Bayesian Monte Carlo Simulation Driven Approach for Construction Schedule Risk Inference

Long CHEN[1], Ph.D. A.M.ASCE; Qiuchen LU[2*], Ph.D. A.M.ASCE; Shuai LI[3], Ph.D.; Wenjing HE[4]; Jian YANG[5], Ph.D.

[1]Lecturer, School of Architecture Building and Civil Engineering, Loughborough University, Loughborough LE11 3TU, UK. Email: l.chen3@lboro.ac.uk

[2]Lecturer, The Bartlett School of Construction and Project Management, University College London, London WC1E 6BT, UK. Email: qiuchen.lu@ucl.ac.uk

[3]Assistant Professor, Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA. Email: sli48@utk.edu

[4]Master student, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China. Email: qzclwenj@sjtu.edu.cn

[5]Professor, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China. Email: j.yang.1@sjtu.edu.cn

ABSTRACT

As the construction of infrastructures becomes increasingly complex, it has been often challenged by construction delay with enormous losses. The delivery of complex infrastructures provides rich source of data for new opportunities to understand and address schedule issues. Based on these data, many efforts have been made to identify key construction schedule risks and predict the probability of risk occurrence. Bayesian network is one of the most useful tools for risk inference. However, there are still two obstacles preventing the Bayesian network from being adopted popularly in construction schedule risk management: (1) the development of directed acyclic graph (DAG) and associated conditional probability tables (CPTs); (2) the lack of observation data to trigger risk inference as evidence at the planning stage. This research aims to develop a novel Bayesian Monte Carlo simulation driven approach for construction schedule risk inference of infrastructures, where the Bayesian network model can be developed in a more convenient way and be used without observation data required. It firstly constructs the key risk network with key risks and links through network theory-based analysis. Then the DAG structure of Bayesian network is developed based on the topological structure of key risk network using deep-first search (DFS) and adapted maximum-weight spanning tree (A-MWST) algorithms. The CPTs are further developed using the leaky-MAX model. Finally, the Bayesian Monte Carlo simulation driven risk inference method is developed for predicting and quantifying the probability of construction schedule risk occurrence. A real

* Corresponding Author, Qiuchen Lu, Email: qiuchen.lu@ucl.ac.uk

infrastructure project is selected as case study to verify this developed approach. The results show that the developed approach is more appropriate to deal with risk inference of infrastructures considering its reliability, convenience, and flexibility. This research contributes a new way to construction schedule risk management and provides a novel approach for quantifying and predicting risk occurrence probability.

Keywords: Construction schedule risks; Network theory-based analysis; Bayesian Monte Carlo simulation

1. Introduction

The construction industry, especially the infrastructure sector, has been rapidly developing in the past decades, where the average annual growth of global built-up areas is 28% between 1990 and 2015 (UN-Habitat, 2016). The construction of infrastructures has attracted many attentions due to the extreme complexity, high risks and long lead time (Fiori and Kovaka, 2005), multiple stakeholders involved (Mok et al., 2017; Valentin et al., 2018), and considerable impacts to the society, economy and natural environment (Zhai et al., 2009). It is thus significant to manage infrastructure projects properly not only for the project itself but for the region concerned (Wang and Yuan, 2016).

As a major objective of infrastructure management, the schedule has been regarded as a vital parameter of the planning and construction of projects (Al-Momani, 2000; Liu et al., 2014). However, the construction schedule delay of infrastructures is not rare but a 'business as usual' with enormous losses and damages to project delivery (Flyvbjerg et al., 2004, 2005). For example, Al-Momani (2000) found that 81.5% of 130 public complex construction projects involved in Jordan had suffered time delay. Ellis and Thomas (2002) found that 55% of highway projects in the US had experienced an average time delay of 44% in excess of the original time specified in the contract. The construction schedule and risk management should be improved to help the delivery of infrastructures within time and budget.

Such complex infrastructures are becoming rich sources of data from multidisciplinary models and systems, raising new opportunities to understand and address schedule issues (Whyte et al., 2016; Chen and Whyte, 2020). Based on these data, many efforts have been made to manage the construction schedule and associated risks of infrastructures. Building information modelling (BIM) has been adopted as a powerful management tool for schedule planning (Liu et al., 2015; Tallgren et al., 2020), control (Moon et al., 2015) and risk management (Sami Ur Rehman et al., 2020). Based on that, extended reality (XR) has been further proposed to assist schedule control (Alizadehsalehi et al., 2020; Fu and Liu, 2018). However, such technologies only focused on pre-defined schedules and risks while ignored the risk impacts on construction schedule. Other methods more focusing on the risk impacts have been also proposed, such as line of balance (LOB), critical path method (CPM), program evaluation and review technique (PERT) and correlated schedule risk analysis model (CSRAM) (Ökmen and Öztaş, 2008). However, they only focused on the diversity of risks or effect of risk on the schedule while ignored the complexity of correlations between risks and uncertainty of risk occurrence, where risk impacts that appear to be minor lead to rippled disruption to the project delivery (Abotaleb and El-adaway, 2018). Bayesian network has been proposed as a useful tool to handle complexity and uncertainty and been adopted in the construction schedule risk management of infrastructures (Nasir et al., 2003; Luu et al., 2009; Khodakarami et al., 2007). It provides a reliable approach to (1) modelling complex risk correlations by cause-effect relationships, and (2) modelling uncertainty of risk occurrence by conditional probability. However, given the richer data, there are still two obstacles preventing the popular adoption of Bayesian network in the practice of construction schedule risk management: (1) it is often impossible in practice to define and use a unified classification code for risk identification and data template for Bayesian network development, especially for the development of directed acyclic graph (DAG) and associated conditional probability tables (CPTs); (2) the observation data is often required as evidence to trigger risk inference which is inapplicable at the project planning stage.

This paper aims to solve these problems through developing a novel Bayesian Monte Carlo simulation driven approach for construction schedule risk inference of infrastructures. This developed approach enables Bayesian network to be developed in a more convenient way and be used without observation data required in practice. It starts with a review of construction schedule risks identified in the past research and theoretical basis of network theory and Bayesian network in construction schedule risk analysis. Then the methodology is explained in terms of strategies of case study and focus group discussions for case data collection. The construction of key risk network is demonstrated using network theory-based analysis, followed by the development of Bayesian network model through DFS and A-MWST algorithms and leaky-MAX model. Finally, the Bayesian Monte Carlo simulation driven risk inference approach is developed for predicting and quantifying the probability of construction schedule risk occurrence. A real underground railway project is selected as case study to verify the developed approach. The results are discussed in terms of reliability, convenience and flexibility, followed by conclusions.

2. Literature Review

2.1 Construction schedule risks of infrastructure project

As the beginning of the risk management, risk identification is a process through which a series of potential risks are identified according to their frequency of occurrence and possible influence, either positively or negatively, on principal project objectives (Perrenoud et al., 2015). Many efforts have been made to identify risks or causes for the construction delay in infrastructures (Table 1). For instance, Lo et al. (2006) investigated the 30 significant factors from seven categories, and further provided corresponding risk mitigation measures in Hong Kong. 'Inadequate resources due to contractor/lack of running capital', 'Unforeseen ground conditions', and 'Exceptionally low bids' have been identified as the three most significant causes for delay in civil engineering projects in Hong Kong. Sambasivan and Soon (2007) also explored the construction delay factors and associated impact on the schedule of large construction project in Malaysia through questionnaire survey and identified 10 most important causes for delay (e.g., contactor's improper planning, poor site management, and inadequate contractor experience.

Some research has been further conducted to reveal insights of construction delay of specific infrastructure project. Han et al. (2009) analysed the construction schedule delay of Korea Train Express (KTX) project in South Korea, and identified five major delay causes for KTX project, including 'Lack of owner's abilities and strategies to manage hi-tech oriented mega project', 'Frequent changes of routes', 'Inappropriate project delivery system', 'Lack of proper scheduling tool' and 'Redesign and change orders of main structures and tunnels'. The time overruns of six FIFA World Cup stadia in South Africa have been investigated by Baloyi and Bekker (2011), examining 11 main factors causing the construction schedule delay. These identified risks in previous research provide a list of potential construction schedule risks for infrastructures both in practice and in this research.

[Insert: **Table 1** Schedule risks for the construction delay of infrastructure project identified in previous research]

2.2 Network theory in risk analysis

Network theory is a view of regarding the dependent elements as the complex network with multiple correlations between them, which is concerned with the 'structure and patterning' of these correlations and seeks to identify both their causes and effects (Yang and Zou, 2014). Evolving from the network theory, the network theory-based analysis can provide an effective way to identify the correlations between system elements and analyse the features and implications of these relational fabrics by integrating mathematical and computational applications (Mok et al., 2017; Dogan et al., 2013). Within the network structure, elements (nodes) of a system are joined by multiple correlations (links) in various manners. The network theory-based analysis accentuates network and relational measures rather than the individual attributes of each element on account of the conception that: (1) the existence of an element can yield effects on or be affected by the presence of other interrelated elements within the system; and (2) the correlations between system elements can the system's strength and behaviours (Fang et al., 2012; Mok et al., 2017).

The network theory-based analysis has been initially applied in the sociometry, and then adopted in construction management. Fang and Marle (2012) developed a simulation-based risk network model for decision support in project risk management (PRM), which defined risks as nodes and correlated influences as links. Furthermore, Fang et al. (2012) analysed the risk network in a large engineering project to distinguish key risks and correlations affecting the project objectives. Yang and Zou (2014) investigated stakeholder-associated risks and their relationships in green building projects to facilitate risk management. Mok et al. (2017) studied stakeholders concerns and intricate interdependencies between them for identifying key challenges in major public engineering projects (MEPs).

Although previous research has showed the viability of network theory-based analysis in analysing the complexity of correlations between risks, it cannot deal with the uncertainty of risk occurrence.

2.3 Bayesian network in risk analysis

The Bayesian Network provides a probabilistic approach to determine the likelihood of occurrence of certain variable conditions (i.e., conditional probability) and can model the uncertainty of risk occurrence (Nasir et al., 2003; Wang and Zhang, 2018). It firstly introduced in the 1970s (McCabe et al., 1998), is a graphical representation of conditional dependence among a group of variables. A Bayesian network model usually consists of two parts: (1) a directed acyclic graph (DAG), which represents the network structure, and (2) an associated set of conditional probability tables (CPTs), which is the network parameter and represents the conditional probability distribution among the variables (Hu et al., 2013). Specifically, the nodes represent variables of the domain, while the arcs represent dependence relationships between the nodes (McCabe et al., 1998). The network is thus constructed by a series of nodes where the nodes are connected by the arcs according to the reasoning direction of decision makers (Kjaerulff, 2008). Based on the Bayes' theorem (Equation (1)), the relationship between each pair of connected nodes is expressed in the form of probability distribution that encapsulates the decision makers' experience (Kjaerulff, 2008). Therefore, for a Bayesian network model, $B = (V, E)$, where V denotes a set of nodes (i.e., variables), and E denotes a set of directed links between pairs of the nodes, a joint probability distribution that can be factorized as:

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \tag{1}$$

$$P(V) = P(X_1, X_2, \cdots X_i, \cdots, X_n) = \prod_{X_i \epsilon V} P(X_i | X_{pa(X_i)}) \tag{2}$$

where $X_{pa(X_i)}$ is the set of parent nodes for X_i, $X_{pa(X_i)} \in V$. Each conditional probability distribution for ith variable, $P(X_i | X_{pa(X_i)})$, consists of a series of conditional probability p_i:

$$p_i = P\left(x_i | x_{pa(X_i)}\right), i = 1, 2, \cdots, I \tag{3}$$

where x_i and $x_{pa(X_i)}$ are the values assigned to X_i and $X_{pa(X_i)}$ respectively, and p_i is the probability of $X_i = x_i$ given the condition $X_{pa(X_i)} = x_{pa(X_i)}$.

The Bayesian network has been adopted in addressing the complex problems under uncertainty in the real world, such as fault diagnosis (Lin et al., 2018), ecosystem assessment (Liu et al., 2014), decision support (Xie and Thomas, 2013), and of course risk analysis of infrastructures. Specifically, many Bayesian network-based studies have been conducted to evaluate the construction schedule risks for infrastructures. For example, Luu et al. (2009) applied the Bayesian network to quality the construction schedule risks and the probability of construction project delays in Vietnam. Nasir et al. (2003) developed an Evaluating Risk in Construction-Schedule Model (ERIC-S) through integrating the PERT technique and Bayesian network model, which could be adopted to determine the lower and upper activity duration values for schedule risk analysis of infrastructure projects. Furthermore, Khodakarami et al. (2007) mapped the CPM to Bayesian networks to provide the prediction for the construction schedule under uncertainty.

Through applying Bayesian network, the complexity of risk correlations and uncertainty of risk occurrence can be modelled properly in construction schedule risk analysis of infrastructures. However, the development and application of Bayesian network is not easy which usually require large amount of data and time for (1) DAG and CPTs development; and (2) risk inference. Unfortunately, there are usually limited time and data provided for Bayesian network development and application in practice, preventing the popular implementation of Bayesian network in the risk analysis of infrastructures. It is time to develop a novel approach for improving the reliability, convenience and flexibility of Bayesian network development and application in construction schedule risk analysis of infrastructures.

3. Research Methodology

The novel approach has been developed and validated in this four-stage research for construction schedule risk inference through integrating network theory-based analysis and Bayesian Monte Carlo simulation (Figure 1).

The network theory-based analysis is firstly conducted to identify key risks (nodes) and correlations (links) and construct the key risk network for construction schedule. The topological structure of key risk network is similar to that of DAG in Bayesian network model, where the nodes represent risks and links represent the cause-effect relationships. The development of DAG in Bayesian network model can then be developed based on the constructed key risk network using DFS and A-MWST algorithms. The CPTs of Bayesian network model can be further worked out using the leaky-MAX model. Based on the developed Bayesian network model, the Bayesian Monte Carlo simulation is conducted for generating reliable probability of each possible state of risk occurrence (i.e., 'Better than expected', 'Expected' or 'Worse than expected'). Finally, the case study is conducted according to a real underground project for approach validation.

[Insert: **Figure 1** Overall flowchart for research]

In this research, the main structure for tunnelling of underground railway project in Greater Bay of China was chosen for case study with the considerations: (1) the project is representative for its type (transportation), complexity, importance and size, (2) the data was accessible to conduct this research, (3) construction schedule delay has already happened or is highly possible to occur in the construction process, and (4) researchers have built a good relationship with the project team, which helped secure the access to the project for further study. Two focus group discussions have been organized with five senior members from the project management team. Although ideally all stakeholders should participate in discussion to achieve a consensus, it is more efficient in practice to have only senior members from the project management team involved to provide enough information (Yang and Zou, 2014). All participants in focus group discussions were selected based on the idea of 'Applicability', in which participants should have rich knowledge and something to say on the discussion topic, have similar socio-characteristics, and be comfortable talking to each other. The data generated based on the synergy of group interactions could thus be applied to the development of hybrid approach.

3.1 1st round focus group discussion

The 1st round focus group discussion was designed for developing key risk network of case project. It lasted for around 2 hours and mainly focused on two aspects: (1) the identification and verification of construction schedule risks within the context of case project; (2) the identification and assessment of links among schedule risks within the context of case project. Before the 1st round focus group discussion, a brief introduction about network theory-based analysis has been presented to the participants, and a list of potential construction schedule risks from literature review (Table 1) has also been provided to participants as a reference to break their cognitive limitations.

The post-discussion log and notes were kept well, which recorded the information related to the double-check of whether the recording is functioning properly, the researcher's reflections and elaborations about the focus group discussion, and the learning from the discussions. The post-discussion notes coupled with the main data collected from the discussion notes ensure the quality and reliability of the data for analysis.

3.2 2nd round focus group discussion

The 2nd round focus group discussion was designed for developing Bayesian network model according to the case project, which also involved the same five senior members participating in the 1st round focus group discussion. It lasted for around 2 hours and mainly focused on two topics: (1) verification and examination of developed DAG structure of Bayesian network model; and (2) development of CPTs of Bayesian network model. The post-discussion log and notes were also kept well with those from the 1st round focus group discussion.

4. Development of Key Risk Network

4.1 Identification of the network boundary

As the foundation of developing key risk network, the boundary (i.e., specific risks) should be identified and examined at first.

The classical experience-based method is one of the most popular methods for risk identification. It includes only core stakeholders to perform the risk identification process, which is conducted based on a stakeholder's or a small group of stakeholders' experiences on 'what are the risk categories' and 'what are the risks' by interviews, surveys or focus group discussions. It is convenient and highly efficient to provide insights into risks according to the rich experience of core stakeholders, but it is difficult for the core stakeholders to break the cognitive limitations and draw the whole set of boundaries (Chen, 2019; Yang and Zou, 2014).

In this research, the classical experience-based method was adopted to identify risks for constructing risk network through the 1st round focus group discussion. Before this focus group discussion, a list of potential construction schedule risks from literature review (Table 1) was also provided to participants as a reference to help them break their cognitive limitations and draw comprehensive boundaries.

4.2 Establishment and assessment of links

After defining the risk network boundary, the links between risks in this research are considered between each pair of risks (Fang et al, 2012). The risk structure matrix (RSM) method is commonly adopted to analyse risk links, which was also adopted in this research.

The RSM (i.e., adjacency matrix) is defined as a square matrix with entry $RSM_{ij} = S_{ij} = I_{ij} \times P_{ij}$ when there is a relationship from ith risk, R_i, to jth risk, R_j, otherwise, $RSM_{ij} = Null$, where S_{ij} is the strength of link, I_{ij} is the intensity of impact from this one risk to the other paired, and P_{ij} is the likeliness of this impact to happen (Mok et al., 2017; Yang and Zou, 2014). The five-point Likert scales are adopted to measure I_{ij} (from 1 = 'No impact' to 5 = 'Extraordinarily significant impact') and P_{ij} (from 1 = 'Never happen' to 5 = 'Always happen').

In order to moderate the confusion and divergence of links establishment and assessment, the 1st round focus group discussion was held to develop the RSM with quantitative assessment (Yang and Zou, 2014). The outcomes can identify and quantify the links between risks.

4.3 Visualisation of network

Once the nodes and links have been identified and assessed, a construction schedule risk network for the target infrastructure project can be developed and mapped as a graph $G(N, K)$, where the identified risks are mapped as N nodes connected by K weighted arrows.

In this research, the NetMiner 4 was used to visualise the risk network for its high competence in the processing and exploratory analysis of huge networks (Furht, 2010). In the network graph $G(N, K)$ presented, nodes represent the risks where different shapes and colours of them indicate different risk categories and sub-categories respectively. The arrows are the links between risks, of which the thicknesses indicate the strength of links.

4.4 Topological analysis of risk network

With risk network mapped as $G(N, K)$, the structural configuration is explored and explained by the metrics of topological analysis (Table 2).

This analysis consists of three levels. Firstly, through the network-level analysis, the network density and cohesion are calculated out to unravel the network structure quantitatively. The value of density indicates how closely the risks are situated in a network, and the value of cohesion implies the complicated of network configuration in terms of node reachability. Then the node-level analysis is further conducted to determine the key risks through examining the direct and/or propagating impacts of nodes, as well as their functions and properties in the influence network. Five node-level metrics were calculated and analysed in this research, namely, degree difference, ego network size, node betweenness centrality, out-status centrality, and total brokerage (Table 2). Finally, the link-level analysis is conducted to measure the extent that a risk link plays a gatekeeper role in governing the influences passing through it based on betweenness centrality (Chen, 2019; Yang and Zou, 2014). A greater centrality value implies a more critical link.

[Insert: **Table 2** Definition of metrics for topological analysis]

4.5 Interpretation of the results

Based on the results of analysis at three levels, the key risks and key risk links can be identified. The key risks are distinguished from the risk network with high values in one or more of nodal metrics, including degree difference (D_d), ego network size (E), betweenness centrality (B), out-status centrality (S), and brokerage. Meanwhile, the key risk links are identified with high values in betweenness centrality at the link-level.

The key risk network thus consists of (1) key risks, (2) key risk links, (3) non-key risks involved in key risk links, and (4) non-key risk links involving key risks. This developed key risk network provides essential information of construction schedule risks for infrastructures but with more concise and manageable structure (Chen, 2019).

5. Development of Bayesian Network

5.1 The construction of DAG structure

The construction of DAG can provide a network structure for Bayesian network model, where two kinds of methods have been commonly adopted, namely the expert knowledge driven structure construction method (Hu et al., 2013; Luu et al., 2009), and the observational data driven structure learning method (Lee et al., 2009). However, the structure learning method is not appropriate to be applied in the field of infrastructures due to (1) the uniqueness and uncertainty of construction schedule risks for infrastructures; and (2) the data provided for training process. Although the structure construction method conforming to the verified causalities is more suitable for risk analysis of infrastructures, it can be time-consuming for construction process and inevitably introduce subjective bias from experts (Hu et al., 2013).

Due to limited time and data, it is reasonable to use the key risk network from network theory-based analysis as basis to generate the DAG structure considering that the topological structure of key risk network is similar to that of DAG in Bayesian network model, where the nodes represent risks and links represent the cause-effect relationships. This novel approach integrating the network theory and Bayesian network is not only more convenient and resource-saving but also reliable for incorporating both expert knowledge and analysis metrics (Table 2).

In order to transform the network from directed cyclic graph (DCG) to DAG properly, it is the key to find the directed cycles in network and eliminate these cycles without essential information loss, where the directed cycles are formed through 'starting at any vertex v and following a consistently-directed sequence of edges that eventually loops back to v again'. In this research, there are two steps developed to construct DAG from key risk network, including (1) searching cycles by DFS algorithm, and (2) constructing DAG by A-MWST algorithm.

5.1.1 Searching cycles by DFS algorithm

The DFS algorithm is adopted as the searching strategy on account of its convenience and rapidity to traverse or search the tree or graph data structures as far as possible (Cormen et al., 2001). The pseudocode of recursive DFS algorithm is as follows (Goodrich and Tamassia, 2006):

```
DFS (G, v)
        label v as discovered
        for all edges from v to w in G.adjacentEdges(v) do
                if vertex w is not labeled as discovered then
                        recursively call DFS (G, w)
```

Based on recursive DFS algorithm, the process of applying DFS algorithm in searching cycles within the key risk network is developed as follows: (1) Define the key risk network as the graph G with vertices v, (G, v), and define $weight == B(R_i \rightarrow R_j)$; (2) Select an unvisited node v with $D_{in}(v) = 0$ as the root node; define $[v] == 0, [w] == 0$, label v as 'visited', and begin the searching procedure; (3) Check whether $G.adjacentEdges(v) == \{\}$ or not, and if the answer is yes, then go to step 6; otherwise, go to step 4; (4) For all the available edges (i.e., links) in $G.adjacentEdges(v)$, select the edge with the maximum weight e which directs to the vertex w; define $[w] == [w] + 1, G.adjacentEdges(v) == G.adjacentEdges(w) \backslash \{e\}$; (5) Define $[v] == [w]$, and check whether the w (or v) has been labelled or not; if the answer is yes, then go to the step 6; otherwise, label w as 'visited' and go to the step 3; (6) Define $[v] == [v] - 1$; check whether $[v] == 0$ (i.e., the root node), if the answer is no, go to step 3; otherwise, further check whether all the nodes have been visited or not, if the answer is no, go to step 2, otherwise, stop the searching procedure; (7) Transform the key risk network (G, v) into spanning tree through integrating the visited nodes and traversed edges (Cormen et al., 2001).

There are usually four types of edges (i.e., links) in the spanning tree (Cormen et al., 2001): (1) the tree edges which belong to the spanning tree itself, (2) the forward edges which point from a node of the tree to one of its nonadjacent descendants, (3) the back edges which point from a node to one of its ancestors, and (4) the cross edges which do neither. The cycle must be existed if the back edge is existed, through which the cycles can be identified in the risk network.

5.1.2 Constructing DAG by A-MWST algorithm

With the cycles identified, the spanning tree transformed from key risk network need to be re-developed to construct DAG structure through eliminating the identified cycles without essential information loss.

The MWST algorithm can highly reserve the structure properties and provide the associated probability distribution closest to the probability distribution of the original network, as measured by the Kullback-Leibler divergence (KLD) (Pearl, 1988). Based on the MWST algorithm, the A-MWST algorithm has been developed in this research to re-developed DAG model from spanning tree:

```
A-MWST_DAG (G, v)
        S ≔ {}
        Loop ≔ {}
        for each vertex v in G
                do v → S
        sort(G.Edges(v → w), B(v → w), 'descend')
        for each edge from v to w in G.Edges(v → w)
                do S ≔ S ∪ {(v → w)}
                if Loop ≔ {}
                        then S ≔ S
                        else S ≔ S\{(v → w)}
        recursively call G.Edges(v → w)
```

In this A-MWST algorithm, the betweenness centrality of link, $B(R_i \rightarrow R_j)$, is a reliable metric for mutual information $I(R_i \rightarrow R_j)$ of corresponding edge and has been defined as the weight of corresponding edge. The process of applying A-MWST algorithm to constructing DAG based on spanning tree is designed as follows: (1) starting from the empty tree over all variables (i.e., nodes); (2) inserting the largest-weight edge (i.e., link); (3) finding the next largest-weight edge and adding it to the tree if no cycle is formed; otherwise, discarding the edge and repeating this step; and (4) repeating the third step until all edges have been selected and an associated DAG is finally constructed whose weight has the maximum value of $\sum_{(v \to w), v, w \in G} B(v \to w)$.

5.2 The development of CPTs

The development of CPTs is another obstacle preventing the adoption of Bayesian network in the practice of construction schedule risk management with limited time and data, whose complexity increases exponentially with the number of parent nodes and possible values (or states) of nodes (Xie and Thomas, 2013; Zagorecki and Druzdzel, 2013). For example, in a multi-valued Bayesian network with m possible values, there can be m^{n+1} conditional probabilities in the CPT for a node with n parent nodes. Apart from the huge amounts of time and data it requires to assess all the probabilities for CPTs, it can also be problematic to what extent experts can be expected to coherently provide the probabilities at the level of detail required (Wisse et al., 2008).

In order to relieve the elicitation task for developing the CPTs, there are two kinds of ways commonly adopted (Wisse et al., 2008). The first one is to provide an easier way for experts to deliver CPTs. For instance, Van Der Gaag et al. (1999) transcribed the CPTs using the scale of both numerical and verbal anchors. However, the efforts to deliver the full CPTs though reduced is still exponential in the number of variables. The second way to reduce these efforts is to reduce the number of probabilistic assessments to be made, for example, reducing the number of variables (e.g., Luu et al., 2009) or limiting the number of possible values (or states) (e.g., Xie and Thomas, 2013). However, such a way will lead to the unavoidable loss of information (Wisse et al., 2008).

Considering both the efforts reduction and information reservation, the canonical probabilistic model was adopted in this research for reducing the number of probabilities to be specified through delivering approximate probabilities (Xie and Thomas, 2013; Wisse et al., 2008). One of the most widely used technique among canonical probabilistic models is the noisy-MAX model introduced by Diez (1993), the generalization applied in addressing multi-valued variables of the noisy-OR model. In this noisy-MAX model, the CPT is derived from the 'marginal conditional' probability distributions specified for each parent using the max function (Diez, 1993). The noisy-MAX model just requires a small number of parameters to specify the entire CPTs, which is linear in the number of conditioning variables rather than exponential (Wisse et al., 2008). It significantly reduces the efforts in knowledge elicitation from experts (Wisse et al., 2008), improves the quality of distributions learned from data (Oniśko et al., 2001), and reduces the special and temporal complexity of algorithms for Bayesian networks (Diez and Galán, 2003).

However, in practice, it is neither feasible nor desirable to model all variables influencing a certain node Y (Diez and Druzdzel, 2006). According to Diez and Druzdzel (2006), in this case, assuming that there is a large Bayesian network that properly represents the real-world domain defined over a set of variables V^R, the reduced model exploited can be defined as V ($V \subset V^R$), and the rest of the variables, $V_I = V^R \backslash V$, are not explicit in the model where the index I means 'implicit' (Figure 2(a)). The leaky model only models the explicit variables to provide useful information for constructing CPTs with reduced efforts.

[Insert: **Figure 2** The example of leaky model: (a) The relationship among the real-world domain V^R, reduced model V, and implicit model V_I; (b) The internal structure of a leaky ICI model, where variable Z_L summarises the effect of V_I]

Before applying the leaky-MAX model, there are two assumptions (i.e., assumption of independence of causal influence (ICI)) for all cause-effect relationships involved: (1) each parent node X_i has a probability p_i of being sufficient to produce an impact Z_i on the child node Y in the absence of all other causes; and (2) the ability of each cause being sufficient is independent of the presence of the other causes (Figure 2(b)). Then the CPT, $P(y|X)$, of leaky-MAX model can be obtained through (Diez and Druzdzel, 2006):

$$P(Y \le y|X) = \sum_{z|f_{MAX}(Z)\le y} \prod_{i|X_i \in X} P(z_i|x_i) \sum_{z_l|f_{MAX}(z,z_L)\le y} P(z_L) = \prod_i \left(\sum_{z_i \le y} c_{z_i}^{x_i}\right)\left(\sum_{z_L \le y} c_{z_L}^{L}\right) \quad (4)$$

$$C_y^L = \sum_{z_L \le y} c_{z_L}^L \quad (5)$$

$$C_y^{x_i} = \sum_{z_i \le y} c_{z_i}^{x_i} \quad (6)$$

$$P(Y \le y|X) = C_y^L \cdot \prod_i C_y^{x_i} \quad (7)$$

$$P(y|X) = \begin{cases} P(Y \le y|X) - P(Y \le y-1|X), y \ne y_{min} \\ P(Y \le y|X), y = y_{min} \end{cases} \quad (8)$$

where z_i are the explicit causes and z_L are the inexplicit causes.

Although the leaky-MAX model is applied under the ICI assumptions, it is good enough to provide the approximation of the CPTs compared with the reduction of efforts (Wisse et al., 2008; Diez and Druzdzel, 2006). Zagorecki and Druzdzel (2006) fitted the noisy-MAX model (i.e., the specification of leaky-MAX model) to existing CPTs of three Bayesian networks and found that the model can provide a good fit for as many as 50% of CPTs, which is powerful enough in practice.

In this research, the leaky-MAX model was thus applied to develop the CPTs $P(y|X)$ under the ICI assumptions for reasonable simplification. The identified key construction schedule risks are taken as the variables in Bayesian network model, while three possible states are assigned to each variable, i.e., 'State 1: Better than expected (B)' defining that the condition is better than expected when risk happens, 'State 2: Expected (E)' defining that the condition is exactly as expected when risk happens, and 'State 3: Worse than expected (W)' defining that the condition is better than expected when risk happens. Ordinal comparison among these states can be defined by the influence degree of each state on construction schedule. Assuming that the 'Expected' state has already considered a certain influence degree of each risk on construction schedule, the states are ordinal according to the influence degree, where '$\mathrm{B} < \mathrm{E} < \mathrm{W}$'.

6. Monte Carlo Simulation Driven Risk Inference

The general problem of computing probabilities of interest from a joint probability distribution is probabilistic inference. With Bayesian network model developed, it provides the basis to predict the probability of risk occurrence where the probabilistic inference can be executed dynamically with two facts (Neil et al., 2005): (1) the effects of observations entered into one or more nodes can be propagated throughout the Bayesian network, in any direction, and the marginal distributions of all nodes are updated; and (2) only relevant inferences can be made in the Bayesian network; The Bayesian network uses conditional dependency structure and current knowledge base to determine those inferences that are valid. According to these two facts, the Junction Tree (JT) algorithm has been developed and commonly adopted in the exact inference of multiply connected networks. Details of JT algorithm can be found in the work of Lauritzen and Spiegelhalter (1988).

In order to start this risk inference process with JT algorithm, the Monte Carlo simulation (MCS) is adopted for simulating the occurrence of risk as evidence according to the updated 'risk sate probability boundary'. This 'risk sate probability boundary' can demonstrate the probability of occurrence of different risk states (i.e., state 1, 2 and 3), which is also the marginal probability of each risk state.

The random number within the scale $[0, 1]$ is firstly generated by MCS based on uniform probability distribution, which is the index of determining the state of target risk. The equal chance of getting any stochastic value between 0 and 1 can model the real system more realistically and accurately (Ökmen and Öztaş, 2008).

$$r_i \sim U[0, 1] \tag{9}$$

$$R_i = \begin{cases} B, r_i \in [0, P_{iB}] \\ E, r_i \in (P_{iB}, P_{iB} + P_{iE}] \\ W, r_i \in (P_{iB} + P_{iE}, 1] \end{cases} \tag{10}$$

where the r_i is the random number generated to determine the state of ith risk; P_{iB}, P_{iE} and P_{iW} are the risk state probability boundaries of ith risk.

In this Bayesian Monte Carlo simulation process, states of risks are simulated in a chain to introduce evidence in risk inference (Figure 3). This chain is defined according to the time sequence, where the priority of simulation will be given to the risks being predecessors or risks will be simulated simultaneously if no predecessor existed. It is reasonable to define this simulation sequence according to the time, which have mirrored the sequence of risk occurrence in practice. For instance, in Figure 3, risks 2 and 4 occurred at the time t_1 are simulated firstly according to equation (9-10). After determining the risk state of these risks, JT algorithm is adopted to propagate the probability along the Bayesian network and update all other risks (i.e., Bayesian process in Figure 3). Then the simulation of risks occurred at the time t_2 (e.g., risk 3 in Figure 3) are followed given the evidence of risk states of risks 2 and 4. The JT algorithm is adopted again for probability propagation. The simulation and propagation processes are repeated until the states of all risks are determined, where one iteration of the Bayesian Monte Carlo simulation process is completed. Multiple iterations are usually needed to provide reliable results for risk inference of infrastructure construction schedule. It is usually sufficient to have 1,000~3,000 iterations for providing reliable results with affordable computational cost (Diaz and Hadipriono, 1993). This research thus had 3,000 iterations and selected the same seed value for generating the same specific sequence of random numbers in every experiment, which makes the simulation become reproducible, and is useful for comparing the results derived from different conditions.

[Insert: **Figure 3** Example of one iteration of Bayesian Monte Carlo simulation process]

7. Case Study

7.1 Case background

The main structure for tunnelling of an underground railway project in Greater Bay of China was selected for case study. This railway project was designed as part of a comprehensive transportation hub in this area, while this research focused on the construction of central ventilating shaft of the metro project, which is also the starting point of tunnel boring. It is an underground three-span two-story structure, whose length is 130.0 m from *YDK46+010.8212* to *YDK46+140.8212*, width is 31.8 m, and depth is 20.0 m. The main construction area of it is 8262 m^2 and affiliated construction area is 797.86 m^2.

This project began on 1st November 2017 and was expected to be completed before 20th June 2018. As part of the transportation hub, the construction schedule should be under control to avoid the occurrence of construction delay. It is thus necessary for this project to conduct the construction schedule risk inference in advance. In the past projects, the project management team mainly relied on experience to manage construction schedule risks, which can identify and classify key risks but cannot quantify probabilities of such risks. The developed approach is needed by the project management team for quantitative risk inference.

7.2 Case data collection

Two rounds of focus group discussions have been held in August 2017 and October 2017 separately to collect data for analysis.

In the 1st focus group discussion, the construction schedule risks were firstly provided and verified by five participants from the project management team using the classical experience-based method. Totally 32 construction schedule risks have been identified as the network boundary of risk network (Table 3). The links between risks were then identified and assessed according to the expert experience and opinions using the RSM method. Totally 262 links between risks have been identified, which constructed the risk network $G(32, 262)$ together with 32 risks. The link strength was also provided for each identified link (Table 4).

[Insert: **Table 3** Construction schedule risks of case project]

[Insert: **Table 4** Example of RSM with link strength of case project]

In the 2nd round focus group discussion held after the construction of DAG structure, the DAG structure was verified and canonical parameters for all risks involved in the DAG were collected to develop CPTs (Table 5 and 6).

[Insert: **Table 5** Canonical parameters for root risks of case project]

[Insert: **Table 6** Canonical parameters for Tw-R1]

7.3 Results and analysis

Based on the data collected, the construction schedule risk network can be constructed and visualised as $G(32, 262)$ shown in Figure 4(a), where 32 risk (nodes) in different categories (shapes) and sub-categories (colours) were connected by 262 links (arrows). In order to construct the key risk network, the topological analysis has been conducted based on the metrics defined in Table 2.

Based on the results of topological analysis at node-level, it was observed that the top three risks with high values of nodal metrics (D_d, E, B, S and total brokerage) were highly overlapped and consistent. Totally 7 key risks have thus been identified (Table 7). Meanwhile, according to the results of link-level topological analysis, it was observed that a sharp decline was occurred at 10 in the L-shape curve of link betweenness centrality, where $B(R_i \rightarrow R_j) = 10$ was set as the cut-off point to distinguish key risk links. Totally 20 key risk links have been selected for their values of betweenness centrality (B) were higher than 10 (Table 7).

In order to construct the key risk network with a simple structure but retaining most of essential information, besides key risks and links, other components involving key risks or involved in key risk links are also necessary to be included in the key risk network, where 14 non-key risks involved in the key risk links and 19 non-key risk links between key risks have thus been counted in (Table 7). As shown in Figure 4(b), the key risk network of case project has been constructed as $G\ (21, 39)$, which consisted of 21 risks (nodes) including 7 key risks and 14 non-key risks and 39 links (arrows) including 20 key risk links and 19 non-key risk links.

[Insert: **Figure 4** (a) Construction schedule risk network of case project, $G(32, 262)$; (b) Key construction schedule risk network of case project, $G\ (21, 39)$]

[Insert: **Table 7** Components of key risk network]

After developing the key risk network $G\ (21, 39)$, DFS algorithm was firstly adopted to search for the cycles if existed in the network through transforming the network into a spanning tree. It was observed that five back edges (i.e., links) were existed in the network, including Tw-R2→S7R4, Tw-R1→Tw-R2, Tw-R1→S7R4, S1R6→S0R2, and S7R4→S1R5, indicating that there were cycles existed in the spanning tree. In order to further transform the cycled spanning tree into DAG structure, the A-MWST algorithm was then applied to eliminating these cycles and constructing the DAG structure. According to A-MWST algorithm, four risk links have been eliminated from $G\ (21, 39)$ due to their low weights (i.e., the betweenness centrality of link), including Tw-R1→Tw-R2, S1R5→S7R4, S1R6→S0R2, and S7R4→Tw-R2. Finally, the DAG structure $G(21, 35)$ consisting of 21 risks and 35 risk links has been developed (Figure 5), which however has no cycle existed compared to the key risk network $G\ (21, 39)$.

[Insert: **Figure 5** DAG structure of case project, $G(21, 35)$]

Following the development of DAG structure, the leaky-MAX model was further adopted to generate the CPTs based on the determined canonical parameters (Table 5 and 6) under the ICI assumptions. According to the equations (4-8), the CPTs for risks of the case project can be figured out conveniently (Table 8).

[Insert: **Table 8** Part of CPTs for Tw-R1]

Finally, the MCS-driven risk inference can be conducted to identify key construction schedule risks and predict the probability of risk occurrence. Based on the construction process (time sequence), the simulation sequence was determined as 'S1R3, S4R5, S4R6, S9R3→S4R4, S1R2→S0R2, S1R6→Tw-R2→Tw-R1→Sp-R4, S6R1, S6R2, S7R4→S1R5, S7R3→S3R1, S7R2→Tr-R3→S4R8, S2R1'. After 3,000 iterations of the simulation, the results provided a good estimation of risk occurrence of case project, quantifying the probability of three states for each risk (Figure 6).

According to the simulated probability of 'State 3: Worse than expected', these 21 risks of case project can be classified into three categories, including high-risky, medium-risky and low-risky, which require different risk management strategies from the project management team. The high-risky ones represent the nine risks with probability of state 3 higher than 50% (i.e., $50\% \leq P_w \leq 100\%$) (e.g., Figure 6a-6c), including S7R3 (89.6%), S7R4 (86.9%), S7R2 (84.1%), Tw-R1 (76.7%), Tw-R2 (72.8%), Sp-R4 (69.6%), S1R6 (66.7%), S4R4 (50.0%), S6R2 (50.7%). The management team should focus on these risks to avoid risk occurrence and prepare plans for mitigating the risk impact if happens. The medium-risky ones represent the eight risks with $20\% \leq P_w < 50\%$ (e.g., Figure 6d-6f), including S6R1 (49.1%), S0R2 (44.2%), S3R1 (43.4%), Tr-R3 (37.6%), S2R1 (36.4%), S1R5 (35.2%), S1R2 (27.9%) and S4R6 (20.2%), where the team need to pay attention to these risks after the high-risky ones and also prepare plans for risk mitigation. The low-risky ones represent the four risks with $0 \leq P_w < 20\%$ (e.g., Figure 6g-6i), including S9R3 (0%) and S4R8 (9.3%), S1R3 (13.1%) and S4R5 (19.4%) where it is not necessary for the team to pay much attention to them but have risk mitigation plans prepared.

Based on the results, the risk management and mitigation for these 21 risks were prioritised for decision-makers to draw a risk management benchmark with four different levels (from 0 to 3) and resources input accordingly. Specifically, the level-0 indicates that no risk happens, and the construction progresses as expected. The level-1 indicates that one or more low-risky risks' states are 'Worse than expected', but the construction schedule is just impacted slightly. The level-2 indicate that one or more medium-risky risks' states are 'Worse than expected', and the construction schedule is impacted moderately. The level-3 indicate that one or more high-risky risks' states are 'Worse than expected', and the construction schedule is impacted seriously. This benchmark can help decision-makers understand the risk interdependencies and dynamic nature of risk propagation, and avoid rippled disruption of project construction and delivery.

To verify the results, these probabilities of risks were back to the project team for further discussion, where the five experts participating the focus group discussion before were invited to review the probability of each risk and assess how appropriate these probabilities are based on their rich experience on similar projects. Each expert was asked the same question "How appropriate are these probabilities to be used to predict risk states?" Through using the five-point Likert scale (from 1 = "Not appropriate at all" to 5 = "Very appropriate"), the results showed that all the probabilities (of state 1, 2 and 3) have been scored over 3 averagely, indicating that the simulation results were believed to be appropriate to be used to predict the risk states and occurrence of this case project.

[Insert: **Figure 6** Examples for probability of occurrence of risk states (P_B, P_E, P_W): (a) S7R2, (b) Sp-R4, (c) Tw-R2, (d) S1R5, (e) S6R1, (f) S3R1, (g) S1R3, (h) S4R8, (i) S9R3.]

8. Discussion

This research contributes to the construction schedule risk inference of infrastructures through developing a state-of-the-art solution, hybrid approach, based on Bayesian Monte Carlo simulation. It outperforms other construction schedule risk analysis methods in its reliability, convenience and flexibility to deal with the complex construction schedule risks.

Firstly, this approach provides more insights into the diversity and interdependency of construction schedule risks, generating more reliable results of risk identification and risk inference. The diversity of and interdependencies between construction schedule risks have been considered as main reasons for complexity of infrastructures (Fang et al., 2012; Chu et al., 2003). It would be necessary and useful to address this complexity for providing reliable prediction of risk occurrence (Raz and Michael, 2001). Compared with previous construction risk analysis methods, such as correlated schedule risk analysis model (CSRAM) (Ökmen and Öztaş, 2008) and system dynamics approach (Wang and Yuan, 2016), this approach steps further to not only identify risk interdependencies but also clarify how these interdependencies impact the risk inference. Through network theory-based analysis and algorithms (DFS and A-MWST), this approach can identify diverse risks and interdependencies which are transformed into DAG structure with essential information preserved. Through Bayesian Monte Carlo simulation, the identified interdependencies are adopted to propagate the impact among risks for risk inference.

Secondly, the developed approach is more convenient in data acquisition and processing for risk inference. Although the data of infrastructure construction is becoming richer recently, it is often impossible in practice to define and use a unified classification code for risk identification and data template for Bayesian network development and training (Lee et al., 2009). The expert knowledge driven structure construction method for Bayesian network development is however time-consuming and will inevitably introduce subjective bias (Hu et al., 2013; Luu et al., 2009). For example, Nasir et al. (2003) used to adopt pre-screening, testing and semi-supervised survey to save time in Bayesian network development but they still claimed that the risk relationship identification and quantification (i.e., DAG and CPTs development) were difficult and even impractical, which took 6 weeks with 69 risks. This approach can fully leverage a small dataset for network theory-based analysis and deal with the exponential growth of the number of parameters in CPTs using leaky-MAX model. It only took totally 4 hours to collect and process data for Bayesian network development. With the network theory-based analysis and proposed algorithms (i.e., DFS and A-MWST), only the strength of link S_{ij} is required to generate the key risk network and further the DAG structure. With the leaky-MAX model, only $(mn \cdot m)$ canonical parameters are needed to generate reliable CPTs rather than $(m^n \cdot m)$ conditional probabilities for a node in multi-valued Bayesian network with m possible values and n parent nodes, significantly reducing the data requirement and computational complexity.

Finally, the Bayesian Monte Carlo simulation method provides more flexibility for construction schedule risk inference of infrastructures in practice. At the project planning stage, the risk inference of construction schedule has been often conducted using observations to provide the predicted probability of risk occurrence in a one-shot way. For example, Khodakarami et al. (2007) developed a Bayesian network solution based on CPM to predict the construction schedule under uncertainty and conducted the scenario analysis for probability of resources level based on hypothetic observations. Nasir et al. (2003) and Luu et al. (2009) both developed a Bayesian network model for quantifying the probability of risk occurrence given the CPTs and hypothetic observations. Compared to previous research, the simulated data are introduced as soft evidence in this research to trigger risk inference at the project planning stage, which enables variables to be simulated individually or simultaneously according to CPTs rather than hypotheses if the observation data are not available. This developed approach is however compatible with the observation data and other datasets (e.g., real-time sensor/visual data informing the risk states) when they are available to trigger risk inference and update the risk network in near real-time. Moreover, by adopting the same seed value, the simulation process can be reproducible and useful for comparing the results derived from different conditions.

9. Conclusions

The construction of complex infrastructures has provided rich data for construction schedule risk management which however has still been challenged by construction delay with enormous losses. This research developed a novel Bayesian Monte Carlo simulation driven approach to predict and quantify the probability of construction schedule risk occurrence. The developed approach efficiently addresses two problems through fully leveraging the dataset from construction schedule risks: (1) the lack of data template for Bayesian network development in practice, and (2) the lack of observation data for triggering risk inference in project planning. It addressed the first problem through developing a data transformation approach based on DFS and A-MWST algorithms and leaky-MAX model and converting a risk network into Bayesian network. It then resolved the second problem by designing a Bayesian Monte Carlo simulation driven risk inference method. The developed approach has been validated by a case study, where the results enabled the project team to prepare look-ahead risk mitigation strategies.

This research makes theoretical contributions to the body of knowledge through analysing the diversity and interdependency of construction schedule risks from a network perspective (i.e., network theory and Bayesian network). This network-oriented approach handles construction schedule risks of complex projects (e.g., infrastructures) in a system engineering approach to avoid rippled disruption of project delivery (Whyte, 2016). The new view and rethinking to manage such risks of complex projects provide deeper insights into the complexity and uncertainty of construction of infrastructures. The practical contributions of this research mainly include: (1) a novel approach developed to construct Bayesian network and conduct Bayesian Monte Carlo simulation for construction schedule risk inference. Compared to traditional methods, it can outperform in reliability of results, convenience of data acquisition and processing and flexibility in practice; and (2) the findings from case study. The constructed key risk network $G\ (21, 39)$ can help decision-makers to identify construction schedule risks and understand risk propagation of infrastructures. The prediction of probabilities of risk occurrence can further help decision-makers prioritise the construction schedule risks and make a risk management benchmark. These findings however are not limited to the case project but can be applied to other similar projects. They can also assist researchers and practitioners in understanding the practicability of this hybrid approach.

There are also some limitations needed to be addressed in the future research. Firstly, the data transformation approach is semi-automatic which requires many efforts to transform key risk network to DAG structure. Secondly, this research only considers the states of risks in time sequence while ignores the dynamic nature of risk states in the longitude. Thirdly, more case studies and more objective assessment methods are needed to validate this developed approach. Future research should be conducted to address these limitations through: (1) developing an automatic approach and system to integrate network theory-based analysis and Bayesian network; (2) developing a dynamic Bayesian network based approach for analysing risk dynamics in the longitude; (3) conducting more case studies of infrastructures both in China and overseas to validate and further improve this developed approach; and (4) integrating the probabilities of risk occurrence with impacts on construction schedule to predict the construction schedule under uncertainty and also to verify this developed approach quantitatively.

Data Availability Statement

All data (network theory-based analysis, canonical parameters and CPTs) and codes (Bayesian Monte Carlo simulation) that support the findings of this study are available from the corresponding author upon reasonable request.

References

Abotaleb, I. S., El-adaway, I. H. (2018). "Managing construction projects through dynamic modeling: Reviewing the existing body of knowledge and deriving future research directions." *Journal of Management in Engineering*, 34(6), 04018033, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000633.

Alizadehsalehi, S., Hadavi, A., Huang, J. C. (2020). “From BIM to extended reality in AEC industry.” *Automation in Construction*, 116, 103254, https://doi.org/10.1016/j.autcon.2020.103254.

Al-Momani, A. H. (2000). “Construction delay: a quantitative analysis.” *International journal of project management*, 18(1), 51-59, https://doi.org/10.1016/S0263-7863(98)00060-X

Baloyi, L. and Bekker, M. (2011). “Causes of construction cost and time overruns: The 2010 FIFA World Cup stadia in South Africa.” *Acta Structilia*, 18(1), 51-67, eISSN: 2415-0487.

Bryman, A. (2004). *Social Research Methods*, 2nd edition. Oxford University Press, Oxford.

Chen, L. (2019). “An integrated construction schedule risk analysis approach for infrastructure projects under uncertainty.” HKU Theses Online.

Chen, L., Lu, Q., and Zhao, X. (2020). “Rethinking the Construction Schedule Risk of Infrastructure Projects Based on Dialectical Systems and Network Theory.” *Journal of Management in Engineering*, 36(5), 04020066, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000829

Chen, L., Whyte, J. (2020). “Analysing Interdependencies of Complex Engineering Systems using a Digital Twin-Driven Design Structure Matrix.” *ASCE Construction Research Congress 2020*, Tempe, Arizona, USA.

Chu, D., Strand, R., Fjelland, R. (2003). “Theories of complexity.” *Complexity*, 8(3), 19-30, https://doi.org/10.1002/cplx.10059.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C. (2001). *Introduction to Algorithms*, 2nd edition. MIT Press and McGraw-Hill.

Diaz, C. F., Hadipriono, F. C. (1993). “Nondeterministic networking methods.” *Journal of Construction Engineering and Management*, 119(1), 40-57, https://doi.org/10.1061/(ASCE)0733-9364(1993)119:1(40).

Diez, F. J. (1993). “Parameter adjustment in Bayes networks. The generalized noisy OR–gate.” *Proceedings of 9th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, pp. 99-105.

Diez, F. J., Druzdzel, M. J. (2006). “Canonical probabilistic models for knowledge engineering.” Technical Report CISIAD-06-01, UNED, Madrid, Spain.

Diez, F. J., Galán, S. F. (2003). “Efficient computation for the noisy MAX.” *International journal of intelligent systems*, 18(2), 165-177, https://doi.org/10.1002/int.10080.

Dogan, S. Z., Arditi, D., Gunhan, S., Erbasaranoglu, B. (2013). “Assessing coordination performance based on centrality in an e-mail communication network.” *Journal of Management in Engineering*, 31(3), 04014047, https://doi.org/10.1061/(ASCE)ME.1943-

5479.0000255.

Ellis, R. D., Thomas, H. R. (2003). "The root causes of delays in highway construction." *82nd Annual meeting of the transportation research board*, Transportation Research Board, National Research Council, Washington, DC.

El-Sayegh, S. M., Mansour, M. H. (2015). "Risk assessment and allocation in highway construction projects in the UAE." *Journal of Management in Engineering*, 31(6), 04015004, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000365.

Fang, C., Marle, F. (2012). "A simulation-based risk network model for decision support in project risk management." Decision Support System, 52(3), 635–644, https://doi.org/10.1016/j.dss.2011.10.021.

Fang, C., Marle, F., Zio, E., Bocquet, J. C. (2012). "Network theory-based analysis of risk interactions in large engineering projects." Reliability Engineering & System Safety, 106, 1-10, https://doi.org/10.1016/j.ress.2012.04.005.

Fiori, C., Kovaka, M. (2005). "Defining megaprojects: Learning from construction at the edge of experience." *Construction Research Congress 2005: Broadening Perspectives*, ASCE, pp. 1-10.

Flyvbjerg, B., Skamris Holm, M. K., Buhl, S. L. (2004). "What causes cost overrun in transport infrastructure projects?" *Transport reviews*, 24(1), 3-18, https://doi.org/10.1080/0144164032000080494a.

Flyvbjerg, B., Skamris Holm, M. K., Buhl, S. L. (2005). "How (in) accurate are demand forecasts in public works projects: The case of transportation." *Journal of the American planning association*, 71(2), 131-146, https://doi.org/10.1080/01944360508976688.

Fu, M., Liu, R. (2018). "The application of virtual reality and augmented reality in dealing with project schedule risks." *Proceedings of Construction Research Congress 2018*, 429–438, https://doi.org/10.1061/9780784481264.042.

Furht, B. (Ed.). (2010). *Handbook of social network technologies and applications*. Springer Science & Business Media.

Goodrich, M. T., Tamassia, R. (2006). *Algorithm design: foundation, analysis and internet examples*. John Wiley & Sons.

Han, S. H., Yun, S., Kim, H., Kwak, Y. H., Park, H. K., Lee, S. H. (2009). "Analyzing schedule delay of mega project: Lessons learned from Korea Train Express." *IEEE Transactions on Engineering Management*, 56, 243-256, 10.1109/TEM.2009.2016042.

Hu, Y., Zhang, X., Ngai, E. W. T., Cai, R., Liu, M. (2013). "Software project risk analysis using Bayesian networks with causality constraints." *Decision Support Systems*, 56, 439-449, https://doi.org/10.1016/j.dss.2012.11.001.

Ke, Y., Wang, S., Chan, A. P. (2010). “Risk allocation in public-private partnership infrastructure projects: comparative study.” *Journal of Infrastructure Systems*, 16(4), 343-351, https://doi.org/10.1061/(ASCE)IS.1943-555X.0000030.

Khodakarami, V., Fenton, N., Neil, M. (2007). “Project Scheduling: Improved approach to incorporate uncertainty using Bayesian Networks.” Project Management Journal, 38(2), 39-49, https://doi.org/10.1177/875697280703800205.

Kjaerulff, U. B. (2008). *BNs and influence diagrams: A guide to construction and analysis*. Springer, New York.

Larsen, J. K., Shen, G. Q., Lindhard, S. M., Brunoe, T. D. (2015). “Factors affecting schedule delay, cost overrun, and quality level in public construction projects.” *Journal of Management in Engineering*, 32(1), 04015032, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000391.

Lauritzen, S. L., Spiegelhalter, D. J. (1988). “Local computations with probabilities on graphical structures and their application to expert systems.” *Journal of the Royal Statistical Society, Series B (Methodological)*, 157-224, https://doi.org/10.1111/j.2517-6161.1988.tb01721.x.

Lee, E., Park, Y., Shin, J. G. (2009). “Large engineering project risk management using a Bayesian belief network.” *Expert Systems with Applications*, 36(3), 5880-5887, https://doi.org/10.1016/j.eswa.2008.07.057.

Lin, S., Chen, X., Wang, Q. (2018). “Fault diagnosis model based on Bayesian network considering information uncertainty and its application in traction power supply system.” *IEEJ Transactions on Electrical and Electronic Engineering*, 13(5), 671-680, https://doi.org/10.1002/tee.22616.

Liu, D., Xuan, P., Li, S., Huang, P. (2014). “Schedule Risk Analysis for TBM Tunneling Based on Adaptive CYCLONE Simulation in a Geologic Uncertainty-Aware Context.” *Journal of Computing in Civil Engineering*, 29(6), 04014103, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000441.

Liu, H., Al-Hussein, M., Lu, M. (2015). “BIM-based integrated approach for detailed construction scheduling under resource constraints.” *Automation in Construction*, 53, 29-43, https://doi.org/10.1016/j.autcon.2015.03.008.

Lo, T. Y., Fung, I. W., Tung, K. C. (2006). “Construction delays in Hong Kong civil engineering projects.” *Journal of Construction Engineering and Management*, 132(6), 636-649, https://doi.org/10.1061/(ASCE)0733-9364(2006)132:6(636).

Luu, V. T., Kim, S. Y., Van Tuan, N., Ogunlana, S. O. (2009). “Quantifying schedule risk in construction projects using Bayesian belief networks.” *International Journal of Project

Management, 27(1), 39-50, https://doi.org/10.1016/j.ijproman.2008.03.003.

McCabe, B., AbouRizk, S. M., Goebel, R. (1998). "Belief networks for construction performance diagnostics." *Journal of Computing in Civil Engineering*, 12(2), 93-100, https://doi.org/10.1061/(ASCE)0887-3801(1998)12:2(93).

Mok, K. Y., Shen, G. Q., Yang, R. J., Li, C. Z. (2017). "Investigating key challenges in major public engineering projects by a network-theory based analysis of stakeholder concerns: A case study." *International Journal of Project Management*, 35(1), 78-94, https://doi.org/10.1016/j.ijproman.2016.10.017.

Moon, H., Kim, H., Kamat, V. R., Kang, L. (2015). "BIM-based construction scheduling method using optimization theory for reducing activity overlaps." Journal of Computing in Civil Engineering, 29(3), 04014048. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000342.

Nasir, D., McCabe, B., Hartono, L. (2003). "Evaluating risk in construction–schedule model (ERIC-S): construction schedule risk model." *Journal of construction engineering and management*, 129(5), 518-527, https://doi.org/10.1061/(ASCE)0733-9364(2003)129:5(518).

Neil, M., Fenton, N., Tailor, M. (2005). "Using Bayesian networks to model expected and unexpected operational losses." *Risk Analysis*, 25(4), 963-972, https://doi.org/10.1111/j.1539-6924.2005.00641.x.

Ökmen, Ö., Öztaş, A. (2008). "Construction project network evaluation with correlated schedule risk analysis model." *Journal of Construction Engineering and Management*, 134(1), 49-63, https://doi.org/10.1061/(ASCE)0733-9364(2008)134:1(49).

Oniśko, A., Druzdzel, M. J., Wasyluk, H. (2001). "Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates." *International Journal of Approximate Reasoning*, 27(2), 165-182, https://doi.org/10.1016/S0888-613X(01)00039-1.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA.

Perrenoud, A. J., Smithwick, J. B., Hurtado, K. C., Sullivan, K. T. (2015). "Project risk distribution during the construction phase of small building projects." *Journal of Management in Engineering*, 32(3), 04015050, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000417.

Raz, T., Michael, E. (2001). "Use and benefits of tools for project risk management." *International journal of project management*, 19(1), 9-17, https://doi.org/10.1016/S0263-7863(99)00036-8.

Sambasivan, M., Soon, Y. W. (2007). "Causes and effects of delays in Malaysian construction industry." *International Journal of Project Management*, 25(5), 517-526, https://doi.org/10.1016/j.ijproman.2006.11.007.

Sami Ur Rehman, M., Thaheem, M. J., Nasir, A. R., Khan, K. I. A. (2020). "Project schedule risk management through building information modelling." International Journal of Construction Management, 1-11, https://doi.org/10.1080/15623599.2020.1728606.

Tallgren, M. V., Roupé, M., Johansson, M., Bosch-Sijtsema, P. (2020). "BIM-tool development enhancing collaborative scheduling for pre-construction." *Journal of Information Technology in Construction (ITcon)*, 25(22), 374-397, https://doi.org/10.36680/j.itcon.2020.022.

Toor, S. U. R., Ogunlana, S. O. (2008). "Problems causing delays in major construction projects in Thailand." Construction management and economics, 26(4), 395-408.

UN-Habitat (2016). "World Cities Report 2016-Urbanization and development: Emerging future." < http://wcr.unhabitat.org/main-report/> (Sept. 16, 2019).

Valentin, V., Naderpajouh, N., Abraham, D. M. (2018). "Integrating the input of stakeholders in infrastructure risk assessment." Journal of Management in Engineering, 34(6), 04018042, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000638.

Van Der Gaag, L. C., Renooij, S., Witteman, C., Aleman, B., Taal, B. (1999). "How to elicit many probabilities." *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, pp. 647-654.

Wang, J., Yuan, H. (2016). "System dynamics approach for investigating the risk effects on schedule delay in infrastructure projects." *Journal of Management in Engineering*, 33(1), 04016029, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000472.

Wang, L., Zhang, X. (2018). "Bayesian analytics for estimating risk probability in PPP waste-to-energy projects." Journal of Management in Engineering, 34(6), 04018047, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000658.

Whyte, J., Stasis, A., Lindkvist, C. (2016). "Managing change in the delivery of complex projects: Configuration management, asset information and 'big data'." *International Journal of Project Management*, 34(2), 339-351, https://doi.org/10.1016/j.ijproman.2015.02.006.

Wisse, B. W., van Gosliga, S. P., van Elst, N. P., Barros, A. I. (2008). "Relieving the elicitation burden of Bayesian Belief Networks." *Proceedings of the 6th Bayesian Modelling Applications Workshop 2008-How biased are our numbers?*, Helsinki, Finland.

Xie, J., Thomas Ng, S. (2013). "Multi-objective Bayesian network model for public-private partnership decision support." *Journal of Construction Engineering and Management*, 139(9), 1069-1081, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000695.

Yang, R. J., Zou, P. X. (2014). “Stakeholder-associated risks and their interactions in complex green building projects: A social network model.” *Building and Environment*, 73, 208-222, https://doi.org/10.1016/j.buildenv.2013.12.014.

Zagorecki, A., Druzdzel, M. J. (2013). “Knowledge engineering for Bayesian networks: How common are noisy-MAX distributions in practice*?” IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(1), 186-195, 10.1109/TSMCA.2012.2189880.

Zhai, L., Xin, Y., Cheng, C. (2009). “Understanding the value of project management from a stakeholder's perspective: case study of mega-project management.” *Project Management Journal*, 40, 99-109, https://doi.org/10.1002/pmj.20099.