buildings

MDPI

Article

Using Network Analysis Theory to Extract Critical Data from a Construction Project

Chih-Han Kao [1,*], Wei-Tong Chen [2,3] and Chung-Kuang Ho [4]

[1] Department of Civil Engineering and Engineering Management, National Quemoy University, Kinmen 89250, Taiwan
[2] Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology, Douliu 64002, Taiwan; chenwt@yuntech.edu.tw
[3] Department of Civil and Construction Engineering, National Yunlin University of Science and Technology, Douliu 64002, Taiwan
[4] Yi Sheng Lai Architect Corporation, Changhua 50048, Taiwan
* Correspondence: al@nqu.edu.tw

Abstract: Construction projects are inherently complex and entail extensive information processing. Thus, they require effective information management, which, in turn, requires the preservation of critical construction data (CD). Although BIM and blockchain methodology use the "change type of query and storage for data management" to improve the service quality of data, data redundancy still causes inefficient retrieval. Moreover, project managers face various source limitations, which prevent the contents of the database from being managed efficiently. This study uses network analysis theory to design an information network (IN). Critical CD were extracted, and an IN structure was built using data from construction practices (network nodes) and data relation (network links). Three metrics were used for performance evaluation of the data references and data delivery. The refurbishment of heritage buildings in Kinmen, Taiwan, was used as a case study to extract critical CD such as the "inspection record checklist" and "architect design plan drawing". Lastly, CD can be applied as the elementary item of a backstage database for BIM and blockchain applications of DM. The combined system of critical DM can play an important role in obtaining comprehensive information for a construction project. Customized metrics of IN analysis can be developed as an integrated composite to decide the priority of CD.

Keywords: data redundancy; refurbishing heritage buildings; network analysis; project information network

check for updates

Citation: Kao, C.-H.; Chen, W.-T.; Ho, C.-K. Using Network Analysis Theory to Extract Critical Data from a Construction Project. *Buildings* **2023**, *13*, 1539. https://doi.org/10.3390/buildings13061539

Academic Editor: Jurgita Antucheviciene

Received: 4 May 2023
Revised: 8 June 2023
Accepted: 10 June 2023
Published: 16 June 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Background

Increases in the complexity and scope of a construction project increase the amount of project information that is generated and required [1]. Such information comes in various forms (e.g., checklists, manuals, codebooks, contracts, and drawings), and is generated through complex connections in the construction project's life cycle. The effective management of such information is a difficult and frequently overlooked aspect of construction project management [2].

Interactive collaboration on an engineering project should be based on the consistent use of information by all members of the project team. Project information is often modified according to feedback from other groups (e.g., change orders or unexpected new discoveries in the subsequent phase of the project's life cycle), requiring continuous revision and updating. During the construction phase, the original design drawings should also be adjusted in response to change orders. Interactions between project groups result in bilateral information flows, which shift across different life cycle stages [3]. As a result, common task data must be properly updated and easily accessible through a centralized database to

ensure that every member obtains consistent and timely information [4]. However, the absence of a uniform and transparent system of construction project information undermines the assurance process and may lead to disputes during a construction project [5].

Poor data management can interfere with important construction tasks, including project monitoring and management. Engineers and managers frequently waste valuable time and effort collecting and processing construction data from engineering documents [6]. To manage document information, the engineering project team continuously updates specific learning databases to share information with all project team members [5,7,8]. The collected data must readily interrogatable by designers and engineers developing data process specifications. Effective data processing can ensure the quality (e.g., compressive, query-efficiency, original data tracking, and correct functioning) of engineering information [8,9]. Therefore, database systems still require significant project management efforts to integrate information and analyze relationships among datasets.

Building information modeling (BIM) is a database approach that is highly effective for knowledge sharing in architecture engineering and construction (AEC) projects [10–13]. However, BIM data storage needs to be integrated to maximize the efficiency of the database process [14]. Since BIM data are locked inside vendor-specific implementations and interfaces [8], nonspatial information remains difficult to query for BIM. Solihin et al. [8] proposed an integrated approach according to the concept of specialized construction data models. The research developed an interface of flexible and efficient queries, including BIM data, with a standard SQL which removed restrictions on predefined queries, effectively transforming the BIM data into an open and query-able database. Furthermore, there is another storage development aspect of data management. Blockchain technology has been adopted to preserve original data when information is offered to all members during a cooperative task of a construction project [5,15,16]. However, the costs of processing data and storage need to be evaluated when a large amount of data are stored as blockchains.

Engineering project documents combine multiple data sources from various stages of an engineering project's life cycle. These engineering project data items, such as forms and handbooks, are created using the same basic design content and requirements, whereas the content in the engineering project document is cross-referenced. Additionally, the traditional database of a construction project is constructed according to the IFC standard [8]. Thus, combining these two practical aspects allows for a relatively large number of entities in the relational database of a construction project to be generated. The differences in the hierarchical natures of object-oriented data (e.g., BIM) and relational data involve performance issues [17], while data redundancy still exists in BIM databases with nonspatial information. Consequently, various databases need to deal with the issue of a relatively large number of entities in the relational database [8]. To enhance construction project data management, a relatively large number of entities need to be eliminated from the relational database, particularly those related to issues of query efficiency and the limited source allocation for data management. To facilitate data integration across different tables in a database, developers add redundant links. Unfortunately, these redundant table links have a negative impact on the database [18–21].

Data redundancy refers to circumstances in which the same data are included in multiple tables in a database system. Non-compressed databases typically lack data redundancy, which has a negative impact on the database's operating performance. However, database developers typically provide multiple extra connections among database tables, which can also negatively impact the database [22]. Some data reduction methods can be used to ensure data integrity and non-repetitiveness [22,23]. Database frames are commonly associated with document-based meta-information, allowing for the reorganization of a database into integrated information structures for specific users or tasks. Such meta-information is reprocessed by extracting critical information from the original documents [24,25] to achieve effective data integration. However, the duplicate entities in a relational database are not removed. Therefore, this research focuses on the developing concept of a relational database structure with regard to the criterion of data elimination (extracting critical data)

from a construction project. The foreign key can be used for evaluating table connections, while the prime key serves as the basic analysis material of a table in a relational database. To eliminate the negative effect of duplicate data in a database, the data need to be properly identified as prime or foreign keys of a relational database according to the criteria of data integration or data reference.

Database development focuses on balancing the information's comprehensiveness and the data capture efficiency to maximize performance and reduce errors. Critical project engineering data can be used as foreign keys of relational databases for engineering projects. To be an appropriate foreign key of a database, critical data must be used in various processes during the project's life cycle. Therefore, critical data must be integrated with many other data types through association relations using prime keys that correspond to foreign keys between different database tables. Screening methods for construction project data can effectively eliminate outdated and redundant data, which are less important for engineering. Such data should be decomposed from the construction project documents, and their utility must be evaluated using engineering practices that connect prime and foreign key databases. Additionally, the critical data need to be assigned with consideration of resource limitation (cost, time, and manpower) for data management, given that adding sources can requires additional data management (e.g., verifying the accuracy of the data and instantly updating data). The prioritized sources and types of data need to be diligently evaluated to achieve efficient management of information and resources.

1.2. Assumption and Procedure of Research

The research aims to develop an information network model for construction projects based on assumptions, including: (1) the critical data of construction projects should consider both information delivery and information derivation for the purpose of performing construction project tasks; (2) all data regarding the construction project should be included in the documents of every construction task during the entire life cycle of the project; (3) the data regarding a construction project must be consistent across the entire life cycle of the construction project.

This study uses information extraction theory for information analysis and information extraction. The significance of extracting robust information from a construction project highlights the importance of this information when setting data management priorities. For this study, the research was divided into four stages, focusing on:

(1) To process the quantity of available data, the information must be decomposed to elementary items.
(2) A conceptual information network should be developed based on the corresponding characteristics of the information flow and network theory. The authors firstly proposed the virtual item of information as the physical individual in traditional network analysis. The information flow structure is indicated via a network graph.
(3) Information network analysis processes are developed based on the analysis data that needs to be transferred from the original practical documents. Connections between data must be indicated in the nodes of the information network.
(4) Finally, data (nodes) are evaluated by network analysis metrics using a case study. Moreover, this study compares the analysis results of the case study against the practical application of critical data for a construction project, and verifies the results obtained using the information extracted from the case study.

Figure 1 illustrates the research flow applied in this study.

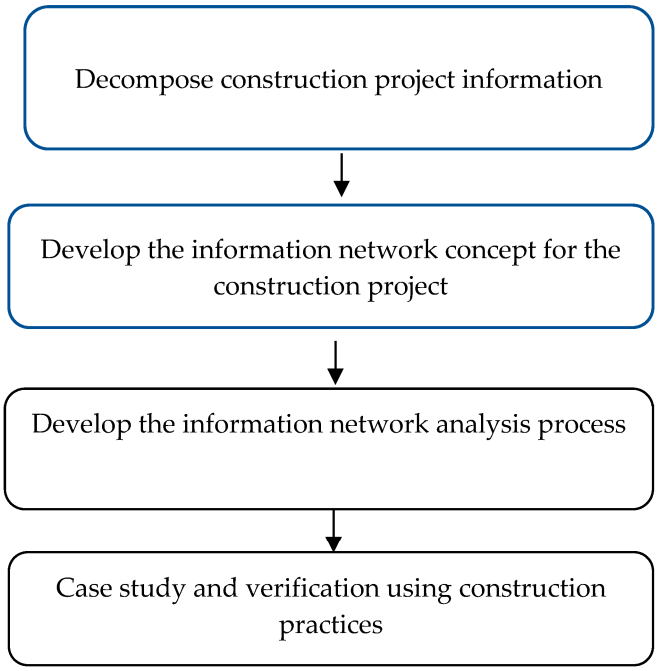

Figure 1. Research flow.

2. Relationship between Network Theory and Information Processes

Construction project information characteristics are cross-referenced and non-standardized using an information structure similar to network graph topology. Since the main goal of this study was to analyze construction project information using network theory, this section highlights how previous studies have used network theory to extract fuzzy information. We then briefly introduce construction project characteristics and summarize the corresponding relationships between the information network and network theory.

Network theory is a methodology for studying the topology and objects of complex networks using quantitative and qualitative graph-based analysis. Network theory has been used to solve problems in social science, computer science, data mining, and other related fields [26–33]. These applications applied the same core concept to demonstrate and analyze relationships between characteristics and actions, using a visual medium to present the relationships between different actors and allowing for full and precise descriptions of authentic conditions without resorting to complex mathematical equations. Network theory includes four major components: graph theory, social networks, online social networks, and graph mining [26]. It has been applied to communities, influence and recommendation, model metrics and dynamics, behavior and relationships, and information diffusion.

Social network analysis (SNA) is a commonly used application in network theory, focused on explicating complex social interactions. Social interactions can be observed in issues related to communication, trust, and consultation [30]. Social group agents are the nodes, while relations between agents are the edges, collectively representing interactions between social groups [28]. In addition, network metrics can be used to evaluate the status of agents in social groups and the state of the entire social group. The network density and the distance between the core agent and other nodes can be used to extract relevant knowledge from practice [31,32].

Cross-referencing activities from other engineering documents affects the data interactions involved in an engineering project. Thus, network theory can be used to solve problems related to graph mining and information diffusion [33]. This study uses network theory to extract critical data from engineering project information. However, previous work has largely focused on the physical individual (human, settlement, or association) as the key network actor, using abstract events (information, communications, or actions) as the network edges [26]. The present study proposes a novel approach to network theory,

focused on the sharing and reuse of virtual information in multiple construction tasks. Therefore, this research creates a new setting for network actors (data) and edges (referencing acts), thus allowing for the precise simulation of information flows to access critical engineering project data.

3. Basic Information Network Concept

Based on the proposed concept of network theory, this study uses an information flow to indicate data relations with indirect and complex connections. The original information network analysis concept is as follows.

3.1. Information Network Components

Since this research focuses on referencing the utility of information for construction projects, the data reference status needs to be indicated specifically. This research develops a unique network structure in which construction project data are the information network nodes (actors), and the corresponding processes between the project data are the network connections (edges). The network diagram can represent the relationships among the project data at both the micro (nodes connection) and macro (network structure) levels.

The significance of critical data is determined based on the frequency with which they are referenced by other data. As a result, the critical data in this study use an information network graph to simulate information flows and interactions. In addition, data which contribute more utility to information networks are more important for construction projects. Thus, the critical data identification criteria suggest that the data have a high level of information utility, and the centrality metric of nodes is used to describe the critical data utility.

3.2. Correlation between Data Utilities and Metrics

Data contribution can be observed through the information flow that presents the data-referencing status. This study aims to evaluate the interactive reference between all data (nodes); thus, the information network metrics should focus on centrality and bridge the issues related to the network analysis metrics. The network analysis centrality issue metric is used to measure the level of influence for all network nodes. The centrality metrics of network nodes indicate their levels of status or power [32,34].

As previously noted, critical data are used as the database's foreign keys, and their features must meet the functional requirements of the foreign key. As a result, the critical data features of a construction project must be linked synchronously to the other project management data. Critical data in construction engineering must be identified using information network centrality metrics, which indicate the referencing importance and direct citation frequency of data for the construction project database. Two centrality issue metrics for network analysis used in this study are degree and closeness.

In a construction project, subsequent phase tasks are based on certain fundamental information from the previous phase. Several information network clusters are formed using the operating act in a single engineering project phase. To maintain the consistency of the information, critical data carry key information from one cluster to the next. Information delivery and indirect reference resemble the "bridge" action of network theory, while the evaluating metric for data bridge utility is the "betweenness." Within the database's comprehensive data set, critical data link to or represent the foreign and prime database keys. Critical data meet one of three metric criteria of an information network; thus, these metrics are not unique, and the result of critical data is the union of superior data from the three criteria. When one of the conditions is met, the data are cited as critical information network data if the condition has high metric value for the central or bridge issue.

4. Analysis Procedure

In this study, we built an information network to evaluate the data utility of a construction project using the following research analysis procedures for information extraction.

We first created a list of all data from the case study's practical documentation. We then constructed an information network of data items by processing the connection between each task. Lastly, the information network was evaluated using the network metric to cite the core node (critical data). Data utility items were evaluated using information network metrics. In addition, this study verified the importance of critical data using practical construction engineering documents. The network construction procedure and network analysis were organized as follows (Figure 2):

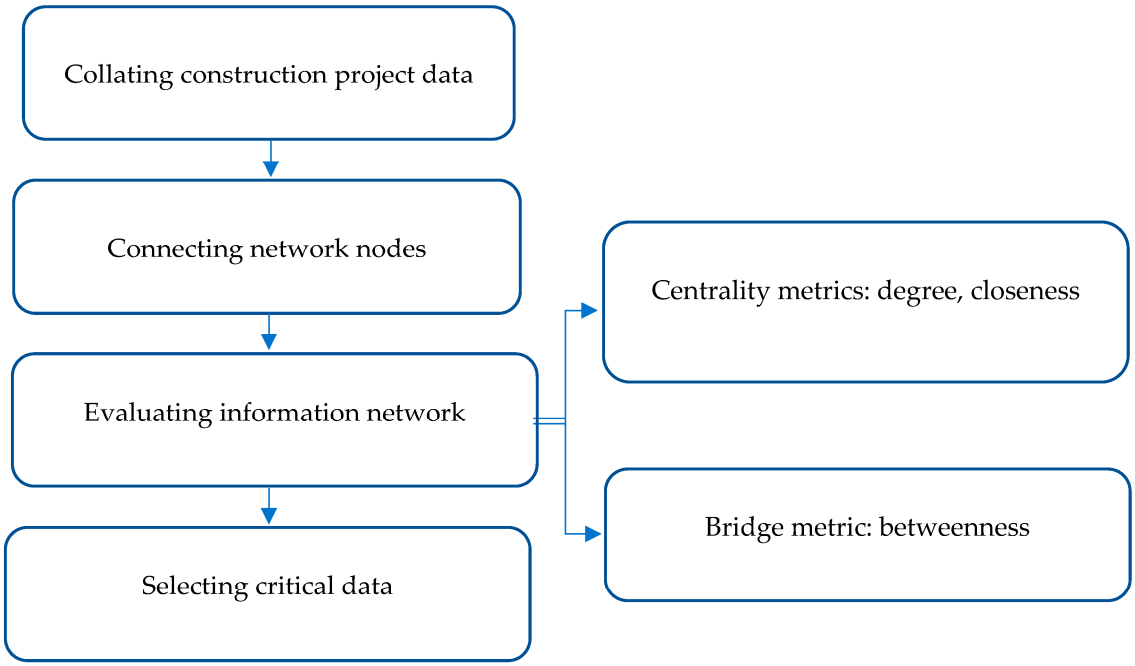

Figure 2. Information network analysis flow chart.

4.1. Establishing an Information Network for Construction Projects

The first phase of this study established a data network for a typical construction project, including the collection of data from practical construction engineering documents and the development of node connections into networks. Data from construction engineering were required to be divided into several individual items from practical construction project documents. To construct a comprehensive information network for the entire life cycle of construction engineering, related documents were included for every completed task.

Construction project documents were collected from all phases of the engineering life cycle. For example, the design drawings, cost estimations, and construction specifications were collected from design phase documents. Key document segments were divided into individual data items to serve as information network elements. The information network nodes served as the construction project data; thus, the design drawings were divided into three parts: plan, elevation, and section. These data were encoded and preprocessed for identification in the network analysis computing process.

The information network analysis phase connected the information network nodes. Engineers with practical experience in construction engineering cited the interactive connection between the two data. The edge between two nodes that had been confirmed and identified referred to the relation of the pairwise data. Moreover, feedback from construction engineering data for the previous project's life cycle phase may have required remedial action. The feedback data were adjusted for the subsequent construction project implementation. This aspect of information interaction has been observed in network analysis. All information interactions were displayed precisely and accessed via a connection matrix.

4.2. Information Network Evaluation

Network theory was used to evaluate each node in the information network. The metrics of information networks serve as a guide for screening critical construction project data. Several network analysis metrics were used to assess the unique data utility, where the data were represented by a node in the construction project information network. In addition, the network's betweenness metric was used to measure the delivery function of the critical data for the construction project. The two metrics, central and bridge, were related to pairwise data cooperation. Furthermore, the network metrics were calculated using the network data.

4.2.1. Centrality Issue Metrics

Two metrics were used to analyze information network centrality, i.e., degree and closeness, which were defined and applied as follows:

A. Degree metric

The degree metric refers to the number of connections to a single node in the information network, where a larger number of connections indicates that it has been referenced by many other data points. Moreover, the degree metric shows that the frequency of the subject data has been directly adopted for the subsequent task. The other individual data comprise the derivative output of those subsequent tasks in the information network. Thus, the degree metric represents the level of importance of the fundamental data for a single node in the information network.

The degree metric equation is as follows:

$$C_D(p_i) = \sum^{N}_{k=1} a(p_i,p_k) \quad (1)$$

$C_D(p_i)$ is the metric value of centrality for the p_i node;

p_i is the ith evaluating and computing node for the pairwise data of the information networks;

p_j is the jth node of data, i.e., the target data;

p_k is the kth node of another related node of the target data;

N is the number of nodes;

$a(p_i,p_k)$ is the direct referring level between p_i and p_k.

B. Closeness

Data influence criteria include not only the number of connections, but also the direct level of citation with which the data has been paired to the critical data. The closeness metric indicates the level of direct reference for the creation of other individual data. Data with a high degree of closeness are useful for the execution of construction projects. A shorter edge distance indicates a more direct correlation for the related data. The closeness metric equation is as follows:

$$C_C(p_i) = (N-1)/(\sum^{N}_{k=1} d(p_i, p_k)) \quad (2)$$

$C_C(p_i)$ is the metric value of closeness for the p_i node;

$d(p_i, p_k)$ is the distance of the referring data between p_i and p_k.

4.2.2. Bridge Issue Metrics

The betweenness metric indicates the bridge act of indirect referencing data. Critical data from the previous engineering phase are delivered to the subsequent phase using a single connection between two data clusters. The "bridge" data have a delivery function to extend information to the subsequent project phase. Without the "bridge" data, it is impossible to create new series of information. The critical information passage was built to connect each data cluster through the bridge data.

The betweenness metric presents the relative ratio of the number of the shortest paths to the total number of possible paths between two groups of nodes. A lower degree of betweenness means that the critical data are in a unique position, and can connect two data clusters. The betweenness metric equation is as follows:

$$C_B(p_i) = \sum_{j=1}^{N} \sum_{k=1}^{-1} (g_{jk}(p_i)/g_{jk}) \quad (3)$$

$C_B(p_i)$ is the betweenness metric value of the ith computing node of the data;
$g_{jk}(p_i)$ is the length of the shortest path between the j and k groups of nodes in p_i;
g_{jk} is the length of the shortest path between the j and k groups of nodes.

5. Case Study for Verification

This study used the refurbishment of heritage buildings in Taiwan's Kimnen County as a case study in order to validate the model used for extracting criterial data for a construction project. There are many precious heritage buildings requiring preservation in Kimnen, Taiwan, resulting in many refurbishment projects being undertaken. Such projects are subject to numerous legal restrictions and building codes (e.g., the Cultural Heritage Preservation Act, Government Procurement Law, and Public Construction Quality Management System). The design, construction, and operating phases for such projects are typically staggered [35], resulting in complex pairwise data interactions and information circulation among practical documents.

This paper used UCINET [29] to analyze the information network involved in the refurbishment of heritage buildings and to verify the critical data extracted using network theory. We also used NetDraw to display the node graph of the information network. The case study analysis results were reviewed from a practical perspective following an analysis process, which will be described as follows.

5.1. Data Adoption for the Case Study

The first step in the analysis process was to create a list of data from the case study's practical documentation. Data were extracted by transferring the practical information from construction engineering for the purpose of analysis by five reviewers, each with at least three years of professional experience in construction engineering and/or management. These reviewers had the necessary skills and background to review and analyze engineering documents, and independently identified the relationships between each data item in the documents. The selected data were compared to ensure the consistence of the five reviewers' opinions. Ninety-eight percent of the data were approved by all reviewers. Furthermore, 2% of the data were rejected by one reviewer. During the review meeting, the reviewers reached a consensus regarding the data list.

Given the large volumes of data involved in the process of refurbishing heritage buildings, the data must be modified using feedback (ex. "change order"). Thus, the amount of data must be adjusted, increasing the likelihood of redundant use of data names during the other content phases. In addition, to avoid fuzzy data identification in the information network analysis process, this study used the combined term "document of phase-data" as the name of the data. The data name can cause a reverse information flow, which is useful for simulating the change order.

Information flow acts require the use of additional computing resources; thus, the case study reviewed herein was required to restrict data adaptation from the design and construction phases as a source of analysis material to construct the information network. This study reviews the entirety of the design and construction phase and presents documentation related to all processes, including design drawing, budget, outsourcing, construction planning, construction, acceptance, and transfer. These documents were reviewed to separate the selected data items from the aforementioned information in the document. Table 1 presents examples of the construction phase data.

Table 1. Data regarding the refurbishment of heritage buildings during the design and construction phases.

Design Phase Documentation	Design Phase Data	Construction Phase Documentation	Construction Phase Data
■ Architect drawing	• Plan view	■ Architect drawing	• Plan view
■ Structure design drawing	• Elevation section	■ Construction plan	• Elevation section
■ Equipment design drawing	• Elevation view	■ Structure design drawing	• Elevation view
■ Component unit price	• Field investigation	■ Equipment design drawing	• Field investigation
■ Amount calculation	• Photos of interior structure damage	■ Main budget table	• Photos of interior structure damage
■ Material specification	• Shop drawing	■ Detail budget list	• Shop drawing
■ Schedule planning	• Equipment drawing	■ Unit price analysis	• Construction method
■ Construction regulations	• Material specification	■ Quantity list	• Equipment drawing
■ Recondition inspection item	• Construction procedure	■ Resource list	• Work procedure
■ Worker qualifications	• Work item	■ Material specification	• Work item
■ Contractor qualifications	• Total price	■ Schedule plan	• Total cost
■ Site management qualifications	• Amount	■ Construction directions	• Work item quantity
■ Subcontractor qualifications	• Item unit price	■ Checklist	• Unit price
	• Specification	■ Worker list	• Specification
	• Calculation of amount	■ Contractor	• Worker
	• Worker requirements	■ Field construction manager	• Construction equipment
	• Construction equipment requirements	■ Subcontractor	• Material inspection
	• Material requirements	■ Construction record	• Material
	• Size of structure element	■ Inspection report	• Material catalog
	• Material brand		• Material sample
	• Material catalog		• Work process description
	• Material sample		• Inspection record
	• Construction procedure		• Detection checklist
	• Random test checklist		• Material specification
	• Materials manual		• Construction record photo
	• Specialist requirements		• Checklist
	• Material test requirements		• Specialist
			• Testing certification
			• Contract ability certification
			• Contractor performance record
			• Business license
			• Subcontractor qualification list
			• Construction equipment acquisition plan
			• Material acquisition plan

5.2. Collating Data for Refurbishing Heritage Buildings

The engineering documents of each phase were collected from each stage of task performance, management, and work supervision. For example, the construction phase included several tasks, such as bidding, purchase, subcontract, construction planning, quality management, schedule management, site management, and safety management. This phase also included documents related to architect drawings, structure design drawings, equipment design drawings, and construction directions, as well as lists of workers, contractors, construction inspections, and field construction managers. Furthermore, the architect drawings were divided into discrete data items, including plan view, elevation section, elevation view, and shop drawings.

5.3. Establishing Connecting Frames of the Information Network

Based on the review of the practical documents, this study created a checklist with which to review the pairwise connections of each data pair that corresponded to the connection criteria. Several types of connection criteria were created, including: (1) detailed spatial data derived from generalized spatial data; (2) basic data for the calculation of new data; (3) inspection document standards translated from design specifications; (4) contract terms translated from project management objectives; (5) qualification determinations based on management records; etc. These practical documents verified whether the other data referred to the document data.

There are several examples of the data pairwise connections that the connecting states examined using the following steps: (1) spatial scale data of images were used for quantitative calculations; (2) design specifications were transformed into quality inspection standards; (3) progress plans were converted into contractual terms and conditions; (4) quantitative calculations were converted into cost calculations; (5) test data were transformed into a quality evaluation report; (6) the quality evaluation report was converted into use permits; and (7) design specifications were converted into maintenance instructions.

To compute the information network metrics using UCINET, a network matrix was constructed to demonstrate the direct reference connections of all data items involved in the refurbishment project. Each column and row in the matrix table included data codes from the project documents (see Figure 3). A matrix table value of "1" indicated that the pairwise data had a referring connection, and otherwise, the values were "0". The network matrix had dimensions of n × n, where n is the number of data nodes in the information network. Similar data in different documents must be updated synchronously using online feedback from the other phases of the project's life cycle to effectively process change orders. For instance, information for heritage building refurbishment is subject to frequent bilateral shifting interactions, and related spatial information has additional applications. Notably, the matrix table can present the statuses of data adjustments by changing the order of the steps of the engineering process. Different data with similar names can thus be properly cited using different values to present the revised data. The matrix connection table is presented in Figure 3. The information network's metadata for the case study was as follows: possible maximum amount of edge = 14,638; amount of edge = 2869; and matrix density = 0.196.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25
1	L11	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	L12	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	L21	0	0	1	1	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	L31	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
5	L32	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0
6	L33	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	L34	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1
8	L41	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1	1	1
9	L42	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1
10	L43	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1
11	L44	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0
12	L45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
13	L46	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0

Figure 3. Part of the matrix table of information network connections.

5.4. Information Network Comparison and Evaluation Results

The case study results are presented as graphs and tables created using UCINET. Furthermore, the results were verified using expert review.

5.4.1. Information Network Metric Graphs

The data metrics were calculated, and the information network was constructed based on the case study. Figures 4–6, created with NetDraw, show the degree, closeness, and betweenness metric graphs of the information network using the data design and the construction phases from the case study. The sizes of the square spots between metric values indicate positive correlations, and connecting lines represent the relationships between the data pairs.

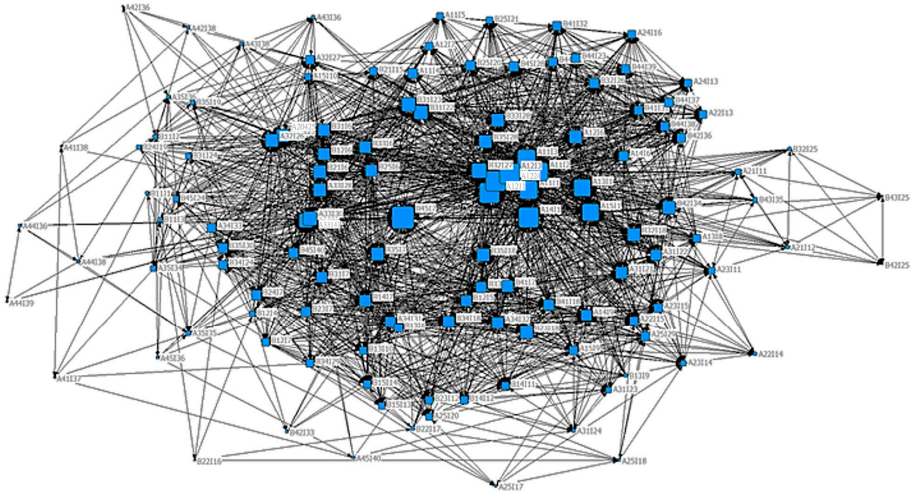

Figure 4. Degree metric graph of the information network using the design data vs. construction phases of the refurbishment of heritage buildings.

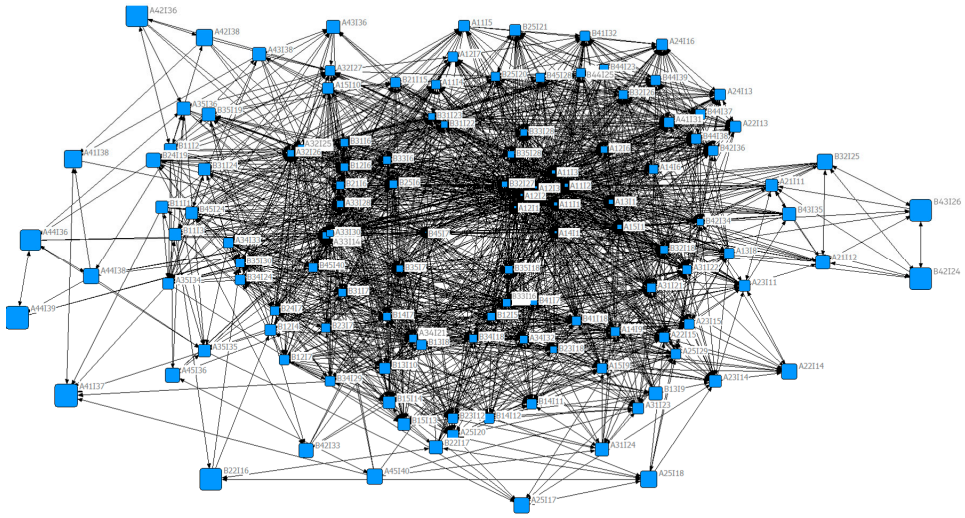

Figure 5. Closeness metric graph of the information network using the design data vs. construction phases of the refurbishment of heritage buildings.

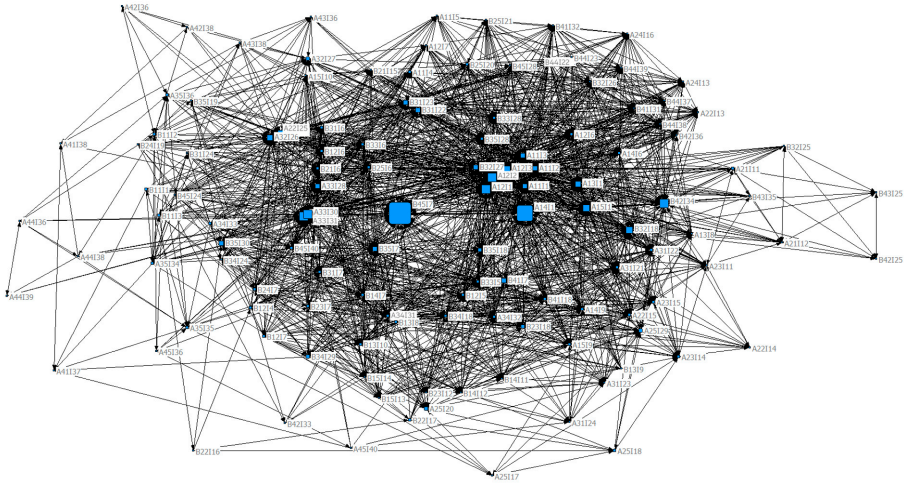

Figure 6. Betweenness metric graph of information network using the design data vs. construction phases of the refurbishment of heritage buildings.

5.4.2. Clustering Superior Data Using Metric Values

To emphasize the metric value trend, the researchers narrowed down the candidate data using the metric rankings of the top 15 information networks in the case study. Table 2 sequentially shows the results of the metric values of each candidate data set. Neighboring data pairs with lower value variation were classified into superior groups, as shown in Table 2. The boundary of the superior group was defined as the place where the gap between one pair of data obviously increased. The superior data groups of the three metrics are cited in the abovementioned description.

Table 2. Rank of the top 15 data items for the information network metrics.

Rank	Data Code	Metric of Degree	Data Code	Metric of Closeness	Data Code	Metric of Betweenness
1	B45I7	0.624	B45I7	0.795	B45I7	0.093
2	A12I1, A14I1, A12I2, A12I3	0.592	A14I1	0.769	A14I1	0.069
3	A11I1, A13I1, A11I2, A11I3	0.520	A12I1, A12I2, A12I3	0.747	A33I29, A33I30	0.034
4	A15I1	0.472	A13I1	0.721	A12I1, A12I2, A12I3	0.030
5	B32I27	0.440	A11I1, A11I2 A11I3	0.716	A32I26	0.025
6	A33I29 A33I30	0.392	A15I1	0.695	A32I25	0.024
7	A33I28	0.368	A33I29, A33I30	0.679	A13I1	0.023
8	A32I26	0.360	B32I27	0.677	A15I1	0.022
9	A32I25	0.352	A33I28	0.670	A11I1, A12I1, A11I3, A33I28	0.021
10	B35I7, B31I23	0.304	A32I26	0.649	B32I27, B35I30	0.017
11	B35I18	0.296	A32I25	0.646	B42I34	0.015
12	B33I28, B42I34	0.288	B31I23	0.642	A32I27	0.012
13	A12I6, A31I21, B31I22, B35I28	0.280	B35I7	0.638	B35I7, B32I18	0.011
14	B41I31	0.264	B33I28	0.635	B31I23	0.010
15	A34I31, B14I7, B41I7, B32I18, B43I36	0.248	B35I28	0.614	B35I18, B31I22	0.009

There were 8, 15, and 9 superior data points, which had, respectively, the highest degree, closeness, and betweenness metric values in the data sequence. To extract all of the critical data, the researchers compared the superior data groups using the top fifteen data points of each of the three metrics (as shown in Table 2). The results suggest that the union of the data from the three groups occurred at the 15th data point. The metric values of the top two data points ("construction record-detect checklist, B45I7" and "architecture drawing–plan view, A14I1") were clearly superior than the other top thirteen data points, as shown in Table 2. To reduce the manuscript's length, this study used practical observations for the two critical data items in order to validate the information network analysis results.

5.5. Verification for the Critical Data Result in the Information Network

To verify the importance of the extracted results, the critical data were compared to the construction information. The focus was information required by engineers and construction managers for heritage building refurbishment projects in Taiwan, based on their professional opinions on construction engineering practices. The verification of the experts' opinions can be described as follows.

5.5.1. Influence Analysis of Direct Citation of Data and Definition of the Core Data

This study found that the "construction record-detect checklist" was a critical data item, with the highest degree metric value of 0.774 and the highest betweenness metric value of 0.097. Both the central and bridge issue metrics show that the "construction record-detect checklist" reflects the discovery of undetected damage beneath the surface elements, thus indicating the modifications required for the original design content. In addition, data from the "construction record-detect checklist" comprise the design content, construction standard, and contract term.

Heritage building restoration projects require records of the original material, as well as the causes and statuses of damage as a reference for future maintenance work. Thus, the "construction record-detect checklist" is frequently used in the contract management, change order, and acceptance check processes. Furthermore, through the information delivery of the "construction record-detect checklist", the design content can be properly utilized in the construction phase. The data shift bilaterally, indicating a type of interactive information flow feedback between the design and construction phases.

The second-most important data item is the "architecture drawing–plan view," which was assessed using the degree, closeness, and betweenness metrics. Based on their professional experience, the engineers described the use of such critical data in practice as follows:

The "Architecture drawing–plan view" presents spatial information along with circulation, wall and floor decoration materials, and windows/doors. Such spatial information is fundamental for the development of many heritage building refurbishment documents. The referenced data type refers directly to the design development of other building elements, contributing to many other critical data. For example, the development of the "elevation view" based on the "plan view" cites two spatial data items related to construction performance, contract management, and outsourcing processes, with the "plan drawing" used to calculate the amount of work, construction site planning, and the work breakdown. Thus, the architectural drawing–plan view has been widely used in the contract management and work assignment processes. For example, according to the layout survey, the spatial data of the architectural drawing–plan view can visually indicate the work site and engineering material. The measurement of material processing and the position of the structural elements are the basic requirements for construction project quality management. The "equipment design" is developed based on the "plan drawing". Additionally, the "maintenance planning of building structure" necessitates the "plan drawing" in order to assess the feasibility of the building structure maintenance plan.

Notably, the types of material and construction methods used for the given specifications directly affect project scheduling and budgets, as they are derived from the critical

data of the "architecture drawing–plan view." As previously mentioned, this research found that the spatial information directly referred to the ordinal standard construction engineering process.

The "construction record-detect checklist" is referred to for design modification in response to undetected damage under the surface elements. Therefore, the "architecture drawing–plan view" is more frequently and directly cited than the "construction record-detect checklist." The superior data of the closeness metric are characterized by their direct influence on many tasks. This is the source of all data, and is referred to as "core data". The content of the core data is the basic material of much other data in the information network (e.g., the specification vs. the checklist).

5.5.2. Important Analysis Sequence for the Three Data Metrics and the Definition of Essential Data

Figure 7 compares the trends of the 3 metrics using their values the top 15 data items, thus presenting a holistic view of the critical data characteristics. The figure shows that the degree and closeness metrics are largely consistent, as is the trend for the three metrics of the top three data items, showing a positive correlation between the degree and betweenness metrics.

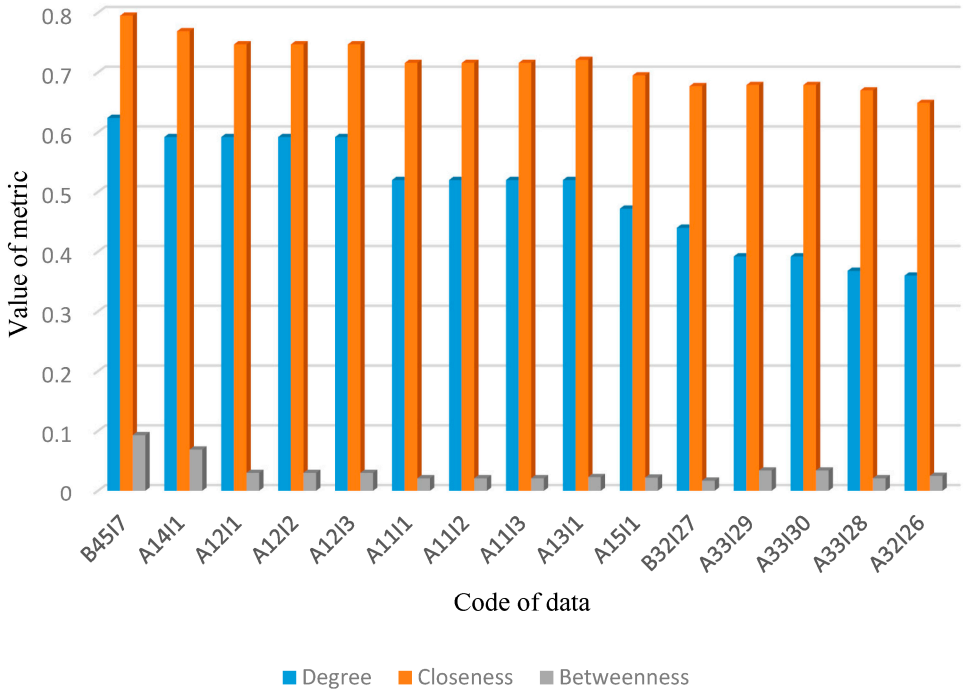

Figure 7. Comparison of metrics between the critical data items of the case study.

Furthermore, this study reviewed the metric equations of degree and betweenness from the perspective of information characteristics. The characteristics of the superior data differed between the degree and betweeness metrics. The superior degree metric data caused many tasks in the construction phase to adopt the data directly, while other tasks in the subsequent phase were required to be based on the same data to process the work (betweenness). These data are referred to as "essential data", and updating the content of the essential data items changes the content of much other data in an information network (e.g., plan drawings). Column 1 (degree metric) and Column 3 (betweenness metric) in Table 2 illustrate a consistent ranking trend of the superior data.

5.5.3. Discussion

The study's results show that the trend lines of the degree metric and the betweenness metric were consistent. from Considering the definition of the degree metric in Section 3.2,

this index represents the state of the critical data in the information network that were directly referenced by many other data items. The definition of the betweenness metric shows that the critical data were directly referenced by a single data item in each of the two clusters in the information network. However, the single data item linked to the critical data was directly referenced by the other data in the same data group, as it served the function of transmitting information in the data group to which it belonged. Therefore, critical data with high betweenness metrics were indirectly referenced by many other data items in the two clusters. These two indicators show that good critical data were directly or indirectly cited, leading to the same trend as that shown in Figure 7.

The closeness metric describes the transfer distance between two data items, and is intended to express the degree of direct association of data references. It is not related to the frequency of data referencing. If the critical data were sorted by their degree and between metric values, the closeness metric values were not in decreasing order. There was no consistency between the trend lines of the degree metric or betweenness metric values, nor the trend lines of the critical data values sorted by the degree metric or the betweenness metric.

In addition, the design concept of the relational database was intended to avoid the huge amount of data stored in hierarchical database forms and long data retrieval paths. Therefore, the data generation results of linked database forms needed to be cross-combined. The number of possible combinations was calculated by taking the number of columns in the previous table (PC) and the number of columns in the next table, minus one (because the prime key column of the duplicated column of the two tables needed to be deducted, FC-1), and multiplying the product of the two numbers [PC∗(FC-1)].

Furthermore, when the degree metric of a critical data item was high, it represented a large number of repeated occurrences in multiple tables in the same database. In this study, we used an example of a database table computation combination to illustrate the data redundancy effect of critical data application and how it improved the database. The same critical data appeared in two related tables, and the hypothetical pattern of removing duplicate fields was used to illustrate the phenomenon of data redundancy reduction. For example, if we were to calculate the original combination of two related tables, Table A (originally designed with 9 columns) and Table B (originally designed with 8 columns), there would have been 63 [=9∗(8-1)] possible combinations of computations. The critical data obtained from the filtering of the two tables according to the degree metric were used to delete the duplicate columns in the A- and B-related association tables. There were 54 [=9∗(8-1-2)] possible combinations of the two tables. As can be seen from this example, when a column of repeated data was reduced, the number of computations equal to the number of columns in the previous form was also reduced. The more the same critical data were repeated in the associated table, the more the reduction in data redundancy increased the computation.

Finally, from the case study, we learned that the spatial data is the critical data with the top rankings according to various metrics. Since the traditional DBMS was designed for the purpose of integrity, the form fields were based on the complete recording of various characteristic values of activities. Therefore, space, time, and individual names were mostly used as prime keys. Spatial data are important basic data for building projects, and have the ability to be used as the prime keys of traditional DBMS tables. This proves the reliability of the analysis model proposed in this study. Moreover, the development concept of the BIM spatial model as an interface for data retrieval can also prove the aforementioned point.

6. Conclusions

This study proposes an information network with customized metrics to resolve the fuzzy structure of information and complex data reference relations associated with practical construction project documents. It also presents a recurring feedback procedure for such projects using customized data processing. The case study shows that spatial data are the most frequently cited in the refurbishment of heritage buildings, and are also the

most important type of bridge data between the design and the construction phases. Two superior data from the case study were cited and verified—the "inspection record-checklist" and the "architect design-plan drawing"—which both showed identical rankings for the three metrics.

Notably, this study found consistent trends in the degree and betweenness metrics for superior data. In addition, critical data were included in the intersection data set of the two metrics of the winning data. When a construction project lacks a coordinated database system, the coordination of core and essential information may not be readily available, causing a loss of previous content and limiting its capacity to capture new information. There are different types of information usage between the essential and core data that must be processed using different data management standards. Properly managing the critical data can ensure that every member of the construction project has a similar version of the information which they can use for collaboration. The critical data for a construction project can be adopted to improve query efficiency, to reduce the scale of data warehouses, to develop standard data management procedures, and to assign sources of data management for BIM and blockchain applications.

Based on the findings and limitations of this study, the following research is recommended. First, critical data content must focus on the use of proper inputting forms and precise indications to ensure effective data utility and task performance. Second, critical data results cannot be used to sort the sequence of importance using the integrated data importance rank, which is a critical issue in determining how to integrate the three metrics of data in the information network. Moreover, the case study used in this research relied on current design and construction phase documents. To obtain a comprehensive view of the critical construction project data, other phases of the construction project must be included in the information extraction analysis. To precisely define the critical data and, thus, to indicate its relative importance, the threshold value of the metric must be objectively set. Moreover, different format types of critical data need to integrate to achieve data linking and to ensure that the information is delivered and combined.

The quality (accuracy and efficiency queries, etc.) of critical data is the basis of a construction project's efficient and successful completion. Additionally, the new challenges of data management are the protection of critical data from unauthorized access or limitations and the preservation of data versions and custody, particularly in the case of dispute resolution. Finally, the authors recommend the following criterial data management processes to enhance project performance. (1) Spatial data need consistent attributes (length, shape, and position) in different drawings. To avoid inconsistency in spatial data in drawings, a consistent 3D digital data model is required. (2) High-quality critical data need a platform on which to protect and exchange newly created content. BIM and blockchains can be applied to share and update information instantly, and the critical data can be contained in the database's backend. The combination application of data management can comprehensively serve to meet the needs of construction projects. (3) One item of data can be decomposed, providing more detail and enhancing the performance of data for a construction project.

Author Contributions: Conceptualization, C.-H.K.; methodology, C.-H.K.; software, C.-H.K.; validation, W.-T.C.; formal analysis, C.-H.K.; investigation, W.-T.C. and C.-H.K.; data curation, C.-K.H.; writing—original draft preparation, C.-K.H.; writing—review and editing, W.-T.C. and C.-H.K.; visualization, W.-T.C. and C.-K.H.; supervision, C.-H.K. and W.-T.C.; project administration, C.-K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cristóbal, J.R.S.; Carral, L.; Diaz, E.; Fraguela, J.A.; Iglesias, G. Complexity and project management: A general overview. *Complexity* **2018**, *154*, 121–132. [CrossRef]
2. He, X.; Liu, R.; Issa, R.R.A.; Anumba, C.J. A Data Tracking Tool to Improve Cost Management Reporting for Large Capital Projects. In Proceedings of the ASCE International Conference on Computing in Civil Engineering 2021, Orlando, FL, USA, 12–14 September 2021.
3. Mak, S. A model of information management for construction using information technology. *Autom. Constr.* **2001**, *10*, 257–263. [CrossRef]
4. Daher, P.; Dias, R.; Ergan, S. Owner requirements in as-built BIM deliverables and a system architecture for FM-specific BIM representation. *Can. J. Civ. Eng.* **2020**, *47*, 215–227.
5. Sheng, D.; Ding, L.; Zhong, B.; Love, E.D.P.; Luo, H.; Chen, J. Construction quality information management with blockchains. *Autom. Constr.* **2020**, *120*, 103373. [CrossRef]
6. Shahi, A.; Safa, M.; Haas, C.T.; West, J.S. Data fusion process management for automated construction progress estimation. *J Comput. Civ. Eng.* **2015**, *29*, 04014098. [CrossRef]
7. Ekanayake, B.; Wong, K.W.; Fini, A.; Smith, P.V. Computer vision-based interior construction progress monitoring: A literature review and future research directions. *Autom. Constr.* **2021**, *127*, 103705. [CrossRef]
8. Solihin, W.; Eastman, C.M.; Lee, Y.; Yang, D.H. A simplified relational database schema for transformation of BIM data into a query-efficient and spatially enabled database. *Autom. Constr.* **2017**, *84*, 367–383. [CrossRef]
9. Chen, W.; Das, M.; Chen, K.; Cheng, C.P. Ontology-based data integration and sharing for facility maintenance management. In Proceedings of the Construction Research Congress 2020, Tempe, AZ, USA, 8–10 March 2020.
10. Beach, T.; Petri, I.; Rezgui, Y.; Rana, O. Management of collaborative BIM data by federating distributed BIM models. *J. Comput. Civ. Eng.* **2017**, *31*, 04017009. [CrossRef]
11. Liu, Z.; Lu, Y.; Peh, L. A review and scientometric analysis of global building information modeling (BIM) research in the architecture, engineering and construction (AEC) industry. *Buildings* **2019**, *9*, 210. [CrossRef]
12. Noor, B.A.; Yi, S.R. Review of Building information modeling in construction industry: Meta-analysis. *Constr. Innov.* **2018**, *18*, 433–452. [CrossRef]
13. Pan, Y.; Zhang, L. A BIM-data mining integrated digital twin framework for advanced project management. *Autom. Constr.* **2021**, *124*, 103564. [CrossRef]
14. Khaja, M.; Seo, J.D.; McArthur, J.J. Optimizing BIM metadata manipulation using parametric tools. *Procedia Eng.* **2016**, *145*, 259–266. [CrossRef]
15. What Is Blockchain Technology? Available online: https://www.ibm.com/topics/blockchain (accessed on 22 May 2023).
16. Li, J.; Greenwood, D.; Kassem, M. Blockchain in the built environment and construction industry: A systematic review, conceptual models and practical use cases. *Autom. Constr.* **2019**, *102*, 288–307. [CrossRef]
17. Sacks, R.; Koskela, L.; Bhargav, A.D.; Owen, R. Interaction of lean and building information modeling in construction. *J. Constr. Eng. Manag.* **2010**, *136*, 14157555. [CrossRef]
18. Leung, N.N.; Chan, S.L.; Issa, R.R.A. Meta-Data-Based Collaboration in Construction Project Management. In Proceedings of the Fourth Joint International Symposium on Information Technology in Civil Engineering, Nashville, TN, USA, 15–16 November 2003.
19. Rob, P.; Coronel, C. *Database Systems: Design, Implementation, and Management*, 13th ed.; Cengage Learning: Boston, MA, USA, 2011; p. 88.
20. Rasdorf, W.J.; Abudayyeh, O.Y. NIAM conceptual data-base design in construction management. *J. Comput. Civ. Eng.* **1992**, *6*, 41–62. [CrossRef]
21. Bansal, S.; Sharma, P.C. 5—Classification criteria for data deduplication methods. In *Data Deduplication Approaches-Concepts, Strategies, and Challenges*; Thwel, T.T., Sinha, G.R., Eds.; Academic Press: Cambridge, MA, USA, 2021; pp. 69–96.
22. Rob, P.; Coronel, C. *Database Systems: Design, Implementation, and Management, Cengage Learning*, 14th ed.; Cengage Learning: Boston, MA, USA, 2022; p. 88.
23. Giorginisa, T.; Ougiarogloubc, S.; Evangelidisc, G.; Dervo, D.A. Fast data reduction by space partitioning via convex hull and MBR computation. *Pattern Recognit.* **2022**, *126*, 108553. [CrossRef]
24. Melo, V.S.; Panisson, A.R.; Bordini, R.H. Meta-information and argumentation in multi-agent systems. *iSys-Braz. J. Inf. Syst.* **2017**, *10*, 74–97. [CrossRef]
25. Chowdhury, S.; Zhu, J. Investigation of critical factors for future-proofed transportation infrastructure planning using topic modeling and association rule mining. *J. Comput. Civ. Eng.* **2023**, *37*, 04022044. [CrossRef]
26. Nettleton, D. Data mining of social networks represented as graphs. *Comput. Sci. Rev.* **2013**, *7*, 1–34. [CrossRef]
27. Tao, B.; Dai, H.N.; Wu, J.; Ho, W.H.; Zheng, Z.; Fong, C.C. Complex network analysis of the bitcoin transaction network. *IEEE Trans. Circuits Syst. II Express Briefs* **2022**, *69*, 1009–1013. [CrossRef]
28. Borgatti, S.P.; Everett, M.G.; Johnson, J.C. *Analyzing Social Networks*; Sage: London, UK, 2022.
29. Brandhorst, R.; Krzyzowski, L. Biographical Reconstructive Network Analysis (BRNA): A Life Historical Approach in Social Network Analysis of Older Migrants in Australia. *Forum Qual. Soc. Res.* **2022**, *23*, 4.
30. Yuret, T. Network analysis of econometric society fellows. *Scientometrics* **2022**, *127*, 7615–7631. [CrossRef]

31. Arif, T.; Ali, R.; Asger, M. Social network extraction: A review of automatic techniques. *Int. J. Comput. Appl.* **2014**, *95*, 975–8887. [CrossRef]
32. Chen, W.; Figueiredo, C.; Azevedo, J. Central nodes and surprise in content selection in social networks. *Comput. Hum. Behav.* **2015**, *51*, 382–392.
33. Mel, A.; Kang, B.; Lijffijt, J.; De Bie, T. Fondue: A framework for node disambiguation and deduplication using network embedding. *Appl. Sci.* **2022**, *11*, 9884. [CrossRef]
34. Brandes, U.; Borgatti, S.P.; Freeman, L.C. Maintaining the quality of closeness and betweenness centrality. *Soc Netw.* **2016**, *44*, 153–159. [CrossRef]
35. Prajnawrdhi, T.A.; Karuppannan, S.; Sivam, A. Preservation of cultural heritage buildings is an action in maintaining the sustainability of architectural diversity in Indonesia. *Procedia Environ.* **2015**, *28*, 557–566. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.