



A multi-association rule-based knowledge graph method for exploring complex research trends in BIM

Gozde Basak Ozturk^{1,*}, Fatih Soygazi², Gamze Gulsun Gazi¹

¹ Aydin Adnan Menderes University, Department of Civil Engineering, Aydın, Türkiye

² Aydin Adnan Menderes University, Department of Computer Engineering, Aydın, Türkiye

* Corresponding author: G. B. Ozturk (gbozturk@adu.edu.tr)

<https://doi.org/10.31462/jcemi.2026.710>

Received 13 April 2026; Revised 16 July 2026; Accepted 8 September 2026; Available online 28 September 2026

Keywords

Association rule mining
Building information modelling
Knowledge graph
Scientometrics
Machine learning

Abstract

Traditional scientometric analysis methods primarily rely on binary keyword associations, which limits the ability to capture complex research structures. This study proposes a multi-association rule-based method (MAR) that enables the extraction of higher-order (n-ary) associations among keywords to address the stated limitation. The method integrates the Apriori algorithm with knowledge graph visualization to quantitatively analyse multi-keyword associations. The proposed method is applied to the Building Information Modelling (BIM) research domain. The bibliometric data, comprising 12,185 publications, is retrieved from the Scopus database following a systematic refinement of the initial search results to the architecture, engineering, construction, and operation (AECO)-relevant literature. The results demonstrate that the MAR method effectively identifies complex research patterns. The framework advances traditional scientometric tools by enabling the identification of multi-keyword semantic associations with quantified strength metrics (support, confidence, lift). Emerging trends revealed by the regression analysis include the convergence of BIM with generative artificial intelligence, real-time data integration, explainable AI, and human-centric approaches. The findings provide a scalable and analytical approach for identifying interdisciplinary research trends and gaps. The proposed method enables a deeper understanding of complex knowledge structures while supporting strategic decision-making and technology adoption.

1. Introduction

The management of construction projects has become increasingly challenging due to their growing scale and complexity. Building Information Modelling (BIM) has emerged as a key solution by providing a shared digital representation of the physical and functional characteristics of any built assets to support decision-making throughout the project lifecycle [1, 2]. BIM enables stakeholders to identify and resolve issues during all project phases by facilitating collaboration across architectural, engineering, construction, operations, and facilities management (AECO-FM) [3, 4].

Artificial intelligence (AI), point cloud, 3D printing, and 3D laser scanning have all been integrated with BIM within the past ten years [5-12]. The rapid integration of innovative technologies has accelerated the growth of BIM literature, making it increasingly difficult to manually track research developments. In response, various scientometric analysis and visualization tools have been developed to enable researchers to map the research field [13-15]. BIM is one of

the most extensively studied topics in construction engineering and management, yet its rapid interdisciplinary expansion has introduced a distinctive analytical challenge for scientometric research. As BIM integrates with AI, IoT, digital twins, sustainability frameworks, and robotics, the keyword space of BIM literature grows not only in volume but in structural complexity. Emerging themes appear not as isolated terms but as multi-concept clusters that co-evolve across publication years. Existing BIM scientometric studies [16-23] have documented this growth through binary keyword co-occurrence analysis. However, these approaches are inherently limited to pairwise relationships, which can identify that two keywords frequently appear together, but cannot capture how three or more concepts form a compound research pattern, nor can they quantify the directional strength of such associations. As BIM research matures into an increasingly interconnected domain, the analytical tools used to map it must evolve accordingly. This study addresses that gap by proposing the MAR method, which enables the

extraction of n-ary keyword associations and their temporal modelling.

Scientometrics aims to map research landscapes at scale without requiring direct reading of individual publications. Since full-text analysis of tens of thousands of papers is computationally and practically infeasible, the field operates on structured metadata (authors, citations, journals, and keywords) that are consistently available across large bibliometric databases. Keywords are particularly central to this approach because they serve as the most compact structural representation of a publication's conceptual focus, encapsulating both the author's own framing of the work (author keywords) and the database's thematic classification identified via machine learning algorithms (index keywords). Within the BIM domain, this keyword-based scientometric approach has been operationalised through tools such as VOSviewer [13], CiteSpace [14], and Pajek [15] to map co-occurrence networks, identify research clusters, and visualise collaboration patterns [16–23].

The prior scientometric literature on BIM reveals both the breadth of the domain and a consistent methodological constraint that motivates the present study. Li et al. [17] provided a foundational bibliometric mapping of BIM knowledge domains through keyword co-occurrence, establishing the analytical template that much subsequent work would follow. Jin et al. [18] extended this approach to BIM-based construction engineering and management research, offering valuable cluster-level insights into the field's thematic structure. Vilutiene et al. [19] narrowed the lens to BIM in structural engineering, demonstrating the value of domain-specific bibliometric analysis, while Ozturk [22, 23] traced interoperability trends across authors and articles. Taken together, these studies document a research frontier that has moved steadily toward integration rather than toward isolated technical refinement, a trajectory further reflected in work coupling BIM with big data analytics and lean facility management [10, 24], BIM-enabled digital transformation of infrastructure projects [11, 25, 26], the semantic and data-integration dimensions of BIM through IoT and GIS [27], BIM-supported digital twin applications for sustainable facility management [28], and the positioning of BIM within cognitive computing and generative AI paradigms [11, 12]. Despite this breadth, all of these studies share a fundamental analytical limitation: they operate on pairwise (binary) keyword co-occurrence, which can identify that two keywords frequently appear together but cannot capture how three or more concepts form a compound, higher-order research pattern, nor quantify the directional strength of such associations. As BIM research matures into an increasingly interconnected domain, the scientometric methods used to map it must evolve accordingly. It is precisely this analytical gap that the proposed multi-association rule-based (MAR) method addresses.

Association rule mining (ARM) is selected as the primary analytical approach in this study because it uniquely enables

the extraction of n-ary ($n \geq 2$) keyword associations with quantifiable co-occurrence strength metrics (support, confidence, and lift). ARM has been widely applied in transactional data mining [29–33], but its application to keyword-level scientometric analysis to model multi-dimensional research trends remains largely unexplored. This constitutes the core methodological contribution of the present study.

This study proposes a multi-association rule-based (MAR) method that captures n-ary ($n \geq 2$) keyword associations. The method combines association rule mining with knowledge graph (KG) visualization. The novelty of the MAR framework lies in three specific contributions: (1) it extends scientometric analysis from binary to n-ary keyword associations, enabling the identification of hidden, high-order research patterns; (2) it formalises a graph-theoretic representation of multi-keyword rules as directed edges in a BIM Knowledge Graph (BIM-KG), providing a semantically enriched and structured model of research knowledge; and (3) it introduces temporal regression over lift values as an exploratory mechanism for analysing the evolution of keyword association strength over time. The resulting knowledge graph shows interconnected keyword associations and enables the identification of complex research patterns not accessible through traditional scientometric analysis (TSA) tools [13–15]. The scope of the paper does not cover the comparison of the current approaches and the proposed approach. Therefore, the paper will not present any statistical comparison among all scientometric tools. The need for more expressive multi-association analyses in the BIM domain leads this study to pose the following research questions:

- o RQ1. How can scientometric analysis be extended beyond binary co-occurrence to capture n-ary ($n \geq 2$) associations among keywords using multi-association rule mining?
- o RQ2. How can association rule mining (e.g., Apriori-based MAR) be effectively integrated with knowledge graph representation to enable structured and interpretable modeling of complex semantic associations in BIM literature?
- o RQ3. To what extent can a scalable multi-association rule framework support interdisciplinary bibliometric analysis and uncover hidden thematic structures and research clusters in large-scale datasets?
- o RQ4. How do multi-metric association measures (support, confidence, and lift) enhance the quantitative interpretation of research dynamics and keyword interdependencies in scientometric studies?
- o RQ5. What insights can the proposed framework reveal regarding the evolution of BIM research, particularly in emerging technologies and interdisciplinary domains?
- o RQ6. How can temporal interpretation of association rule outputs support the analysis of evolving research associations and provide a foundation for future predictive scientometric studies?

The remainder of this paper is organized as follows. Section 2 examines methodological background explaining association rule-based method and knowledge graphs. Section 3 explains layer by layer methodology application of the hybrid approach. In Section 4, findings of applied multi-association rule-based (MAR) method steps are presented. Binary and n-ary associations are represented as a summary of the findings of the BIM literature data analysis. Section 5 discusses the trends and gaps of the BIM research field, stating prospects and emphasizing limitations. Section 6 concludes with the proposed justification of the proposed solutions, main contributions of the research, and highlighting limitations to suggest a new perspective for future studies.

2. Methodological Background

2.1. Multi-association rule mining

Association rule mining is a data mining method for examining data from the past to support future research. The expression for the association rule is $X \rightarrow Y$. Y is quite likely to be present in a transaction if X is present [29, 30]. Association rules extract unidentified associations and generate outcomes that can serve as the foundation for prediction and decision-making [31, 32]. The prerequisite for finding rules via association rule mining is to set minimum threshold for support and confidence values. The association rule mining has a two-step process: (1) finding all frequent keywords, and (2) generating strong association rules from the frequent keywords. The first step includes generation of all keywords set whose support values are larger than predetermined minimum support count. The second step includes generation of high confidence rules from each frequent keyword set, where each rule is a binary partitioning of a frequent keywords set.

Support: The support (supp) value indicates the significance of an association pattern in the transaction population. The support value is determined by the likelihood that the association pattern is present in the process population [31].

$$\text{Supp}(X) = P(X) \quad (1)$$

Confidence: The probability in the transaction population that includes both X and Y is the confidence (conf) value of the $X \rightarrow Y$ rule [31].

$$\text{Conf}(X \rightarrow Y) = \frac{\text{Supp}(X \cup Y)}{\text{Supp}(X)} = P(Y | X) \quad (2)$$

Lift: The lift value is a measure that determines how much more often the antecedent (A) and consequent (B) of a rule occur together than to be expected if they were statistically independent. Lift is calculated as the ratio of the observed

support of A and B occurring together to the support that would be expected if they were independent [31].

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Supp}(X \cup Y)}{\frac{(\text{Supp}(X) \times \text{Supp}(Y))}{\text{Supp}(Y | X)}} = \frac{\text{Supp}(X \cup Y)}{\text{Supp}(Y)} \quad (3)$$

where $\text{Supp}(\cdot)$ denotes support and $P(\cdot)$ represents probability in the transaction dataset.

Interpretation of values;

If $\text{Lift} > 1$, X and Y are positively associated.

If $\text{Lift} < 1$, X and Y are negatively associated.

If $\text{Lift} = 1$, X and Y are statistically independent.

The Apriori algorithm identifies frequently used keywords and extracts association rules between them. Therefore, the Apriori algorithm was used in this research. The contribution of this study is to use this algorithm for knowledge graphs [34], whereas the Apriori algorithm is commonly used for tabular data.

2.2. Knowledge graph (KG)

A knowledge graph is a directed labelled graph in which the labels carry well-defined meanings. It consists of nodes, edges, and labels, where nodes may represent any entity of interest (e.g., people, organizations, or concepts), edges connect pairs of nodes to capture the relationship between them, and labels specify the semantic meaning of that relationship [35, 36]. More formally, given a set of nodes N and a set of labels L, a knowledge graph is a subset of the cross-product $N \times L \times N$, where each member of this set is referred to as a triple. In the context of this study, the nodes correspond to BIM-related keywords extracted from the bibliometric dataset, the edges correspond to the association rules derived through the Apriori algorithm, and the labels correspond to the direction and strength of these associations as quantified by the support, confidence, and lift metrics (Fig. 1). Constructing the KG from multi-association rules rather than from pairwise co-occurrence frequencies is what enables it to represent n-ary ($n \geq 2$) relations: a single rule such as $\{\text{information theory, sustainability}\} \rightarrow \{\text{BIM, sustainable development}\}$ is preserved as a coherent multi-keyword structure within the graph, rather than being fragmented into isolated keyword pairs, thereby retaining the compound semantic patterns that characterise interdisciplinary research domains. Because its applications are built on learning from nodes and the relationships between them, a KG supports a variety of downstream operations such as association extraction and KG-completion, which has attracted considerable research interest [37, 38]. In practice, KGs function as an evolving layer of data within a community or organization.

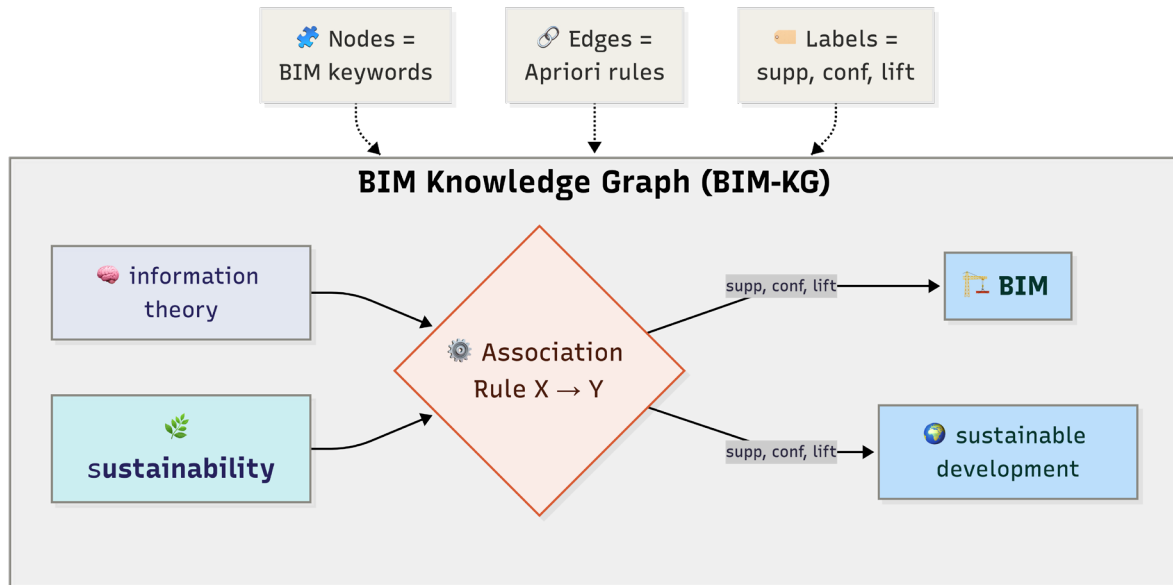


Fig. 1. Conceptual structure of the BIM Knowledge Graph (BIM-KG) representing n-ary keyword associations

3. Methodology

3.1. Proposed solution (Apriori $\rightarrow$ MAR $\rightarrow$ Knowledge graph)

The hybrid approach integrates the Apriori Algorithm, Multi-Association Rules (MAR), and Knowledge Graph construction to systematically transform unstructured BIM data into structured knowledge (Fig. 2). The Apriori

Algorithm iteratively generates and prunes candidate sets based on support thresholds. Frequent keyword-sets are generated and extended into multi-association rules. The extracted rules are mapped into a graph structure. In this graph, keywords form nodes and associations form edges. The output is a BIM Knowledge Graph (BIM-KG) that supports advanced semantic analysis, pattern discovery, and decision-making.

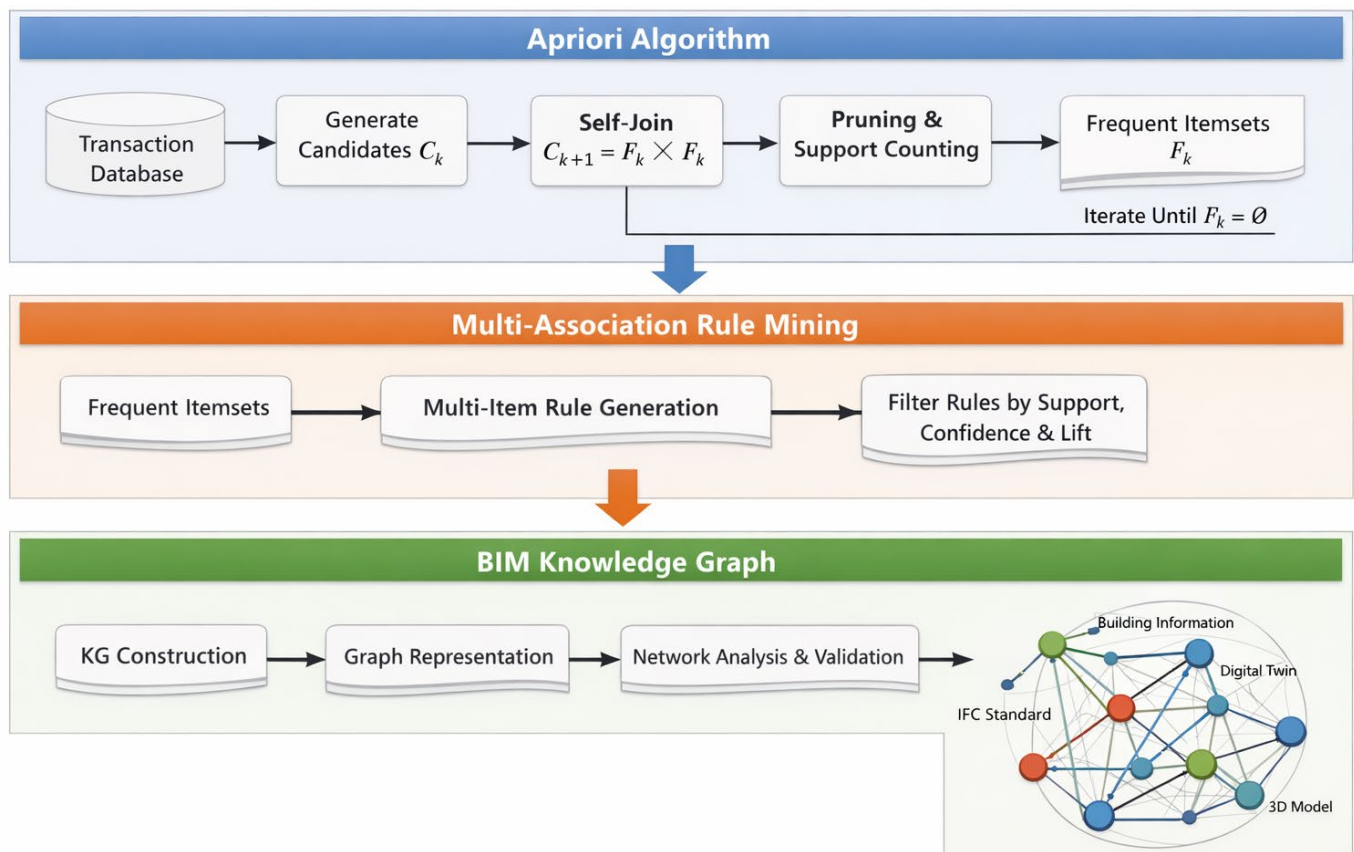


Fig. 2. Hybrid approach (Apriori $\rightarrow$ MAR $\rightarrow$ Knowledge Graph)

3.2. Methodology phases

This study presents an association rule mining method that is applied on a BIM Knowledge Graph (BIM-KG). The proposed multi-association rule-based method includes five sequential steps (Fig. 3). First of all, the Data Acquisition phase gathers raw bibliometric data from Scopus database.

The Data Preprocessing phase consists of cleaning and normalizing the data such as; merging synonymous keywords. The Creating BIM Knowledge Graph (BIM-KG) phase is for representing keywords with nodes and representing patterns with edges. The Extracting Associations is realized for n-ary ($n \geq 2$) association rules derivation. Through the application of The Apriori Algorithm, association rules are measured by support, confidence, and lift metrics (Eqs. 1-3). At the last step, the Knowledge Graph Visualization displays derived associations as an interactive network, enabling the examination of complex multi-keyword associations, and research trends. This systematic workflow expands upon traditional binary scientometric analysis by recognizing high level associations in the BIM literature. In this research, NLTK [39], NumPy [40], pandas

[41], matplotlib [42], Apyori [43], and arulesViz [44] libraries were used to perform scientific calculations, unify semantically equivalent keywords, convert the normalized data into a rule-based method, and apply visualization techniques.

3.2.1. Data acquisition

BIM research data was acquired from the Scopus database. Scopus was selected because it is the most widely used database in BIM scientometric research [17, 22, 23] and provides comprehensive coverage of construction engineering and management publications with consistent structured metadata. Both author keywords and index keywords were used to maximise keyword coverage. Fig. 4 summarizes the bibliometric search input for the field of BIM research. The consolidated dataset of 12,185 publications, representing 33.70% of the initial 36,163 records retrieved for the term "BIM", provides sufficient domain coverage for association rule-based scientometric analysis. This retention rate should be understood as the outcome of a deliberate scoping process rather than as a limitation of coverage.

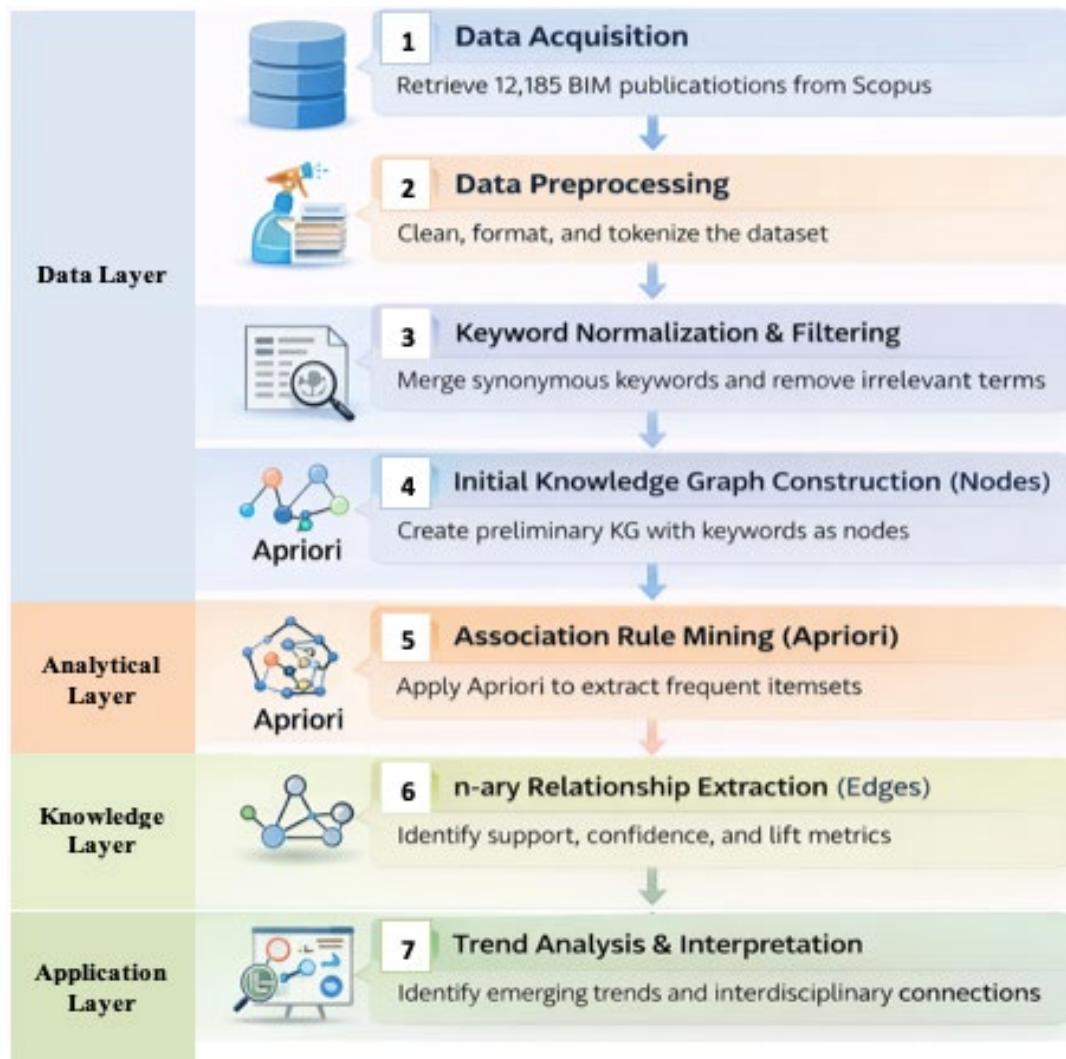


Fig. 3. The workflow of the multi-association rule-based method for BIM knowledge graph construction and analysis

```
{BIM} OR {building information modeling} OR {building information modelling} AND AND (
LIMIT-TO ( LANGUAGE,"English" ) ) AND ( LIMIT-TO ( SRCTYPE,"j" ) ) AND ( LIMIT-
TO ( DOCTYPE,"ar" ) ) AND ( LIMIT-TO ( SUBJAREA,"ENGI" ) OR LIMIT-TO (
SUBJAREA,"COMP" ) OR LIMIT-TO ( SUBJAREA,"SOCI" ) OR LIMIT-TO (
SUBJAREA,"ENVI" ) OR LIMIT-TO ( SUBJAREA,"ECON" ) OR EXCLUDE (
SUBJAREA,"PHYS" ) ..... )
```

Fig. 4. Research input summary

The initial unfiltered pool contained a substantial number of records in which the "BIM" abbreviation refers to concepts unrelated to the AECO industry, which introduce semantic noise into keyword co-occurrence structures. The systematic refinement process removes these ambiguous records and yields a domain-coherent subset directly relevant to the AECO context of this study. This is methodologically significant for association rule mining, because support, confidence, and lift are computed as proportions of the total transaction population; a larger but noisier corpus produces a higher incidence of spurious associations, whereas a systematically curated corpus yields more reliable and interpretable rules. Prior BIM scientometric studies have drawn conclusions from considerably smaller datasets: Li et al. [17] analysed 2,813 publications, Jin et al. [18] worked with 1,469 records, and Vilutiene et al. [19] examined portable datasets of similar scale. At 12,185 records, the present dataset is among the largest employed in BIM keyword-level scientometric analysis to date, providing a robust empirical foundation for the extraction of statistically meaningful association rules. The 33.70% retention rate following systematic refinement is not a limitation but a methodological strength: it reflects a principled scoping of the analysis to publications directly relevant to the AECO industry context of BIM, eliminating noise introduced by the ambiguous use of the BIM abbreviation across unrelated disciplines. A dataset that is large, domain-coherent, and systematically curated provides a more reliable basis for association rule extraction than a larger but noisier corpus would.

3.2.2. Data preprocessing

Since, the unwanted keywords are the keywords that cause ambiguity, an initial step involves the removal of unwanted keywords from the analysis. For instance, "BIM-Business Information Management" is an unwanted keyword in "BIM-Building Information Modelling" search. Because of this, a feature has been developed that enables the user to identify the keywords that contribute to data noise. A rule-based filtering mechanism was applied to remove ambiguous or irrelevant keywords (e.g. BIM – Business Information Management). The path to the reference file may optionally be left unspecified; in such a case, no filtering is applied. The unwanted keywords specified in the reference file are extracted, and any unnecessary space characters within these entries are removed. Subsequently, the complete set of candidate keywords is compared against the list of unwanted

terms. Any keyword found in the unwanted list is excluded from further analysis.

3.2.3. Keyword normalization and filtering

Keywords with the same meaning but various spellings can lead to data loss in keyword-based scientometric analysis, which can have a negative impact on accuracy. Because of this, several synonymous terms that are useful for the BIM literature are integrated in this study. All keywords were acquired from the Scopus database. Keywords with the same meaning but different spellings are refined to a single keyword. "BIM" as an abbreviation cannot be consolidated into a specific phrase because it can mean both "building information modelling" and "building information model" as shown below.

```
keyword_map = {"building information modelling":
"building information modelling (BIM)", "building
information modeling": "building information modeling
(BIM)", "building information modeling (bim)": "building
information modelling (bim)". . . }
```

3.2.4. Initial BIM knowledge graph construction

The Apriori algorithm is employed to extract association rules from the BIM keyword dataset. Prior to this step, the data is preprocessed through cleaning and synonym normalization to ensure consistency and reduce information loss. Following the preprocessing process, a knowledge graph (KG) is constructed. Nodes as keywords and their co-occurrence patterns as associations are represented in the BIM-KG [29]. Apriori-based pruning enables extraction of significant associations. This process facilitates the extraction of complex multi-keyword associations which forms the basis for subsequent knowledge graph analysis.

3.2.5. Association rule mining with the apriori algorithm

The Apriori algorithm, as outlined in the pseudocode below [33], reveals the association rules between the keywords and keyword groups. It is used in this study to examine the keywords in the BIM domain. In the data preprocessing stage data is prepared to create a keywords knowledge graph by cleaning the data and synchronizing synonyms. This preliminary preparation process is for preventing the loss of value as much as possible by considering the presence of keywords in the data and their relevance to the subject. After the data was made to be operable, a KG was created by using the Apriori Algorithm. The Apriori Algorithm has succeeded in revealing the associations between each other by seeing the keyword or keyword groups in the data as items.

Input: D – transaction database (each transaction = keyword set of one publication)
 min_supp – minimum support threshold
 min_lift – minimum lift threshold (1.0)
 max_conf – maximum confidence threshold (0.95)
Output: R – set of association rules

```

1:  $C_1 \leftarrow \{ \{i\} \mid i \in I \}$  // candidate 1-keyword-sets
2:  $F_1 \leftarrow \{ c \in C_1 \mid supp(c) \geq min\_supp \}$  // frequent 1-keyword-sets
3:  $k \leftarrow 2$ 
4: while  $F_{k-1} \neq \emptyset$  do
5:    $C_k \leftarrow generate\_candidates(F_{k-1})$  // join step
6:   for each  $c \in C_k$  do
7:     if any  $(k-1)$ -subset of  $c \notin F_{k-1}$  then
8:        $C_k \leftarrow C_k \setminus \{c\}$  // prune step (Apriori property)
9:     end if
10:  end for
11:  for each transaction  $t \in D$  do
12:    for each  $c \in C_k$  with  $c \subseteq t$  do
13:       $count(c) \leftarrow count(c) + 1$ 
14:    end for
15:  end for
16:   $F_k \leftarrow \{ c \in C_k \mid supp(c) \geq min\_supp \}$  // frequent  $k$ -keyword-sets
17:   $k \leftarrow k + 1$ 
18: end while
19:  $F \leftarrow \bigcup_k F_k$  // all frequent keyword-sets
20:  $R \leftarrow \emptyset$ 
21: for each  $f \in F$  with  $|f| \geq 2$  do
22:   for each non-empty proper subset  $X \subset f$  do
23:      $Y \leftarrow f \setminus X$  //  $X \rightarrow Y$  is a candidate rule
24:      $conf \leftarrow supp(f) / supp(X)$ 
25:      $lift \leftarrow supp(f) / (supp(X) \cdot supp(Y))$ 
26:     if  $lift \geq min\_lift$  and  $conf \leq max\_conf$  then
27:        $R \leftarrow R \cup \{ X \rightarrow Y \}$ 
28:     end if
29:   end for
30: end for
31: return  $R$ 

```

Apriori's pruning method pruned the KG by pruning the associations between keywords and this pruning continued until the lowest threshold was determined. In this way, the association rules that remain hidden in the data could also be determined.

Initially, the support value of each keyword is calculated and listed. This is called the 'frequent one-keywords sets' that are candidates of the final n -keywords set. The minimum support value is compared to each keyword. Keywords under the threshold are pruned from the F_k keywords set. The candidates of the two-keyword sets are generated from the remaining keywords in F_k . The minimum support value for two keyword sets are counted again and the frequent keyword sets that have higher support values than minimum threshold remains. Then, the three-keywords sets are generated with the unique keywords combined with frequent two keyword sets. This operation continues until no frequent keyword sets are found to be lower than the threshold.

3.2.6. n-ary association rule extraction

The Apriori algorithm is employed to identify both binary and n -ary ($n \geq 2$) associations among keywords through an iterative process, as outlined in the pseudocode above. The algorithm progressively generates candidate keyword-sets, which are then pruned based on predefined thresholds. The minimum lift threshold (Eq. 3) of 1.0 was selected as the standard criterion for positive association in ARM literature, ensuring that only keyword pairs that co-occur more frequently than chance are retained [31, 32]. A maximum confidence (Eq. 2) of 0.95 was applied to exclude trivial or tautological rules, in keyword-level scientometric analysis, confidence values approaching 1.0 typically reflect insignificant patterns (e.g., a keyword always appearing with itself or a near-synonym) that provide no novel semantic insight. The minimum support threshold (Eq. 1) was calibrated to the dataset size (12,185 records) to extract rules appearing in a sufficient number of publications. The sensitivity of the extracted rule set to threshold selection was examined across all three parameters. With respect to lift,

reducing the minimum threshold from 1.0 to 0.8 increases the number of extracted rules by approximately 40%, but the additional rules are low-strength associations ($0.8 \leq \text{lift} < 1.0$) that fall below the positive-association criterion and introduce spurious co-occurrences; conversely, raising the threshold to 2.0 yields a more focused set dominated by high-confidence, high-lift patterns. With respect to support, lowering the minimum support count admits a larger number of infrequent keyword combinations, which increases coverage at the cost of statistical reliability, since rules supported by only a handful of publications are more susceptible to noise; raising it progressively restricts the rule set to associations that recur across a larger share of the corpus. The confidence ceiling of 0.95 primarily affects the exclusion of near-tautological rules rather than the overall rule count. The thresholds adopted in this study therefore represent a deliberate balance between coverage and specificity, calibrated to the size and composition of the 12,185-record dataset.

It should be noted that formal statistical significance testing of individual association rules is not standard practice in keyword-level scientometric analysis, since the support, confidence, and lift metrics themselves serve as the accepted quantitative measures of association strength and reliability in the association rule mining literature [31, 32]. Nevertheless, the integration of statistical significance testing and formal robustness procedures into scientometric rule mining represents a valuable methodological direction, and is identified in Section 5.3 as an avenue for future work. It is important to note that keyword co-occurrence associations identified by the MAR method reflect statistical patterns in the published literature and should not be interpreted as causal, conceptual, or semantic equivalence relationships between research topics. High lift values indicate that keyword groups co-occur significantly more often than chance, which is a meaningful scientometric signal.

12,185 publications were acquired from Scopus database to obtain the BIM keyword dataset. Thereby, BIM domain-knowledge association rules are extracted through applying the Apriori algorithm. Since there are thousands of rules extracted, only a few high-impact rules ($\text{lift} \geq 10$) could be reported in the findings section as a representative of possible insights.

3.2.7. Trend analysis and visualization

A regression-based approach was employed using yearly aggregated lift values derived from the multi-association rule (MAR) analysis to investigate the temporal evolution of semantic associations within BIM-domain data. The lift metric represents the strength of co-occurrence between keywords and was used as the primary indicator of semantic association intensity over time.

Let t denote the year and L_t represent the average lift value corresponding to that year. A linear regression model was formulated as follows:

$$L_t = \alpha t + \beta \quad (4)$$

where α represents the rate of change in lift values over time, and β is the intercept term.

It is important to note that this linear regression model is intended as an exploratory baseline for visualising directional trends in lift values, and not as a definitive predictive model. The linear formulation cannot capture non-linear, cyclical, or disruptive dynamics in research trends. Given the nine-year observation window, methods suited to short time series such as polynomial or piecewise regression, exponential smoothing, or Bayesian regression, are recommended for future work requiring higher predictive accuracy. A 95% confidence interval was used to quantify prediction uncertainty.

arulesViz was used to visualize the BIM-KG [44]. The visualization enables users to set thresholds for support, lift, and confidence interactively (Fig. 5).

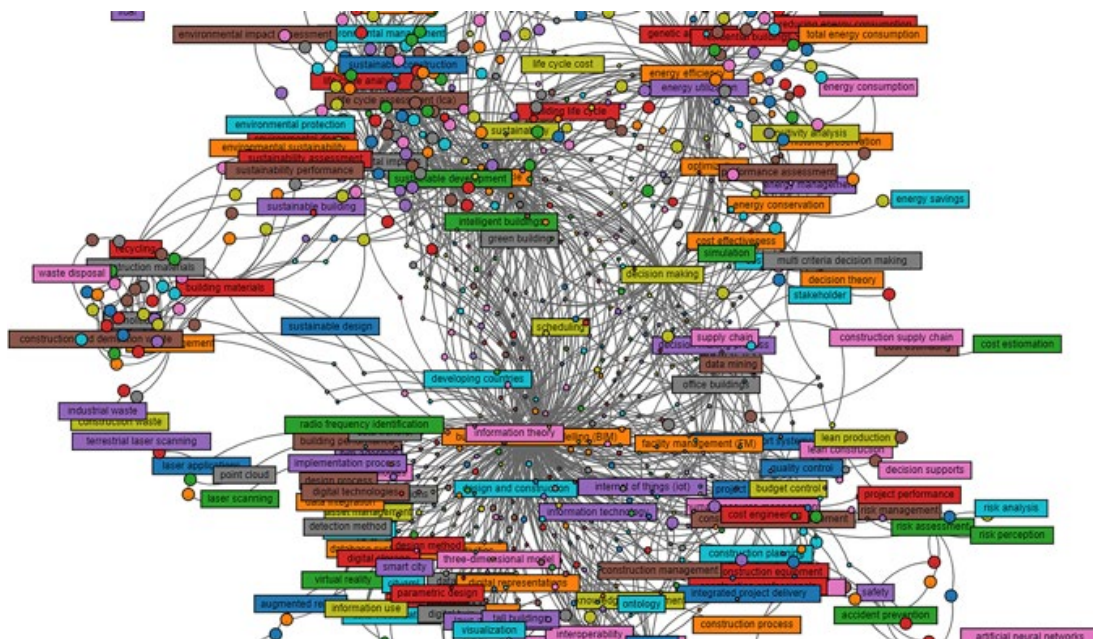


Fig. 5. Visualization of BIM-KG

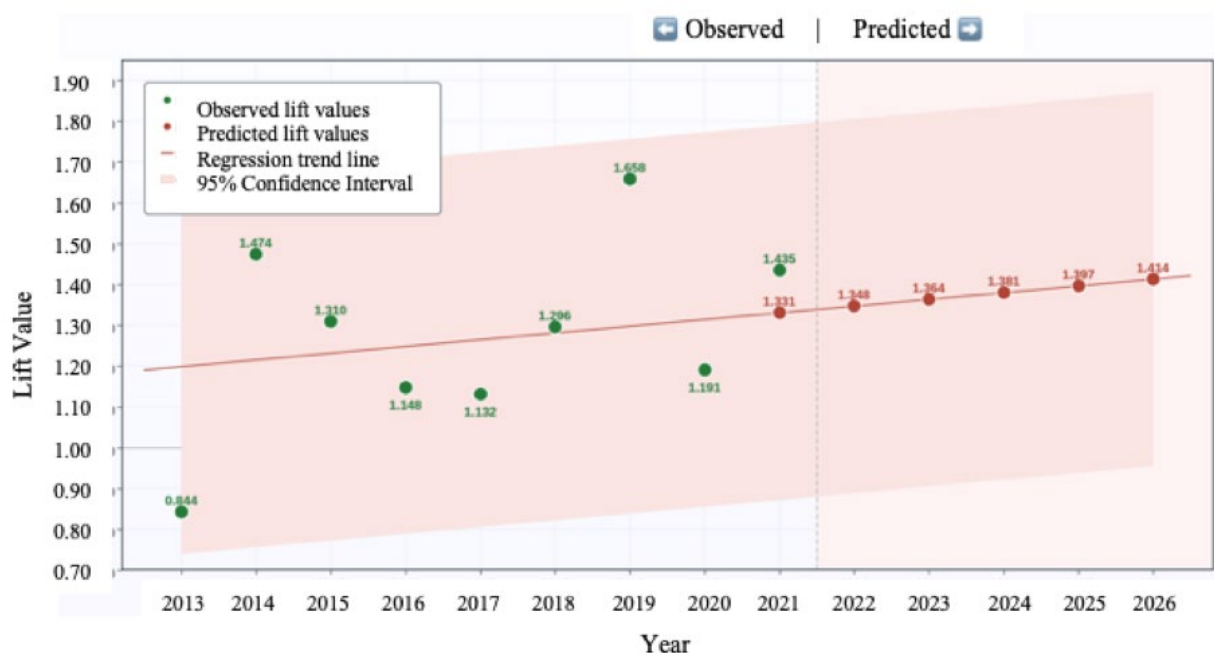
4. Findings

4.1. Sample association rule-based research trend prediction (binary) result

This section presents the temporal prediction of future research trends via association rule mining metrics. A sample association rule-based research trend prediction for ‘BIM→Decision-Making’ is shown in Fig. 6 and Table 1, based on the linear regression model defined in Eq. (4). The linear regression slope (α , Eq. 4) of 0.0165 corresponds to an annual increase of approximately 0.016–0.017 units in lift value. In Fig. 6, observed lift values (2013–2021) are shown alongside predicted values (2022–2026) based on a linear regression model. The shaded region represents the 95% confidence interval. It indicates increasing association strength and emerging semantic consolidation on this association rule.

2013–2015 is the emergence period of BIM-Decision making association. The lift value of 0.844 in 2013 fell below the 1.0 threshold, suggesting that BIM and decision-making were not yet meaningfully associated in the literature. This is consistent with BIM research at that time focusing primarily on technical aspects such as modelling, clash detection, and IFC standards. By 2014–2015, lift values rose above 1.0 (1.474 and 1.310), reflecting the emerging framing of BIM as a decision-support platform [45].

During the consolidation (2016–2018) of BIM-decision making focused research, small, variable residuals (+0.078, −0.101, −0.133, +0.015) indicate that the association stabilised and became normalised in the literature during this period, with near-zero deviation in 2018 suggesting full alignment with the linear trend.



Regression: Lift = 0.0165 * Year + (-32.0155) | SE = 0.2338 | 95% CI shown as shaded band

Fig. 6. Temporal evolution and prediction of lift values for ‘BIM-decision making’ keyword association (binary association)

Table 1. Lift values (2013–2026) of association rule-based research trend prediction for ‘BIM-decision making’

Year	Observed Lift	Predicted Lift	Residual	95%CI
2013	0.844	1.199	-0.355	[0.741, 1.657]
2014	1.474	1.216	+0.259	[0.757, 1.674]
2015	1.310	1.232	+0.078	[0.774, 1.690]
2016	1.148	1.249	-0.101	[0.790, 1.707]
2017	1.132	1.265	-0.133	[0.807, 1.723]
2018	1.296	1.282	+0.015	[0.823, 1.740]
2019	1.658	1.298	+0.360	[0.840, 1.756]
2020	1.191	1.315	-0.123	[0.856, 1.773]
2021	1.435	1.331	+0.104	[0.873, 1.789]
2022	NA	1.348	NA	[0.889, 1.806]
2023	NA	1.364	NA	[0.906, 1.882]
2024	NA	1.381	NA	[0.922, 1.839]
2025	NA	1.397	NA	[0.939, 1.855]
2026	NA	1.414	NA	[0.955, 1.872]

In the peak divergence (2019), the observed lift of 1.658 is 0.360 above the regression prediction, which represents the strongest association in the observation period. This is consistent with the convergence of multiple concurrent developments, including the post-2016 UK BIM mandate translating into published empirical output and increasing AI-BIM integration studies. These contextual interpretations are offered as plausible explanations rather than empirically verified causes.

The subsequent years (2020–2021) represent the disruption and recovery period. The dip to 1.191 in 2020 is consistent with documented slowdowns in construction research output during the COVID-19 pandemic period. The subsequent rise to 1.435 in 2021 aligns with reports of accelerated digital transformation in the mid-pandemic and post-pandemic construction industry.

The forecast values show a steady increase from 1.348 (2022) to 1.414 (2026) all remaining above the positive association threshold of 1.0. The confidence interval (approximately ± 0.46) reflects the uncertainty inherent in a model fitted to 9 years of observation data. This interval is expected to narrow as the dataset expands.

The regression model accurately represents the long-term linear growth of the ‘BIM–decision-making’ association. However, each outlier is not random noise. It can be a result of structural external factors such as policy shifts, technological waves, or global crises. This provides evidence that a meaningful interpretation can be achieved through integrating the quantitative trend with its external context.

4.2. Sample multi-association rule-based research trend prediction (n-ary)

The temporal evolution of sample high-impact n-ary multi-association rules provides deeper insights into BIM research dynamics (Fig. 7 and Table 2). The consistently increasing

predicted lift values suggest these associations are accelerating and structurally embedded in the BIM ecosystem.

Rule R4 (information theory, sustainability → BIM, sustainable development) exhibits the highest observed and predicted values. This highlights the rapid convergence between sustainability-focused research and information-centric BIM frameworks. Rules R2, R7, R8, and R9 form a life cycle and environmental impact cluster. These rules indicate that BIM-LCA integration has become institutionalized in the research domain. Rules R1 and R10 highlight the strengthening association between BIM and project management. Rules R3 and R6 indicate that BIM’s conceptual framework is increasingly becoming data-driven. This, in turn, suggests that interoperability, data integration, and semantic structuring are key drivers of advanced BIM functionality.

These results confirm that the proposed MAR method does not only identify strong associations but also highlights BIM’s evolving role as a key enabler of smart, integrated, and sustainable built environment systems.

Observed (2013–2021) and predicted lift values (2022–2026) in Fig. 7 indicate increasing association strength and emerging semantic consolidation in BIM domain association rules. The results reflect a long term structural upward trend that reflects the evolution of the BIM domain towards higher levels of knowledge integration and interoperability.

These influential results demonstrate that BIM research is moving toward a tightly interconnected, lifecycle and sustainability focused, and data-driven ecosystem. The dominance of “information theory” further shows the shift toward an information centric paradigm. Thereby, BIM promotes new research directions in data integration, processing, and interpretation.

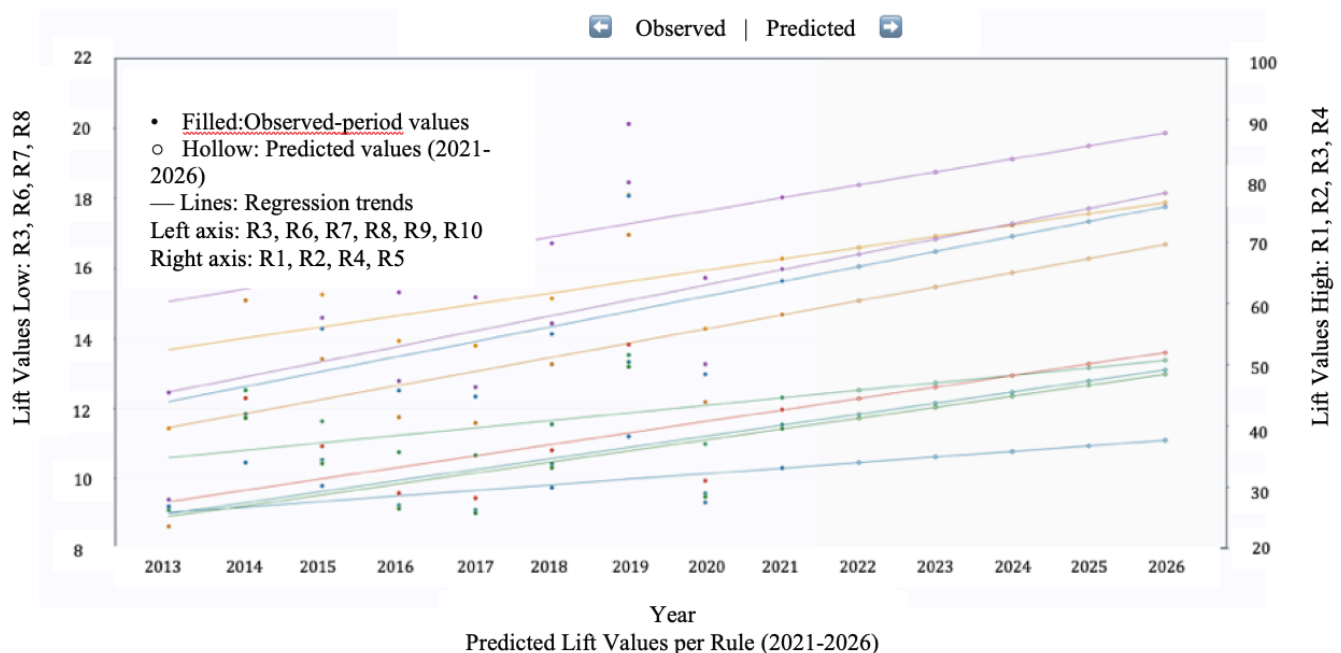


Fig. 7. Temporal evolution and prediction of lift values for ‘BIM-related keywords’ associations (n-ary association)

Table 2. Sample high impact multi-association rules (Lift Value ≥ 10), corresponding metric values, and predicted lift values (2021-2026)

Rule ID	Association Rules	Conf.	Supp.	Observed Lift	Predicted Lift					
					2021	2022	2023	2024	2025	2026
R1	Construction management, information theory $\rightarrow$ BIM, project management	0.906	0.002	29.479	33.111	34.019	34.927	35.835	36.743	37.651
R2	Environmental impacts, information theory $\rightarrow$ BIM, life cycle	0.941	0.001	39.691	44.581	45.804	47.026	48.249	49.471	50.694
R3	BIM, information systems, life cycle $\rightarrow$ information theory	0.929	0.001	10.645	11.957	12.284	12.612	12.94	13.268	13.596
R4	Information theory, sustainability $\rightarrow$ BIM, sustainable development	0.926	0.002	68.763	77.235	79.353	81.471	83.589	85.707	87.825
R5	FM, information theory $\rightarrow$ BIM, office buildings	0.882	0.003	59.878	67.255	69.1	70.944	72.788	74.633	76.477
R6	BIM, GIS, information systems $\rightarrow$ information theory	0.895	0.001	10.257	11.521	11.837	12.153	12.469	12.784	13.1
R7	BIM, decision making $\rightarrow$ life cycle	0.882	0.001	14.214	15.965	16.403	16.841	17.279	17.717	18.154
R8	BIM, environmental impacts, sustainable development $\rightarrow$ life cycle	0.867	0.001	13.912	15.626	16.055	16.483	16.912	17.34	17.769
R9	BIM, environmental impacts $\rightarrow$ life cycle	0.811	0.003	13.061	14.67	15.073	15.475	15.877	16.279	16.682
R10	BIM, construction management $\rightarrow$ project management	0.85	0.001	10.159	11.411	11.724	12.036	12.349	12.662	12.975

These findings confirm that the proposed MAR method does not only identify strong associations but also highlight BIM's evolving role as a key enabler of smart, integrated, and sustainable built environment systems. This in turn, provides a robust analytical foundation for understanding the dynamic transformation of the interdisciplinary BIM domain.

5. Discussion

5.1. The designed MAR-based scientometric analysis tool and comparison with existing approaches

The primary contribution of this study is the proposal of a MAR-based method capable of extracting n-ary ($n \geq 2$) keyword associations from bibliometric data — a capability that extends beyond the binary co-occurrence paradigm common in existing scientometric analysis tools. The aim is not to assess or improve existing tools' outputs or process but to offer a complementary analytical framework that addresses a structural gap in the representation of multi-keyword research relationships. Existing scientometric tools (VOSviewer [13], CiteSpace [14], and Pajek [15]) operate on binary co-occurrence frequency matrices derived from keyword pairs. VOSviewer generates co-occurrence network maps where nodes are keywords and edge weights reflect pairwise frequency. CiteSpace provides burst detection and tree-ring visualization of citation and keyword trends. Pajek

enables network clustering and centrality analysis. These tools provide valuable insights into research structure but share a fundamental constraint in that they cannot model associations among three or more (n-ary) keywords simultaneously, and they do not produce quantified measures of association strength (confidence, lift, support) beyond 'co-occurrence frequency' and 'total link strength' values.

The proposed MAR framework addresses these limitations by

- (1) enabling the extraction of n-ary ($n \geq 2$) keyword associations through Apriori-based rule mining;
- (2) quantifying association strength via lift, confidence, and support metrics; and
- (3) representing n-ary (multi-keyword) rules as directed edges in a formal knowledge graph structure.

This provides a richer and more expressive representation of research knowledge structures [46]. The addition of temporal regression over lift values further differentiates the proposed method by introducing a prospective, exploratory dimension absent from standard scientometric tools. The BIM-KG enables the definition of all keywords and their frequent associations based on defined thresholds, providing quantitative metrics to support interpretation and decision-making.

5.2. The trends in the BIM research

BIM supports decision-making across the building lifecycle. Recent BIM research focuses on project management, interoperability (IFC), and lifecycle-oriented analysis. The findings reveal emerging trends that collectively shift toward the following clusters:

- *Cognitive BIM and Knowledge Driven Systems*: BIM is evolving toward cognitive systems capable of contextual reasoning and knowledge-based decision-making with semantic knowledge graphs and artificial intelligence [24, 47-49].
- *Real-Time Data Fusion and Edge Computing*: Integration of BIM with IoT data and edge computing provides dynamic, scalable, and low-latency systems [11, 25, 27].
- *Generative Artificial Intelligence (GenAI) Integrated BIM*: Design automation and natural language processing (BIM-LLM) integration is rapidly gaining traction [12, 48, 49, 50-53].
- *Explainable AI (XAI) in BIM-based Decision-Making*: Growing demand for XAI in BIM-based systems to ensure transparency, trust, and accountability in automated decision-making processes [53-55].
- *Interoperability 2.0 (Semantic and AI Driven)*: BIM compatibility is shifting toward semantic and AI-driven frameworks enabling intelligent data exchange [24, 48, 56, 57].
- *Human-Centric BIM*: Collaborative intelligence bridging human expertise and AI-powered analytics [25, 38, 49].
- *BIM for Resilience and Sustainability*: Predictive risk modelling, LCA, and carbon footprint tracking integrated within circular economy frameworks [58-62].
- *Industry 5.0 and Robotics*: Robotic construction site applications and 3D scanner integration for autonomous processes [26, 63-65].

5.3. Limitations of the study

This study has several limitations that should be considered when interpreting its findings. First of all, the bibliometric data is retrieved exclusively from the Scopus database. While Scopus is the most widely used source in scientometric research [17, 22, 23], the exclusion of Web of Science, IEEE Xplore, and other datasets may limit the completeness of keyword coverage and introduce database-specific bias in extracted association rules. Future studies are recommended to replicate the MAR approach across multiple databases to assess the robustness of the identified patterns.

The regression model used for lift trend prediction employs a simplified linear formulation, which cannot fully capture the non-linear or potentially disruptive dynamics of research trend evolution. The observation window in this study spans nine years, which constrains the class of models that can be meaningfully applied, since data-intensive forecasting approaches require substantially longer series to produce stable parameter estimates and would risk overfitting at this scale. Methods better suited to short series therefore

offer a more realistic path forward. These include polynomial or piecewise (segmented) regression to accommodate curvature and structural breaks such as those observed in 2019 and 2020, exponential smoothing techniques such as Holt's linear trend method, weighted regression that assigns greater influence to more recent years, and Bayesian regression, which quantifies parameter uncertainty explicitly rather than relying on asymptotic assumptions. As the bibliometric record accumulates and the observation window lengthens, more data-intensive time-series approaches will become viable and are recommended for subsequent studies.

The practical utility of the proposed BIM-KG has not been validated through expert evaluation, user testing, or real-world case studies. Therefore, the framework's outputs should be considered decision-support aids that require further validation by practitioners before being adopted for research strategy or policy-making. Nevertheless, a comprehensive validation through expert assessment and real-world applications remains necessary.

The contextual interpretations linking lift fluctuations to external events (policy mandates, technological waves, the COVID-19 pandemic) are offered as plausible explanations consistent with the quantitative patterns observed. They are not empirically validated causal claims.

Furthermore, the incorporation of formal statistical significance testing and robustness procedures into the rule extraction process (though not standard in current scientometric ARM practice) would strengthen the validation of extracted associations and is recommended for future methodological development.

5.4. Theoretical implications

This research contributes to the advancement of scientometric analysis by extending traditional binary association analysis approaches towards a multi-association rule (MAR) method. This method is capable of capturing n-ary associations among keywords. From a theoretical perspective, the integration of MAR with KG representation establishes a hybrid analytical method that bridges data mining and semantic network analysis. This approach provides a richer and more expressive representation of knowledge structures. Furthermore, the temporal extension with regression-based analysis introduces a dynamic dimension. This allows the identification of existing patterns and the prediction of their evolution in time.

5.5. Practical implications

The proposed MAR-based BIM-KG offers several practical benefits for researchers, practitioners, and decision-makers in the built environment. It provides a systematic tool for identifying emerging research trends and interdisciplinary connections. The research output enables researchers to position their work more strategically within the evolving BIM domain. These practical benefits should be understood as complementary to, rather than replacing, domain expertise and traditional review methods. The BIM-KG outputs require

qualitative interpretation by subject-matter experts before being used to guide research strategy or technology adoption decisions, consistent with established principles holding that quantitative research metrics should support rather than replace expert assessment [66, 67].

5.6. Comparison with existing literature

Widely used tools mainly rely on network representations derived from binary co-occurrence frequencies, often overlooking multi-dimensional associations among research concepts. In contrast, the proposed MAR-based approach enables the extraction and analysis of n-ary associations. The integration of association rules with knowledge graph construction provides a structured and semantically enriched representation. The addition of temporal regression analysis further differentiates the proposed method through its exploratory predictive capabilities.

5.7. Suggestions for future studies

Future research can use various hybrid algorithms combinations to work with big data. Techniques for combining synonymous words would enhance the outcome. Prospective research may incorporate data on average publication years and citations to improve the KG. Author and document-based citation analyses can be developed. In recent studies, the “noise” issue encountered when confidence values exceed a certain threshold can be overcome using more advanced pruning algorithms or hybrid algorithms. Thus, future research can reach more multiple association rules with fewer data. Furthermore, temporal metric can be integrated into the analysis to examine the dynamic structure of the knowledge graph. Future studies can produce tabular data of all metrics by developing an interface that allows users to export values automatically. Validation through expert evaluation and user testing is a high-priority direction for future work, as is the replication of the MAR approach across multiple bibliometric databases to assess cross-database robustness. Short-series regression methods, such as piecewise or Bayesian regression, are recommended for lift value prediction to overcome the limitations of the linear baseline employed in this study. As the observation window lengthens, more data-intensive time-series approaches will also become applicable.

6. Conclusions

This study presents a multi-association rule-based (MAR) method that extends traditional scientometric analysis beyond binary co-occurrence methods. The proposed method enables the systematic extraction and interpretation of n-ary ($n \geq 2$) keyword associations. It integrates the Apriori algorithm with knowledge graph (KG) visualization to provide a structured

and scalable mechanism to understand complex semantic associations in BIM research.

The findings demonstrate that the MAR method significantly enhances scientometric studies by quantifying multi-keyword associations through support, confidence, and lift metrics. The proposed method captures hidden, high level associations that reveal emerging interdisciplinary trends. In particular, the consistent increase in lift values across both binary and n-ary association rules indicates a progressive semantic consolidation within the BIM domain, reflecting a transition toward more integrated and mature knowledge structures. In the BIM domain, these include the convergence of approaches such as BIM with digital twins, generative AI, robotics, IoT, life cycle assessment, information-centric paradigms, and cognitive system of systems perspectives. Such insights highlight BIM's evolving role within the AECO-FM industry, transitioning from a data management tool to an intelligent, predictive, and integrative platform.

The regression-based temporal extension of association rule mining enables the interpretation and prediction of evolving BIM research dynamics. The observed gradual increase in predicted lift values suggests that the identified association rules represent long-term structural trends rather than transient patterns. The linear regression model is intended as an exploratory baseline. To improve forecast accuracy and capture non-linear dynamics, future work should employ predictive models appropriate to the available series length, such as piecewise or Bayesian regression. This predictive capability enhances the analytical depth of scientometric studies by linking historical patterns with future research trajectories.

Despite its contributions, this study has limitations including the reliance on a single bibliometric database, keyword-based representation, and the use of a linear regression model that cannot fully capture non-linear temporal dynamics. These limitations point to important directions for future research. In particular, the integration of automated semantic enrichment techniques, full-text semantic analysis, advanced machine learning-based prediction models, multi-database coverage, and expert validation of the BIM-KG outputs are needed to further enhance the robustness and generalizability of the results.

The proposed MAR-based knowledge graph method provides a complementary tool, enabling a comprehensive and multidimensional examination of research domains. The application of higher-order association mining and temporal prediction method to the BIM domain demonstrates its potential to uncover complex knowledge patterns in rapidly evolving interdisciplinary fields and to support evidence-based decision-making processes.

Declarations

Conflict of Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research received no external funding.

Author Contributions

G. B. Ozturk: Conceptualization, Methodology, Software, Visualization, Validation, Formal Analysis, Investigation, Data Curation, Writing- Original draft, Writing - Reviewing and Editing. F. Soygazi: Conceptualization, Methodology, Software, Visualization, Validation, Formal Analysis, Investigation, Data Curation, Writing - Original draft, Writing - Reviewing and Editing. G. G. Gazi: Methodology, Software, Data Curation, Visualization, Software, Writing - Original draft.

Acknowledgments

Not applicable.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Ethics Committee Permission

Not applicable.

Use of Generative AI and AI-assisted Technologies

The author(s) used Claude Sonnet 5 to translate text and to check grammar, consistency and formatting, and have reviewed and take full responsibility for the final content.

References

- [1] Volk R, Stengel J, Schultmann F (2014) Building information modeling (BIM) for existing buildings—Literature review and future needs. *Automation Construction* 38:109–127. <https://doi.org/10.1016/j.autcon.2013.10.023>.
- [2] Succar B (2009) Building information modelling framework: A research and delivery foundation for industry stakeholders. *Automation Construction* 18(3):357–375. <https://doi.org/10.1016/j.autcon.2008.10.003>.
- [3] Azhar S (2011) Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry. *Leadership and Management Engineering* 11(3):241–252. [https://doi.org/10.1061/\(ASCE\)LM.1943-5630.0000127](https://doi.org/10.1061/(ASCE)LM.1943-5630.0000127).
- [4] Ozturk GB (2019) The relationship between BIM implementation and individual level collaboration in construction projects. In: *Proceedings of IOP Conference Series: Materials Science and Engineering*. IOP Publishing. <https://doi.org/10.1088/1757-899X/471/2/022042>.
- [5] Siebert S, Teizer J (2014) Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. *Automation Construction* 41:1–14. <https://doi.org/10.1016/j.autcon.2014.01.004>.
- [6] Gheisari M, Esmaili B (2019) Applications and requirements of unmanned aerial systems (UASs) for construction safety. *Safety Science* 118:230–240. <https://doi.org/10.1016/j.ssci.2019.05.015>.
- [7] De Melo RRS, Costa DB, Álvares JS, Irizarry J (2017) Applicability of unmanned aerial system (UAS) for safety inspection on construction sites. *Safety Science* 98:174–185. <https://doi.org/10.1016/j.ssci.2017.06.008>.
- [8] Darko A, Chan AP, Adabre MA, Edwards DJ, Hosseini MR, Ameyaw EE (2020) Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. *Automation Construction* 112:103081. <https://doi.org/10.1016/j.autcon.2020.103081>.
- [9] Alves JL, Palha RP, de Almeida Filho AT (2025) Towards an integrative framework for BIM and artificial intelligence capabilities in smart architecture, engineering, construction, and operations projects. *Automation Construction* 174:106168.
- [10] Ozturk GB, Tunca M (2020) Artificial intelligence in building information modeling research: Country and document-based citation and bibliographic coupling analysis. *Celal Bayar University Journal of Science* 16(3):269–279. <https://doi.org/10.18466/cbayarfbe.770565>.
- [11] Ozturk GB (2021) The evolution of building information model: Cognitive technologies integration for digital twin procreation. In: Yitmen I (ed) *BIM-enabled Cognitive Computing for Smart Built Environment*. CRC Press, Boca Raton, pp. 69–94. <https://doi.org/10.1201/9781003017547>.
- [12] Ozturk GB, Brilakis I, Kookalani S (2024) Generative AI research fields in the AEC-FM industry. In: *Proceedings of the 8th International Project and Construction Management Conference (IPCMC)*. İstanbul, Türkiye.
- [13] van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84(2):523–538. <https://doi.org/10.1007/s11192-009-0146-3>.
- [14] Chen C (2014) The CiteSpace Manual. <http://cluster.ischool.drexel.edu/~cchen/citespace/CiteSpaceManual.pdf>.
- [15] Batagelj V, Mrvar A (2002) Pajek—Analysis and visualization of large networks. In: Mutzel P, Jünger M, Leipert, S (eds) *Graph Drawing*. GD 2001. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45848-4_54.
- [16] Wei F, Grubestic TH, Bishop BW (2015) Exploring the GIS knowledge domain using citespace. *The Professional Geographer* 67(3):374–384. <https://doi.org/10.1080/00330124.2014.983588>.
- [17] Li X, Wu P, Shen GQ, Wang X, Teng Y (2017) Mapping the knowledge domains of Building Information Modeling (BIM): A bibliometric approach. *Automation Construction* 84:195–206. <https://doi.org/10.1016/j.autcon.2017.09.011>.

- [18] Jin R, Zou Y, Gidado K, Ashton P, Painting N (2019) Scientometric analysis of BIM-based research in construction engineering and management. *Engineering, Construction and Architectural Management* 26(8):1750–1776. <https://doi.org/10.1108/ECAM-08-2018-0350>.
- [19] Vilutiene T, Kalibatiene D, Hosseini MR, Pellicer E, Zavadskas EK (2019) Building information modeling (BIM) for structural engineering: A bibliometric analysis of the literature. *Advances in Civil Engineering* 2019(1):5290690. <https://doi.org/10.1155/2019/5290690>.
- [20] Chen D, Xiang P, Jia F, Guo J (2022) A systematic review of current status and trends of mega-infrastructure projects. *Ain Shams Engineering Journal* 13(6):101773. <https://doi.org/10.1016/j.asej.2022.101773>.
- [21] Marzouk M, Elshaboury N (2022) Science mapping analysis of embodied energy in the construction industry. *Energy Reports* 8:1362–1376.
- [22] Ozturk GB (2020) Interoperability in building information modeling for AECO/FM industry. *Automation Construction* 113:103122. <https://doi.org/10.1016/j.autcon.2020.103122>.
- [23] Ozturk GB (2020) Trends in interoperability in building information modeling (BIM) research: A scientometric analysis of authors and articles. *AJZ ITU Journal of the Faculty of Architecture* 17(3):169–183. <https://doi.org/10.5505/ituja.2020.79026>.
- [24] Ozturk GB, Brilakis I, Ozen B, Soygazi F (2024) Ontology research fields in the cultural heritage domain. In: *Proceedings of 8th International Project and Construction Management Conference (IPCMC2024)*, İstanbul, Turkey. <https://doi.org/10.17863/CAM.108662>.
- [25] Ma T, Xiao F, Zhang C, Zhang J, Zhang H, Xu K, Luo X (2025) Digital twin for 3D interactive building operations: Integrating BIM, IoT-enabled building automation systems, AI, and mixed reality. *Automation in Construction* 176:106277. <https://doi.org/10.1016/j.autcon.2025.106277>.
- [26] Ozturk GB (2021) The integration of building information modeling (BIM) and immersive technologies (ImTech) for digital twin implementation in the AECO/FM industry. In: Yitmen I (ed) *BIM-enabled Cognitive Computing for Smart Built Environment*. CRC Press, pp. 95–129. <https://doi.org/10.1201/9781003017547>.
- [27] Ozturk GB, Ozen B, Soygazi F (2024) Novel technology use for digital transformation of cultural heritage. *Journal of Construction Engineering, Management & Innovation* 7(3):172–188. <https://doi.org/10.31462/jcemi.2024.03172184>.
- [28] Ozturk GB, Soygazi F (2024) Generative AI use in the construction industry. In: Lyu, Z (ed) *Applications of Generative AI*. Springer International Publishing, Cham, pp. 161–187.
- [29] Agrawal R, Imieliński T, Swami A (1993) Mining association rules between sets of items in large databases. In: *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*. <https://doi.org/10.1145/170035.170072>.
- [30] Hong T-P, Lee Y-C (2008) An overview of mining fuzzy association rules. In: Castillo O, Melin P, Kacprzyk J, Pedrycz W (eds) *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*. Studies in Fuzziness and Soft Computing. Springer, Berlin, Heidelberg, pp. 397–410. https://doi.org/10.1007/978-3-540-73723-0_20.
- [31] Han J, Pei J, Tong H (2022) *Data Mining: Concepts and Techniques*, 4th Edition. Burlington, MA. ISBN-10:9780128117606.
- [32] Al-Maolegi M, Arkok B (2014) An improved Apriori Algorithm for association rules. *International Journal on Natural Language Computing (IJNLC)* 3(1):21–29. <https://doi.org/10.48550/arXiv.1403.3948>.
- [33] Sammut C, Webb GI (2017) *Encyclopedia of Machine Learning and Data Mining*. Springer Publishing Company, Incorporated. ISBN: 978-1-4899-7685-7.
- [34] Jia B, Dong C, Chen Z, Chang K-C, Sullivan N, Chen G (2018) Pattern discovery and anomaly detection via knowledge graph. In: *Proceedings of the 2018 21st International Conference on Information Fusion (FUSION)*. Cambridge, UK, pp. 2392–2399. <https://doi.org/10.23919/ICIF.2018.8455737>.
- [35] Hogan A, Gutiérrez C, Cochez M, de Melo G, Kirrane S, Polleres A, Navigli R, Ngomo ACN, Rashid SM, Schmelzeisen L, Staab S, Blomqvist E, d’Amato C, Gayo JEL, Neumaier S, Rula A, Sequeda J, Zimmermann A, et al. (2022) *Knowledge Graphs*. Springer Cham. ISBN 978-3-031-00790-3. <https://doi.org/10.1007/978-3-031-01918-0>.
- [36] Çiftçi O, Soygazi F, Tekir S (2024) Enrichment of Turkish question answering systems using knowledge graphs. *Turkish Journal of Electrical Engineering and Computer Sciences* 32(4):516–533. <https://doi.org/10.55730/1300-0632.4085>.
- [37] Wang Q, Mao Z, Wang B, Guo L (2017) Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29(12):2724–2743. <https://doi.org/10.1109/TKDE.2017.2754499>.
- [38] Geng Y, Chen J, Pan JZ, Chen M, Jiang S, Zhang W, Chen H (2023) Relational message passing for fully inductive knowledge graph completion. In: *Proceedings of the 2023 IEEE 39th International Conference on Data Engineering (ICDE)*. Anaheim, CA, USA, pp. 1221–1233. <https://doi.org/10.1109/ICDE55515.2023.00098>.
- [39] Bird S, Loper E, Klein E (2009) *Natural Language Processing with Python*. O’Reilly Media. Sebastopol, CA.
- [40] Harris CR, Millman KJ, van der Walt SJ, et al. (2020) Array programming with NumPy. *Nature* 585(7825):357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
- [41] McKinney W (2010) Data structures for statistical computing in Python. In: *Proceedings of 9th Python in Science Conference*, pp. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>.
- [42] Hunter JD (2007) Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* 9(3):90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- [43] Ymoch: Apyori. <https://github.com/ymoch/apyori>. Accessed 24 Sep 2026
- [44] Hahsler M (2017) arulesViz: Interactive visualization of association rules with R. *The R Journal* 9(2):163–175. <https://doi.org/10.32614/RJ-2017-047>.
- [45] Eastman CM, Teicholz P, Sacks R, Liston K (2011) *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*, Second Edition. John Wiley & Sons, Inc.
- [46] Tsaneva S, Dessi D, Osborne F, Sabou M (2025) Knowledge graph validation by integrating LLMs and human-in-the-loop. *Information Processing & Management* 62(5):104145. <https://doi.org/10.1016/j.ipm.2025.104145>.

- [47] Kang R, Zhu J, de Silva L, Brilakis I (2026) Context-aware knowledge graph reasoning for road maintenance decision-making. *Advanced Engineering Informatics* 74:104618. <https://doi.org/10.1016/j.aei.2026.104618>.
- [48] Lu L, Luo P, Sheil B, Brilakis I (2026) LLM-driven multi-agent framework for enhancing human-digital twin interaction in built infrastructure management. *Automation in Construction* 187:106896. <https://doi.org/10.1016/j.autcon.2026.106896>.
- [49] Pan Y, Wang M, Lu L, Lamsal R, Pärn E, Zlatanova S, Brilakis I (2026) LLM-enabled multi-agent framework for natural language interaction with graph-based digital twins. *Automation in Construction* 183:106791. <https://doi.org/10.1016/j.autcon.2026.106791>.
- [50] Kookalani S, Parn E, Brilakis I, Dirar S, Theofanous M, Faramarzi A, Feng Q (2024) Trajectory of building and structural design automation from generative design towards the integration of deep generative models and optimization: A review. *Journal of Building Engineering* 97:110972. <https://doi.org/10.1016/j.jobe.2024.110972>.
- [51] Kookalani S, Parn E, Brilakis I, Pan N, Wang M, Alavi H (2024) Exploring deep generative models in building design. In: *Proceedings of Conference 2024 EC³ European Council on Computing in Construction*. Chania, Crete, Greece.
- [52] He Z, Wang Y-H, Zhang J (2025) Generative AIBIM: An automatic and intelligent structural design pipeline integrating BIM and generative AI. *Information Fusion* 114:102654. <https://doi.org/10.1016/j.inffus.2024.102654>.
- [53] Luo P, Parn E, Kookalani S, Brilakis I (2026) TailorAlert: Large language model-based personalized alert generation system for road infrastructure management with digital twins. *Journal of Computing in Civil Engineering* 40(2):04025154.
- [54] Sönmez N, Kuruoğlu M, Maçka Kalfa S, Tokdemir OB (2025) Explainable AI-Driven analysis of construction and demolition waste credit selection in LEED projects. *Architecture* 5(4):123. <https://doi.org/10.3390/architecture5040123>.
- [55] Kiavarz H, Jadidi M, Rajabifard A, Sohn G (2024) An explainable & prescriptive solution for space-based energy consumption optimization using BIM data & genetic algorithm. *Journal of Building Engineering* 92:109763. <https://doi.org/10.1016/j.jobe.2024.109763>.
- [56] Vittori F, Tan CF, Pisello AL, Chong A, Piselli C, Miller C (2025) BIM-to-BRICK: Using graph modeling for IoT/BMS and spatial semantic data interoperability within digital data models of buildings. *Energy and Buildings* 348:116368. <https://doi.org/10.1016/j.enbuild.2025.116368>.
- [57] Argasiński K, Tomczak A (2025) Enhancing semantic interoperability of heritage BIM-based asset preservation. *Heritage* 8(10):410. <https://doi.org/10.3390/heritage8100410>.
- [58] Albanese PM, Baglivo C, Congedo PM (2025) Integrating LCA with BIM-based technologies in the building construction context: A review. *Buildings* 16(1):168. <https://doi.org/10.3390/buildings16010168>.
- [59] Nasir MM, Prakash PR, Suman M (2025) An automated BIM-GRIHA15-LCA framework for building sustainability assessment. *Journal of Building Engineering* 101:111908. <https://doi.org/10.1016/j.jobe.2025.111908>.
- [60] Bolognesi C, Bassorizzi D, Balin S, Manfredi V (2025) An overview on LCA integration in BIM: Tools, applications, and future trends. *Digital* 5(3):31. <https://doi.org/10.3390/digital5030031>.
- [61] Birgonul Z (2024) A user-centric analysis and survey-based approach for early design decision-making criteria of the ‘Symbiotic Data Platform’: Integration of BIM with IoT for occupants’ thermal comfort. *Journal of Construction Engineering, Management & Innovation* 7(1):56-76. <https://doi.org/10.31462/jcemi.2024.01056076>.
- [62] Azima M, Seyis S (2022) Evaluating the impact of building envelope on energy performance: Cooling analyses. *Journal of Construction Engineering, Management & Innovation* 6(1):48-56. <https://doi.org/10.31462/jcemi.2023.01048056>.
- [63] Almusaed A, Yitmen I, Almssad A, Myhren JA (2024) Construction 5.0 and sustainable neuro-responsive habitats: Integrating the brain-computer interface and building information modeling in smart residential spaces. *Sustainability* 16(21):9393. <https://doi.org/10.3390/su16219393>.
- [64] Kasperzyk C, Kim MK, Brilakis I (2017) Automated re-prefabrication system for buildings using robotics. *Automation in Construction* 83:184-195.
- [65] Kurien M, Kim M-K, Kopsida M, Brilakis I (2018) Real-time simulation of construction workers using combined human body and hand tracking for robotic construction worker system. *Automation in Construction* 86:125-137. <https://doi.org/10.1016/j.autcon.2017.11.005>.
- [66] Hicks D, Wouters P, Waltman L, De Rijcke S, Rafols I (2015) Bibliometrics: The leiden manifesto for research metrics. *Nature* 520(7548):429-431. <https://doi.org/10.1038/520429a>.
- [67] Paulheim H (2016) Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8(3):489-508. <https://doi.org/10.3233/SW-160218>.