite solar cell technology. *Technological Forecasting and Social Change*, *146*, 432–449. https://doi.org/10.1016/j.techfore.2019.01.012

Li, J., Greenwood, D., & Kassem, M. (2019b). Blockchain in the built environment and construction industry: A systematic review, conceptual models and practical use cases. *Automation in Construction*, *102*, 288–307. https://doi.org/10.1016/j.autcon.2019.02.005

Li, X., Fan, M., Zhou, Y., Fu, J., Yuan, F., & Huang, L. (2020). Monitoring and forecasting the development trends of nanogenerator technology using citation analysis and text mining. *Nano Energy*, *71*, 104636. https://doi.org/10.1016/j.nanoen.2020.104636

Liu, Z., Jiang, L., Osmani, M., & Demian, P. (2019). Building Information Management (BIM) and Blockchain (BC) for sustainable building design information management framework. *Electronics*, *8*(7), 724. https://doi.org/10.3390/electronics8070724

Lu, Y., Wu, Z., Chang, R., & Li, Y. (2017). Building Information Modeling (BIM) for green buildings: A critical review and future directions. *Automation in Construction*, *83*, 134–148. https://doi.org/10.1016/j.autcon.2017.08.024

Isikdag, U. (2015). BIM and IoT: A synopsis from GIS perspective. *The International Archives of the Photogrammetry*, *10*, 28–30. https://doi.org/10.5194/isprsarchives-XL-2-W4-33-2015

Markatou, M., & Vetsikas, A. (2015). Innovation and crisis: An analysis of its impact on the Greek patenting activity. *Procedia-Social and Behavioral Sciences*, *195*, 123–132. https://doi.org/10.1016/j.sbspro.2015.06.419

Mcaleer, M., & Slottje, D. (2005). A new measure of innovation: The patent success ratio. *Scientometrics*, *63*(3), 421–429. https://doi.org/10.1007/s11192-005-0222-2

Miettinen, R., & Paavola, S. (2014). Beyond the BIM utopia: Approaches to the development and implementation of building information modeling. *Automation in Construction*, *43*(7), 84–91. https://doi.org/10.1016/j.autcon.2014.03.009

Mignard, C., & Nicolle, C. (2014). Merging BIM and GIS using ontologies application to urban facility management in ACTIVe3D. *Computers in Industry*, *65*(9), 1276–1290. https://doi.org/10.1016/j.compind.2014.07.008

Mohd, S., & Ahmad Latiffi, A. (2013). Building Information Modeling (BIM) application in construction planning. In *Seventh International Conference on Construction in the 21st Century* (*CITC-VII*), Bangkok, Thailand. https://doi.org/10.14455/ISEC.res.2014.79

Park, C. S., Lee, D.-Y., Kwon, O.-S., & Wang, X. (2013). A framework for proactive construction defect management using BIM, augmented reality and ontology-based data collection template. *Automation in Construction*, *33*, 61–71. https://doi.org/10.1016/j.autcon.2012.09.010

Perng, Y. H., & Huang, Y. Y. (2016). Investigation of technological trends in shading devices through patent analysis. *Journal of Civil Engineering and Management*, *22*(6), 818–830. https://doi.org/10.3846/13923730.2014.914091

Pilkington, A., Lee, L. L., Chan, C. K., & Ramakrishna, S. (2009). Defining key inventors: A comparison of fuel cell and nanotechnology industries. *Technological Forecasting and Social Change*, *76*(1), 118–127. https://doi.org/10.1016/j.techfore.2008.03.015

Seo, J. O., Han, S. U., Lee, S. H., & Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, *29*(2), 239–251. https://doi.org/10.1016/j.aei.2015.02.001

Sharma, P., & Tripathi, R. C. (2017). Patent citation: a technique for measuring the knowledge flow of information and innovation. *World Patent Information*, *51*, 31–42. https://doi.org/10.1016/j.wpi.2017.11.002

Sheikh, N. J., & Sheikh, O. (2017). Forecasting of biosensor technologies for emerging point of care and medical IoT applications using bibliometrics and patent analysis. *Portland International Conference on Management of Engineering and Technology* (pp. 3082–3093), Portland, USA. https://doi.org/10.1109/PICMET.2016.7806585

Sheng, D., Ding, L., Zhong, B. T., Love, P. E. D., & Chen, J. G. (2020). Construction quality information management with blockchains. *Automation in Construction*, *120*, 103373. https://doi.org/10.1016/j.autcon.2020.103373

Sohrabi, B., Vanani, I. R., & Abedin, E. (2018). Human resources management and information systems trend analysis using text clustering. *International Journal of Human Capital and Information Technology Professionals*, *9*(3). https://doi.org/10.4018/ijhcitp.2018070101

Turk, Z., & Klinc, R. (2017). Potentials of blockchain technology for construction management. *Procedia Engineering*, *196*, 638–645. https://doi.org/10.1016/j.proeng.2017.08.052

Volk, R., Stengel, J., & Schultmann, F. (2014). Building Information Modeling (BIM) for existing buildings—literature review and future needs. *Automation in Construction*, *38*, 109–127. https://doi.org/10.1016/j.autcon.2013.10.023

Wang, Y. B., & Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems*, *105*, 87–95. https://doi.org/10.1016/j.dss.2017.11.001

Yang, J. M., Liao, W. C., Wu, W. C., Yin, & C. Y. (2009). Trend analysis of machine learning-A text mining and document clustering methodology. In *Proceedings-2009 International Conference on New Trends in Information and Service Science, NISS 2009* (pp. 481–485). https://doi.org/10.1109/NISS.2009.176

Ye, Z., Yin, M., Tang, L., & Jiang, H. (2018). Cup-of-Water theory: A review on the interaction of BIM, IoT and blockchain during the whole building lifecycle. In *Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC 2018)*, Berlin, Germany. https://doi.org/10.22260/isarc2018/0066

Yoon, J., Choi, S., & Kim, K. (2011). Invention property-function network analysis of patents: A case of silicon-based thin film solar cells. *Scientometrics*, *86*(3), 687–703. https://doi.org/10.1007/s11192-010-0303-8

You, H., Li, M., Hipel, K. W., Jiang, J., Ge, B., & Duan, H. (2017). Development trend forecasting for coherent light generator technology based on patent citation network analysis. *Scientometrics*, *111*(1), 297–315. https://doi.org/10.1007/s11192-017-2252-y

Zhang, L. (2011). Identifying key technologies in Saskatchewan, Canada: Evidence from patent information. *World Patent Information*, *33*(4), 364–370. https://doi.org/10.1016/j.wpi.2011.06.002

Zhang, Y., Guo, S. L., Han, L.N., & Li, T. L. (2016). Application and exploration of big data mining in clinical medicine. *Journal of China Medicine*, *129*(6), 731–738. https://doi.org/10.4103/0366-6999.178019

Zhong, B. T., Hei, Y. J., Li, H., Rose, T., & Luo, H. B. (2019). Patent cooperative patterns and development trends of Chinese construction enterprises: A network analysis. *Journal of Civil Engineering and Management*, *25*(3), 228–240. https://doi.org/10.3846/jcem.2019.8137

Zhong, B. T., Pan, X., Love, P. E. D., Ding, L. Y., & Fang, W. (2020). Deep learning and network analysis: classifying and visualizing accident narratives in construction. *Automation in Construction*, *113*, 103089. https://doi.org/10.1016/j.autcon.2020.103089