Reliability Engineering and System Safety 186 (2019) 194–208

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Combining association rules mining with complex networks to monitor coupled risks

Zhou Ying[a], Li Chenshuang[a], Ding Lieyun[a], Przemyslaw Sekula[b], Peter E.D. Love[c], Zhou Cheng[a,*]

[a] School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
[b] Department of Civil & Environmental Engineering, University of Maryland, College Park, MD 20742-3021, USA
[c] School of Civil and Mechanical Engineering, Curtin University, GPO Box U1987, Perth, WA 6845, Australia

ARTICLE INFO

Keywords:
Association rule mining
Deep-pit foundations
Complex network
Risk

ABSTRACT

Due to geotechnical uncertainties, existing underground infrastructure, the construction of deep-pit foundations in dense urban areas is particularly challenging as there is a propensity for building and structural settlement to occur. Recognizing the need to proactively manage safety risks during construction, a new risk analysis approach that combines complex networks and association rules mining (ARM) is proposed. An improved *Apriori* algorithm is developed to unearth abnormal monitoring types. Then, complex network theory is introduced to examine the characteristics of the coupled relationships existing between different types of abnormal monitoring types. This research identifies and examines complex network measures to understand the topology of settlement networks. It is revealed that settlement networks confirm to both scale-free and small-word properties indicating that risks are not random events. This new approach of combining ARM with complex network is applied to examine deep foundation pits that are constructed for a subway project in Wuhan, China. It is demonstrated that proposed approach can successfully reveal the association rules between safety risk monitoring types and the coupling of risks. Preventative actions can therefore be undertaken in advance to mitigate against potential risks that are identified from the abnormal monitoring combinations.

1. Introduction

Subway transportation plays an integral part of metropolitan infrastructure systems as it provides the means to efficiently convey passengers from one location to another. With increasing levels of urbanization, pressure is placed on existing subway networks, which influences the demand for new lines [1]. The formation of deep foundation pits, for example, form an integral part of the construction process for new subway systems. Managing and controlling risk and ensuring peoples safety on a construction site and the general public is a pervasive challenge. In particular, the complexity of geological conditions, the presence of underground utilities and services and the potential for building settlement and structural collapse are significant safety risks that need to be considered [2,3].

It is therefore necessary to take precautionary measures to avoid accidents from occurring by determining the uncertainties and risks that can materialize [4,5]. To address the challenges associated with subway construction, a number of studies have identified that equipment should be installed to monitor the health of structures surrounding the deep-foundation pit and within its workspace [6]. Through the effective real-time dynamic monitoring of deep-pit foundations and the analysis of the data that is obtained, additional information for risk analysis can be acquired to improve safety [7]. Risks are interdependent and correlation analysis is needed to identify those that are coupled. However, there are a limited number of risk management methods that have been developed that can address this issue [8]. To better understand the coupling of risks that can materialize during the construction of deep-pit foundations and improve safety, a new risk analysis method that combines association rule mining (ARM) and complex network is proposed and demonstrated using a real-life project.

2. Literature review

The construction of underground subway systems is a complex and high-risk environment due to issues such as ground subsidence, foundation pit collapse, tunnel destruction, and damage to surrounding buildings. To mitigate such risks numerous, several studies have

* Corresponding author.
E-mail addresses: ying_zhou@hust.edu.cn (Y. Zhou), dly@hust.edu.cn (L. Ding), psekula@umd.edu (P. Sekula), p.love@curtin.edu.au (P.E.D. Love), chengzhou@hust.edu.cn (C. Zhou).

https://doi.org/10.1016/j.ress.2019.02.013
Received 29 June 2018; Received in revised form 29 January 2019; Accepted 12 February 2019
Available online 13 February 2019
0951-8320/

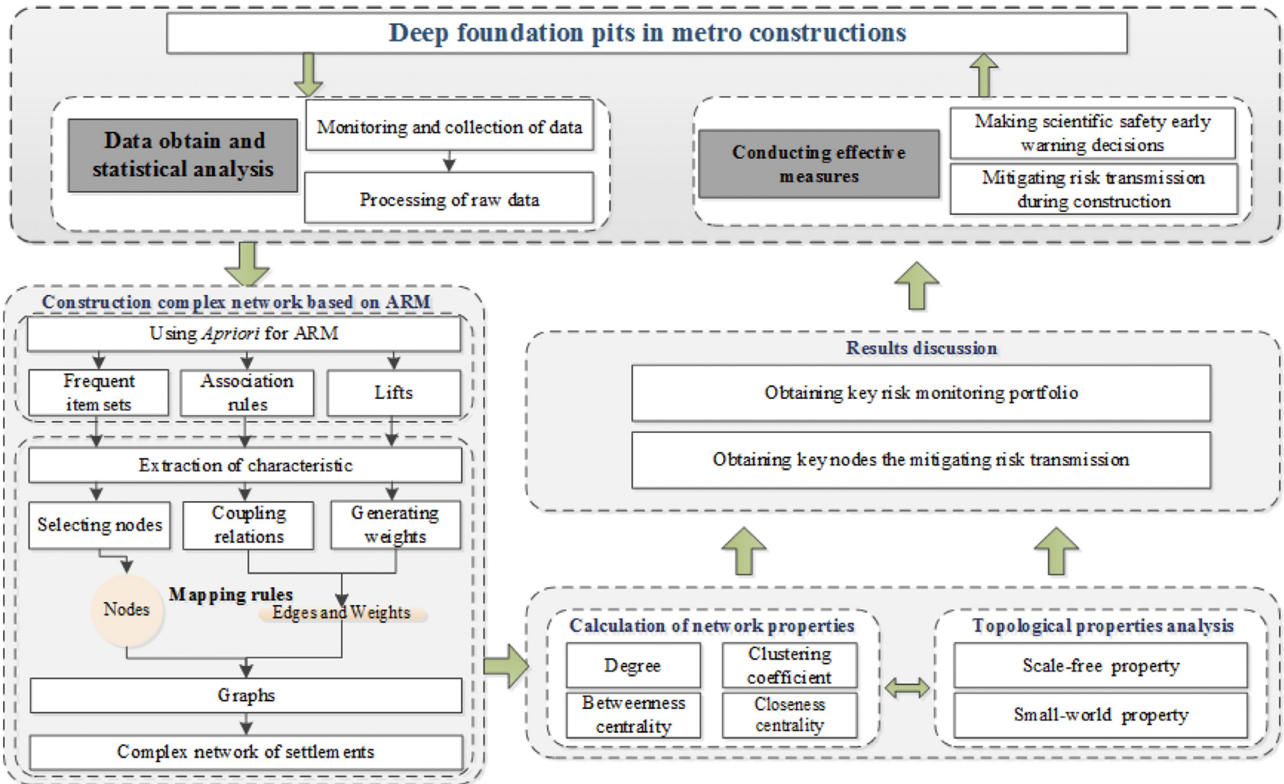


Fig. 1. The process of using complex network based on ARM for risks analysis.

advocated the need for the incorporation of risk-based analysis into safety management practice [9–11].

2.1. Risks analysis of deep foundations pits

Traditional risk analysis methods such as accident and fault trees have been extensively used to assess risks in deep foundation pits [11,12]. Commonly used risk analysis methods in construction include support vector machine (SVM) [13–15], Bayesian network (BN) [16,17], and random forest (RF) [18,19]. While such methods have demonstrated to be effective they also have a number of limitations. For example, SVM can be used to model small sample sizes where data is unavailable [23], but is unable to process those of a large-scale and solve multi-classification problems [20,21]. BNs can be used to construct dynamic models from data and/or expert opinion and employed for a wide range of tasks [5,11,20]. However, BNs are unable to solve the problem of coupling relationship that typically exist between risk factors [22,23]. The RF is confronted with similar issues as SVM as it is also used for small of samples data [19]. However, when RF is used for classification or regression problems, it is prone to overfitting [24,25]. Therefore, SVM, RF and BN methods are unable to solve the complex problem of coupling relationship between risk factors [21–25].

Numerical simulation and computer monitoring technology have also been applied to monitor and obtain data about the deformation of foundation pits. Common monitoring data analysis methods can be divided into two categories that focus on: (1) large volumes of statistical data [26,27]; and (2) the use of system science methods [28–30]. In addition, risk factors are interrelated with one another [31]. Thus, being able to determine the risk factors that are coupled with one another can enable strategies to be put in places to control accidents [32,33]. Within construction, several risk coupling models have been proposed, which include the: (1) Boston Matrix coupling model [34]; (2) coupled fault tree [35]; (3) Work Breakdown Structure-RBS coupled matrix analysis method [36]; (4) interaction matrix [37]; and (5) N–K model [38]. Yet these methods have several limitations when used in risk coupling analysis as they: (1) consider risks to be independent rather than interdependent; and (2) primarily based subjective ranking of risk events; and (3) ignore patterns and causal structures that may prevail.

2.2. Association rule mining

Association rule mining (ARM) is a procedure that aims to determine frequent patterns, correlations, associations, or causal structures from data sets found in various kinds of databases and repositories. To describe and analyze unknown relationships in data, ARM has been applied to assess safety risks [39]. Essentially, ARM is a simple and practical data mining technique that can be used to identify frequent itemsets between uncertain factors and generate strong association rules from a large dataset, particularly auxiliary data from engineering operations. ARM is able to unearth frequent itemsets and association rules between different abnormal situations that have been monitored [39].

Data mining technology based on association rules has been applied to several areas associated with the construction of deep-pit foundations. For example, Sousa examined monitoring data to reveal the maximum deformation rate for a deep-pit [40]. Similarly, Zhou performed a numerical simulation of dewatering deep foundation pits, which assisted risk assessments to be identified for land subsidence and subsequently controlled [41]. Contrastingly, Li and Chen applied the association rule algorithm to analyze the collapse of subway station foundation pits and determine the frequent itemsets of risk between "human-machine-environment-management" [39].

2.3. Complex network

Complex network (CN) can be used to effectively analyze the coupled relationship that exists between risk factors [42–44]. It has been widely used in to examine urban rail transit for network reliability, route analysis and passenger flow assignment [45,46]. In the context of subway systems, CN it has been used to examine the safety and reliability of safety systems particularly the coupling of risks associated with workers, equipment, management and environment [38].

3. Methodology

The research aims to address two issues: (1) the development of a methodology for establishing CN of deep foundations pits based on ARM; and (2) to scientifically analyze the risks of deep foundation pits based on CN theory. The process of developing a risk-coupling model based on monitoring data is presented in Fig. 1. Monitoring data for deep foundation pits needs to be collected and processed. The CN that is established is based on the association rules generated by the *Apriori* algorithm of ARM. Then, the CN based risk analysis for deep foundations pits is conducted. Effective measures to optimize of deep foundation pits are also presented.

3.1. Association rule mining

ARM aims to discover strong rules from databases using some measures of interestingness [47]. To define and select rules from a set of possibilities, ARM requires constraints on various measures of significance to be identified, which can take the following forms:

Definition 1. (Transaction): Let t_i be a collection of all the items that a customer takes in a transaction. The elements $i_k (k = 1, 2, \cdots n)$ that make up t_i are called *Item*. Let the set of all items in D be $I = \{i_1, i_2, \cdots, i_k\}$.

Definition 2. (Association rule): If both A and B are the itemsets of D, and $A \cap B = \varnothing$, the $A \Rightarrow B$ logical implication is called an association rule.

Definition 3. (Support): The association rule $A \Rightarrow B$'s support is the percentage of all transaction items in the transaction that contain both items set A and item B. That is, the support of the items $A \cup B$. It indicates the frequency at which the association rule $A \Rightarrow B$ appears.

$$Support(A \Rightarrow B) = \frac{|\{T: A \cup B \subseteq T, T \subseteq D\}|}{|D|}$$

Definition 4. (Confidence): The confidence of the association rule $A \Rightarrow B$ is the percentage of the transaction that contains both the item A and B in all transactions containing the item A, which indicates the intensity of the association rule $A \Rightarrow B$. The calculated expression is as follows:

$$Confidence(A \Rightarrow B) = \frac{|\{T: A \cup B \subseteq T, T \subseteq D\}|}{|\{T: A \subseteq T, T \subseteq D\}|} = \frac{Support(A \Rightarrow B)}{Support(A)}$$

Definition 5. (Lift): The lift of the association rule $A \Rightarrow B$ is the percentage of the transaction that the observed support to that expected if A and B were independent, which indicates the effectiveness and importance of the association rule $A \Rightarrow B$. Lift > 1 means that the B is likely to appear with A, which means these rules potentially useful for predicting the consequent in future data sets. The calculated expression is as follows:

$$Lift(A \Rightarrow B) = \frac{Supp(A \cup B)}{Supp(A) \times Supp(B)}$$

In discovering meaningful association rules, two thresholds are required [48]:

1. *Minimum support* (*minsup*): The minimum support that the association rules must statistically satisfy; and
2. *Minimum confidence* (*minconf*). The minimum confidence that the association rule must satisfy and also be reliable [49].

Support and confidence are ideal quality measures in ARM. Consequently, they are proposals regularly used to determine minimum thresholds. Exceeding quality thresholds, however, does not guarantee that the rules are relevant, and therefore alternative quality measures may have to be considered. The assessment of association rules based on support and confidence as measures of importance and accuracy has several drawbacks. In addressing this issue, a *lift* can be used, which is ratio of the confidence of the rule and its expected confidence. It refers to dependencies between the antecedent and consequent of an association rule. Therefore, the thresholds *lift* considers both the confidence of the rule and the overall data set. Hence, it can help to filter and obtain effective results [50].

Algorithm. (*Apriori*): In all association rule algorithms, *Apriori* is one of the most influential for mining frequent itemsets with Boolean rules [51,52]. It has been used to mine the frequent itemsets of safety monitoring types. The algorithms process contains four steps [53]:

1. Set minimum confidence and support according to user requirements.
2. Scan the database D once; calculate the support of all individual items; generate frequent 1-itemsets L_1.
3. Use L_1 to find frequent 2-itemsets L_2; this continues until frequent itemsets are unable to be extended and the algorithm stops. In the kth cycle, the set C_k of the candidate k-item sets is generated first, and then the database is scanned to obtain L_k. Put together all the frequent itemsets.
4. On the basis of the frequent itemsets, strong association rules emerge that satisfy the minimum confidence.

3.2. Establishment of complex network

A network can be stored as a two-dimensional matrix $A_{n\times n} = \{a_{ij}\}$. This is commonly referred to as an adjacency matrix, where n is the total number of nodes in a network. If an edge exists between nodes v_i and v_j, then $a_{ij} = 1$; otherwise, $a_{ij} = 0$. Networks without any self-loops are denoted as $a_{ii} = 0$. In weighted networks, edges are always associated with weights that differentiate them in terms of their strength, intensity, or capacity. The risk factors of monitoring data have different influence on the process of deep foundations pits. As for many other CNs, knowing details of the information flow (coupling risk) is a crucial factor for a network. To accommodate the risks of deep foundations pits, a weighted CN is constructed using ARM. Based on the results of the ARM, frequent itemsets and association rules for deep foundations settlement can be obtained. Then, the frequent itemsets and association rules can be mapped into CNs for analysis. For weighted networks, these weights are defined by a matrix (A^w), in which each element (w_{ij}) represents the *lift* of association rule i to j. There are three steps to map the frequent itemsets and association rules map into CNs:

Step 1: Each monitoring type in the above association rule represents a node of the network.
Step 2: Each association rule includes two nodes to form the edge of the network.
Step 3: The lift of the association rule is chosen to be the weight of the edge. A lift > 1 means that the rule is valid and potentially useful for predicting the consequent in future data sets.

Fig. 2 provides an example of the mapping association rules for a weighted and undirected network. For instance, {F ⇒ C} and its *lift* are shown at the top of the table in Fig. 2. This association rule indicates that there is effective relationship between F and C and the *lift* measures the degree to which those two monitoring risk factors are dependent on one another. Therefore, F and C can be regarded as two nodes and there is an edge (F, C). The nodes and relationship between them can be seen in the weighted matrix and CN. The dots of the weighted matrix are

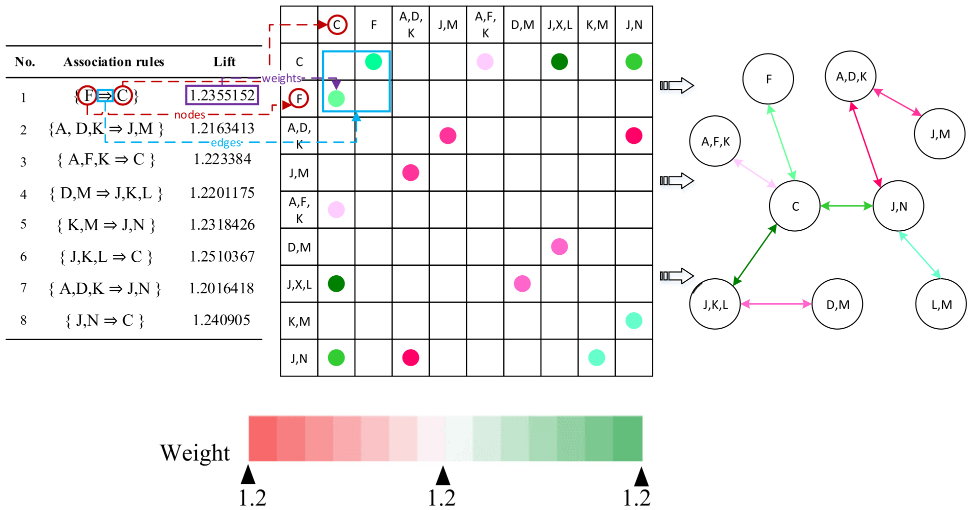


Fig. 2. Example of mapping association rules into a weighted network. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

scaled to weights of the edge of CNs, as denoted in Fig. 2. In addition, the dots of the weight matrix and edges of the CN in Fig. 2 vary in color as their *lifts* are different. The other association rules are the same and the final matrix and network are represented in Fig. 2.

3.3. Complex network analysis

Each class of network presents specific topological features that characterizes its connectivity, interactions and dynamic processes [55]. The topological properties of the constructed network were adopted to measure its main characteristics monitoring points. Analyzing the topology of risks within a network is important as it:

- enables the identification of the most efficient way of engineering the network's structure. and
- provides a systematic way of identifying the critical nodes and their dynamic interdependencies in the network.

3.3.1. Network parameters

3.3.1.1. Degree and degree distribution. They are the common measurements to analyze the local properties (considering only the direct neighborhood of a node) of CNSs. The complementary cumulative distribution function can be expressed as:

$$p(k_i > X) = \sum_{i}^{n} p(k_i)$$

3.3.1.2. Clustering coefficient. The clustering coefficient is an intermediate property (considering the individual neighborhoods of within a node) and measures the aggregation degree between network nodes. The clustering coefficient is defined as:

$$C = 2k_i'/k_i(k_i - 1)$$

where, the degree of node i and k_i' denotes the number of existing edges among its neighbors.

3.3.1.3. Average shortest path length. The average shortest path length is a global property (considering all nodes) of CNs. It quantifies the level of segregation throughout the network, and is defined as:

$$L = \frac{1}{n(n-1)} \sum_{i<j} l_{i,j}$$

3.3.1.4. Betweenness centrality. Betweenness centrality represents the ability of a node to provide medium bridge [54]. It means the importance of individual nodes for the transport of information or matter, assuming that both typically travel through the network on shortest paths. The betweenness centrality of node *i* can be defined as [45]:

$$B_i = \sum_{i \neq s \neq t} l_{st}^i / l_{st}$$

where l_{st} is the number of the shortest paths between node s and t and l_{st}^i is the number of the paths which pass through node i among all the shortest paths between node s and t.

Table 1 summarizes the key metrics used to characterize a network and their important roles in a risk network

3.3.2. Network classification

Most network studies rely on measures that can characterize relevant topological features to identify the unifying principles and statistical properties commonly found in empirical networks. Based on their topological features, the characteristics of different types of networks are summarized in Table 2.There are two kinds of networks that deserve special attention:

- *Scale-free network:* The degree distribution follows a power law. In a CN, if the degree distribution $p(k)$ is followed by a power law distribution, the corresponding complex network is scale free, which indicates that there is a high probability that only a limited number of well-connected nodes known as the hubs of the network will manifest strong impacts on the other nodes; and
- *Small world network:* The degree distribution is akin to that of a random network. This network, however, exhibits shorter average path length and higher clustering coefficient. In the processing of network growth, newly added nodes have a preferential attachment instead of random attachment in small world network.

4. Case study

A case study example is used to demonstrate the proposed new risk analysis approach. The selected case is the Wuhan Rail Transit Network Subway Line 6 at the Li Miao-lu station. The station is located in Jiangan District of Wuhan City. The station is underground and has a two-tier and two-span structure. The length of the station's main pit is approximately 293 m and its width ranges from 20.9 m to 24.74 m. The deep foundation pits excavation depth of ranges from 16.99 to 18.94 m.

Table 1
Property and statistics of a complex network.

Category	Network property	Symbol or equation	Statistics of network property
Local	Node degree	$k_i = \sum_{j=1}^{v} a_{ij}$	Number of edges incident with nodes
	Average degree	$\langle k \rangle = \frac{1}{n}\sum_{i=1}^{n} k_i$	The average number of neighbors a node has in the network
	Maximum degree	k_{max}	Maximum value of degree
	Degree standard deviation	k_σ	Standard deviation value of degree
	Degree distribution	$p(k) = \frac{n_k}{n}$	The probability of a node with degree
	Cumulative distribution of degree	$p(k_i > X) = \sum_i^n p(k_i)$	The degree distribution is important for studying both real-world networks, and theoretical random networks which have very different degree distributions
Intermediate	Clustering coefficient	$C = 2k'_i/k_i(k_i - 1)$	Measures aggregation degree between network nodes and represents the network's transitivity
	Average clustering coefficient	$\langle C \rangle = \frac{1}{n}\sum_{v_i \varepsilon V} C_i$	Average value of clustering coefficient
	Clustering coefficient Standard deviation	$C(k)_\sigma$	Standard deviation value of clustering coefficient
Global	average path length	$L = \frac{1}{n(n-1)}\sum_{i<j} l_{i,j}$	Quantify the level of integration/segregation throughout the newtwork
	Closeness centrality	$C_c(i) = \frac{n-1}{\sum_{v_j \varepsilon n, i \neq j} d_{ij}}$	Measures the extent to which a node is close to all the other nodes along the shortest path and reflects its accessibility in a given network
	Betweenness centrality	$B_i = \sum_{i \neq s \neq t} l_{st}^i / l_{st}$	Measures the importance of individual nodes for the transport of information or matter,

4.1. Monitoring and processing data

The data required for the association rules in this research are derived from the Wuhan Subway Safety Early Warning System. The system was developed by the Construction Management Institute of Huazhong University of Science and Technology in Wuhan, China. A portion of the collected data collected is presented in Table 3.

The monitoring types and related national standards are derived from the *Technical specification for urban rail transit engineering monitoring (GB 50911-2013)*, which are presented in Table 4. Fig. 3 identifies various risk types from the monitoring data used to enable their early detection: (1) green-general, (2) yellow-significant, (3) orange-high, (4) red-serious. The risk factors of monitoring data are replaced by A, B C and so on, as denoted in Table 5.

The classification of four risk types is based on the standard GB 50911-2013. The original monitoring data of the Li Miao-lu Station has been discretised and is presented in Fig. 4. Three sub-graphs represent the yellow, orange and red alerts. Respectively, different colors of the discount segment represent distinct types of monitoring. The horizontal axis represents time, and the vertical the number of occurrences. It can be seen in Fig. 4 that in July and August a series of data anomalies occurred compared to other months. This may be related to external conditions such as weather and climate characteristics of the city.

4.2. Association rules mining

The rule that satisfies both minimum support and confidence is referred to as being a "strong association rule", which are categorized as being effective and valid. The confidence results of generated association rules by ARM in this paper are very close and are not useful thresholds to filter association rules. Therefore, the minimum support and lift are selected to filter valid association rules. Within subway construction, the sample size, the minimum support degree and lift as well as other parameters can influence the quality of the data that is collected. Therefore, the selection of parameters is important. To select the appropriate parameters, a sensitivity analysis on the support and lift threshold values using the *Apriori* algorithm was conducted. Fig. 5 identifies the number of association rules excavated under the different degrees of support and lift.

The association rules resulting from applying the *Apriori* algorithm are not all valid. To obtain effective and useful rules, the *minimum support* and *lift* should be set appropriately. If the values for the minimum support and lift are set a low level, irrelevant rules will be generated. If they are set too high, no rules will be generated. Therefore, the setting of these two parameters requires the actual data set to be provided. By comparing the number of association rules in Fig. 8 with the input parameter values, the balance between the two extremes and select the compromise data can be considered. The minimum support and lift input parameters are as follows:

$$\begin{aligned} &\textit{Minimum support} = 45\% \\ &\qquad \textit{Lift} \geq 1.1 \end{aligned}$$

There are 2123 association rules generated by ARM based on the *Apriori*. The rules were sorted by *lift*, and are prioritized (Table 6). For example, the first association rule {*J, M, D, A*⇒ *C, L, K*} ({*Yellow-Structural Settlement, Orange-Concrete Strut Axial Forces, Yellow-Building Settlement, Yellow-Surface Settlement*⇒ *Red-Surface Settlement, Yellow-Concrete Strut Axial Forces, Red-Structural Settlement*}) (Support = 67.5%, Confidence = 80.4%, Lift = 1.26), implies that when the structural, building, surface settlement simultaneously portray a yellow alert and concrete strut axial forces portray an orange alert, then there is potential for the site to be subjected to abnormal settlement (red alert).

The authenticity and effectiveness of the conclusions of the ARM are verified by conducting a site inspection. The Kang Yi garden building is chosen to validate the results. This building resides in the vicinity of the foundation pit under construction and thus may be subjected to settlement. The settlement observation points are located in the position that can reflect the deformation characteristics of the building, such as the wall angle of the building. The results indicate that the deformation

Table 2
Characteristics of various networks.

Type of network	Characteristics path length (L)	Clustering coefficient (C)	Degree distribution ($P(k)$)
Regular	Long	Large	Point to point
Random	Short	Small	Poisson or Binomial
Small-world	Short	Large	Exponential or power-law
Scale-free	Short	Large	Power-law
Real-world	Short	Large	Similar power-law

Table 3
Original monitoring data.

No.	Surface settlement (mm)	Building settlement (mm)	Groundwater level (mm)	Structural settlement (mm)	Column settlement (mm)	Underground continuous wall Inclinometer (mm)	Steel Strut Axial Forces (kn)	Earth pressure (Mpa)	Concrete Strut Axial Forces (kn)
1	21.55705	22.5513	20.4356	22.2621	5.51	0.0	8.88	13.27	4.79
2	21.6017	22.55905	20.652	22.26539	4.58	0.0	11.24	11.89	1.34
3	21.60244	23.86117	20.7842	22.29162	5.07	0.0	19.98	19.28	0.54
4	21.63255	2.61	21.0689	22.2816	4.09	0.0	16.89	22.04	0.81
5	21.56928	2.96	21.0546	22.3203	6.1	0.0	23.19	23.8	0.33
6	21.63311	2.87	20.9891	22.3144	3.74	0.0	2.67	20.42	0.17
7	21.65076	3.89	21.4553	22.30661	6.47	0.0	3.89	20.63	0.3
8	21.59056	5.1	21.7564	22.3006	7.69	0.0	11.53	20.2	0.14
9	21.58611	5.91	21.7932	22.31538	5.55	0.0	27.6	18.9	0.15
10	21.61477	7.63	21.5424	22.33983	5.27	0.0	31.39	18.1	0.03
…	…	…	…	…	…	…	…	…	…
142	21.59436	9.3	20.4568	22.2316	4.63	0.0	42.75	3.12	2.54
143	21.61069	10.76	20.743	22.2299	5.1	0.0	3.91	17.88	2.82
145	21.54060	11.43	21.0482	22.2424	7.41	0.0	7.83	17.62	3.04
146	21.51432	12.37	21.6495	22.2951	5.39	0.0	14.68	4.55	3.27
147	21.67449	12.91	21.7123	22.2621	7.94	0.0	37.46	5.21	4.03
148	21.88796	14.47	22.3026	22.2614	6.32	0.0	36.42	3.43	3.74
149	21.85632	15.99	20.6845	22.1416	7.59	0.0	8.89	2.98	3.38
150	21.89206	16.66	22.483	22.0769	7.89	0.0	16.17	4.34	3.27

of the deep foundation pit is affected by precipitation, wall deformation and support time. While excavating earth, the soil in the pit can be displaced, which can contribute to its uplift resulting in surface settlement. Surface deformation will then cause deformation of the supporting structure, which significantly increases the value of its axial force (Fig. 6). When the support force reaches the alert value, the deformation outside the pit also continues to increase. It can be seen that there is a coupling relationship between surface subsidence, building settlement and the concrete supports axial force. Rules can also be pruned by either the minimum or maximum number of monitoring types involved. In Table 6, the set of four types with the highest lift are presented.

4.3. Establishment and complex network based on ARM

Underground engineering of urban rail systems must monitor the supporting structure, surrounding rock and soil and environment during construction (GB50497-2013). Three specific sedimentary monitoring types serve as guiding examples: (1) surface settlement, (2) building settlement, and (3) structural settlement. The association rules for each of the three monitoring types are examined hereinafter.

To conduct risk analysis for each of the different settlement types, association rules need to be selected. For example, if the CN of C (Red-Surface Settlement) needs to be established, then its target monitoring type is chosen and the right of association rules related to it are selected. Then, C (Red-Surface Settlement), F (Red-Building Settlement) and K (Red-Structural Settlement) are chosen to be the target monitoring types, respectively. After choosing the target monitoring type, the association rules for the varying settlements can be obtained. Thus, the association rules of surface settlement, building settlement, and structural settlement are 819, 1863 and 753 respectively.

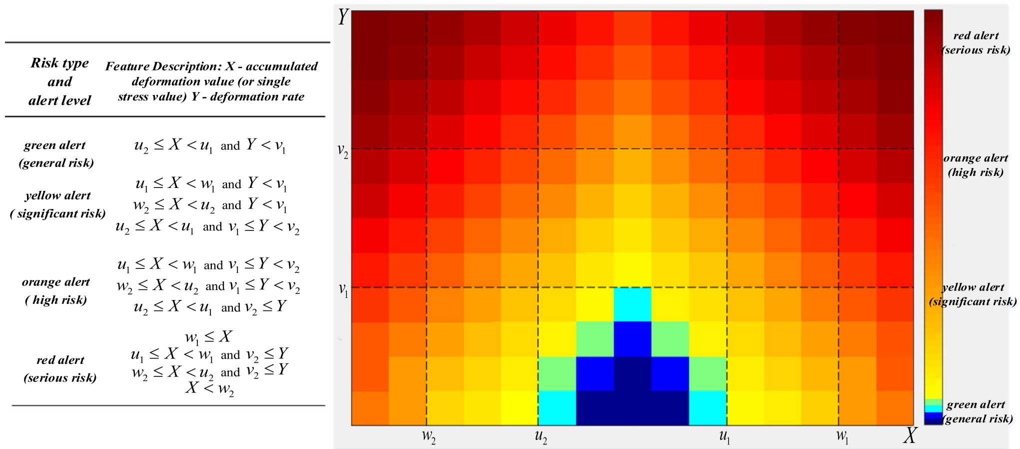
Each monitoring type in the above association rule represents a node of the network. Each association rule that causes the target monitoring type to form the edge of the network. Lift is chosen to be the weight of the network. The association rules of surface, building, and structural settlement are then mapped into three weighted networks. The risk factors of three different settlements are constructed into complex networks using the visibility graph algorithm. The weight (lift) values of the top 20 edges of the settlement association rules are matched to different colors for the various weight matrices. The color red denotes a higher value of lift, while blue shows a lower value, which refers to differences in strength for the risk factors. The weight matrix and complex network are shown in Fig. 7.

4.4. Topology properties of the complex network

The network structure-based analysis is related to macro-level complex features, such as “small world” and scale-free phenomenon. To determine the characteristics of settlement networks and evaluate their complexity, these features were analyzed. The node influence-based analysis involves their micro-level influence indexes. For example, node degree, closeness and betweenness were used to identify keynote which have vital influence on the whole network. Features of complex network for three specific sedimentary monitoring types are shown in

Table 4
Monitoring types and related standards.

Monitoring types	Absolute critical value		Absolute alert value		Relative value	
	Positive $\mathbf{u}_1$(mm|kn)	Negative $\mathbf{u}_2$(mm|kn)	Positive $\mathbf{w}_1$(mm|kn)	Negative $\mathbf{w}_2$(mm|kn)	Critical $\mathbf{v}_1$(mm/d|kn/d)	Alert $\mathbf{v}_2$(mm/d|kn/d)
Surface settlement	8	−24	10	−30	3	5
Building settlement	8	−24	10	−30	2	3
Groundwater level	800	−800	1000	−1000	400	500
Structural settlement	7	−40	10	−56	4	6
Column settlement	8	−8	10	−10	2	3
Underground continuous wall inclinometer	32	−32	40	−40	3	5
Steel strut axial forces	2005.6	−2005.6	2507	−2507	400	500
Earth pressure	10,000	−10,000	100,000	−100,000	300	500
Concrete strut axial forces	7712	−7712	9640	−9640	300	500



Risk type and alert level	Feature Description: X - accumulated deformation value (or single stress value) Y - deformation rate
green alert (general risk)	$u_2 \le X < u_1$ and $Y < v_1$
yellow alert (significant risk)	$u_1 \le X < w_1$ and $Y < v_1$ $w_2 \le X < u_2$ and $Y < v_1$ $u_2 \le X < u_1$ and $v_1 \le Y < v_2$
orange alert (high risk)	$u_1 \le X < w_1$ and $v_1 \le Y < v_2$ $w_2 \le X < u_2$ and $v_1 \le Y < v_2$ $u_2 \le X < u_1$ and $v_2 \le Y$
red alert (serious risk)	$w_1 \le X$ $u_1 \le X < w_1$ and $v_2 \le Y$ $w_2 \le X < u_2$ and $v_2 \le Y$ $X < w_2$

Fig. 3. Alert level and grading standards.

Table 5
The alphabet corresponding to the risk factors.

New	Old	New	Old
A	Yellow-Surface Settlement	L	Yellow-Concrete Strut Axial Forces
B	Orange-Surface Settlement	M	Orange-Concrete Strut Axial Forces
C	Red-Surface Settlement	N	Red-Concrete Strut Axial Forces
D	Yellow-Building Settlement	O	Yellow-Groundwater Level
E	Orange-Building Settlement	P	Orange-Groundwater Level
F	Red-Building Settlement	Q	Yellow-Soil Pressure
G	Yellow-Column Settlement	R	Orange-Soil Pressure
H	Orange-Column Settlement	S	Red-Soil Pressure
I	Red-Column Settlement	T	Yellow-Underground Continuous Wall Skew
J	Yellow-Structural Settlement	U	Yellow-Steel Support Axial Force
K	Red-Structural Settlement	V	Orange-Steel Support Axial Force

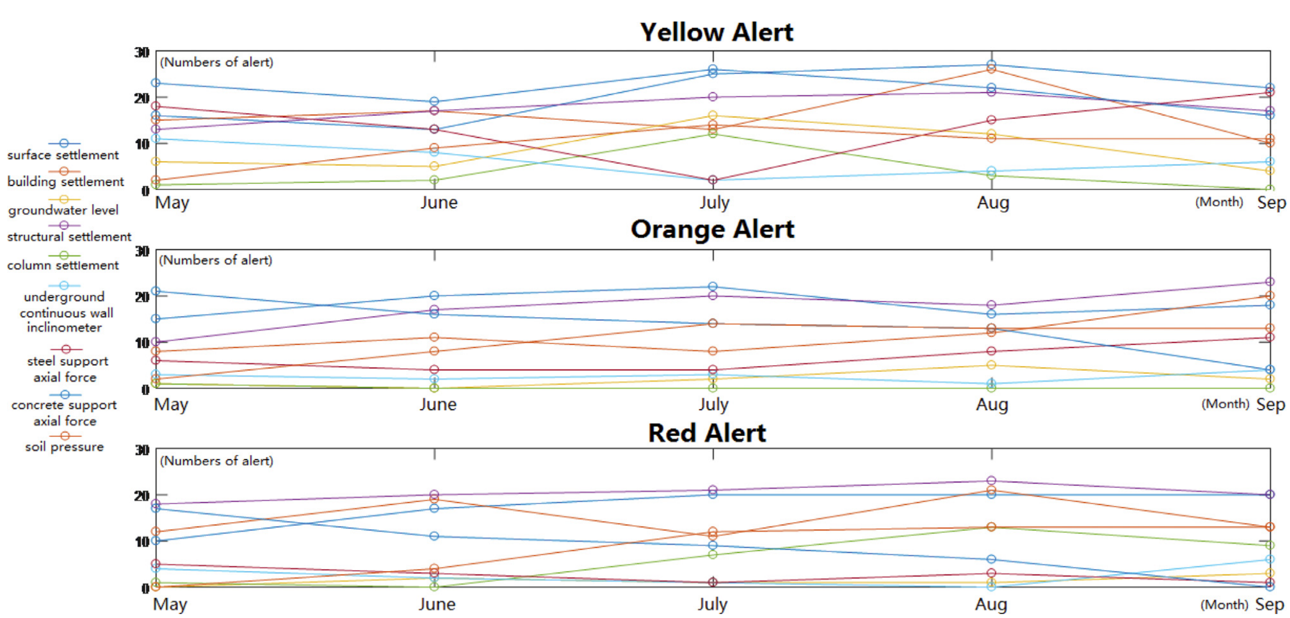


Fig. 4. Abnormal data statistics.

Table 7.

4.4.1. Network structure based analysis

The occurrence of power law distributions on the nodes throughout the entire network signifies the aggregation of risk factors between settlements. This suggests that the CNs for inter-relationship of risk factors for settlement are not random and dependent but possess a universal property of scale-free. The connectivity of nodes is unequal in the scale-free network. Thus, the nodes with larger degree in related network are more vulnerable to failure in the whole the system.

As can be seen from Fig. 8, the node degree distribution shows the different influences between them. The more connections a node has, the greater its influence in the network. The nodes in the building settlement network are generally larger than those within the surface

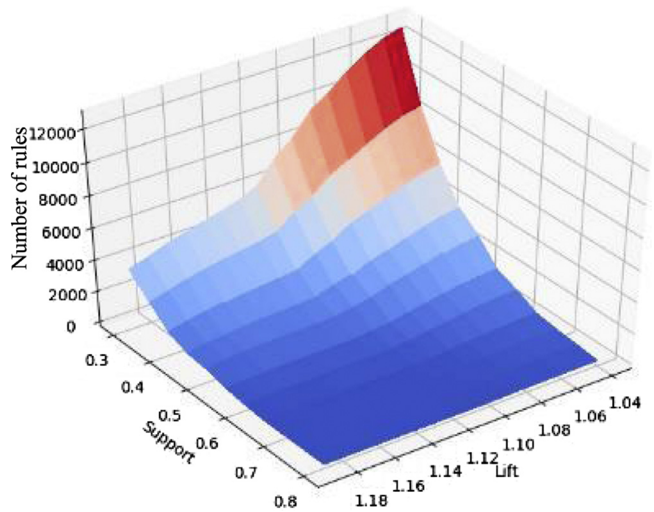


Fig. 5. The total number of rules that can be generated, based on different values of support and lift.

Table 6
Top six rules by lift.

Left-hand side of the rules	Right-hand side of the rules	Support	Confidence	Lift
J, M, D, A	C, L, K	0.675496689	0.803921569	1.264502
C, L, K	J, M, D, A	0.635761589	0.854166667	1.264502
A, K, M	D, C, N	0.754966887	0.684210526	1.259949
D, C, N	A, K, M	0.543046358	0.951219512	1.259949
J, C, D, N	M, K, A	0.536423841	0.950617284	1.259151
M, K, A	J, C, D, N	0.754966887	0.675438596	1.259151

structural settlement network (Fig. 8). This implies the presence of high-degree nodes, referred to as hubs, in constructing a settlement network that describes the influence. Moreover, the probability distribution of node degree and its complementary cumulative probability distribution are shown in Fig. 9. It can be seen from Fig. 9 that the surface, building and structural settlement networks all conform to a power law exponent.

The nodes with high betweenness centrality often take the role of a bridge and connect different parts of the network. The betweenness centralities indicate that the networks are highly clustered. It can be inferred that the networks with smaller path lengths yet higher clustering coefficients, also indicate "small-world" characteristics.

The distribution of the clustering coefficient and betweenness centrality of nodes are represented in Figs. 10 and 11. Here the networks for risk factors of settlement possess "small-world" characteristics. In comparison with the building settlement, the node's betweenness centrality is generally lower and clustering coefficient are higher in surface and structural settlement networks. This signifies that surface and structural settlement networks are more clustered and aligned with the characteristics of the small world. The risk transfer path is shorter and speed is faster in these two networks, which suggests that their transitivity is strong. Minor changes to these vital nodes can drastically change the performance of the network. Thus, effective control of risk factors within these networks needs to be undertaken as an increase the average path length can slow down the diffusion efficiency of risk events.

4.4.2. Node influence based analysis

In this research, different centrality measures (e.g., degree, betweenness, and closeness) have been used to identify critical components that may reside within networks. In a CN, the node with the largest degree and betweenness centrality has the strongest control over the information flow. These nodes with high betweenness centrality often take the role of bridging disparate parts of the network. So, by discovering the critical nodes, the propagation of information in risk networks can be controlled, and therefore used to prevent failures in deep foundation pits.

Therefore, network-based ranking algorithms use network representations of the data to compute the inherent value, relevance, or importance of individual nodes in the system [55]. In this research, the ranking algorithms and centrality metrics are used interchangeably.

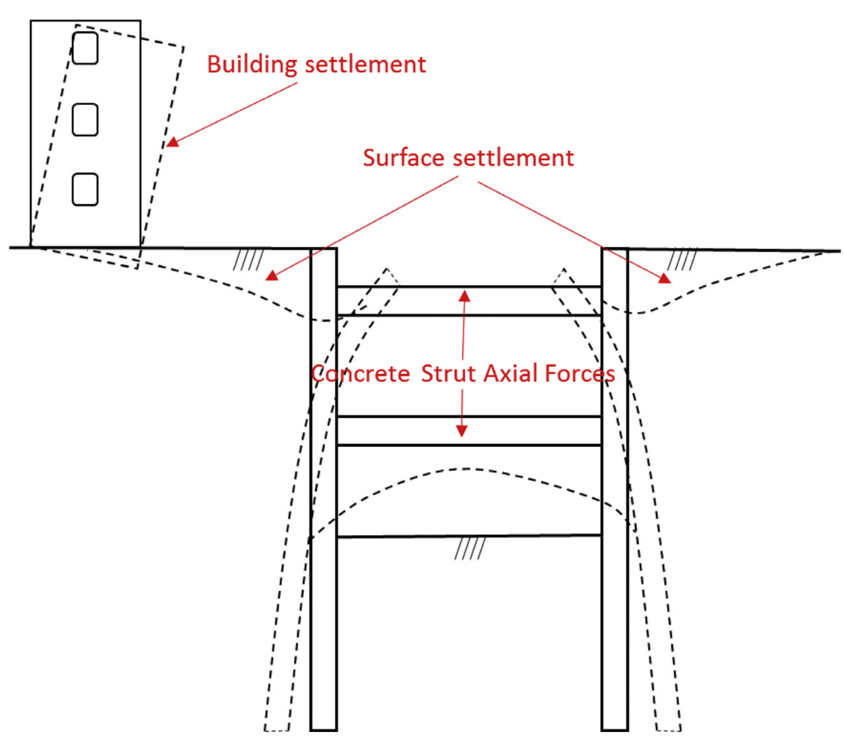


Fig. 6. Retention of bottom pit support.

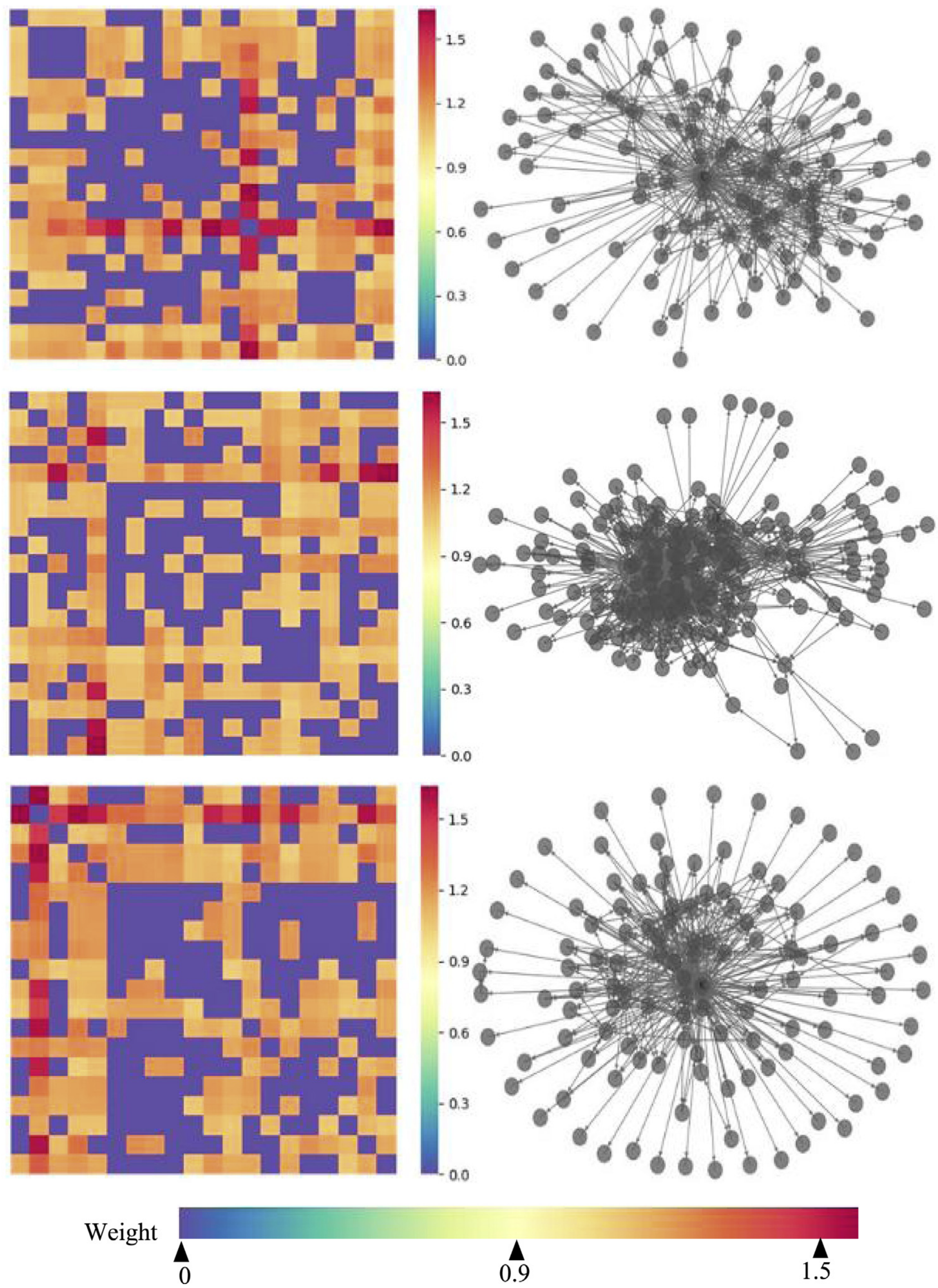


Fig. 7. Main phases establishment for complex network based on ARM.

Table 7
Features of complex network for three specific sedimentary monitoring types.

Settlement type	n	k_{max}	k	k_∂	C	C_∂	L	Network structure
Surface	105	45	6.942	7.209	0.467	0.307	2.2	SF,SW
Building	180	59	10.402	11.601	0.372	0.391	2.9	SF,SW
Structural	118	45	5.483	6.593	0.459	0.332	2.3	SF,SW

The node ranking is divided into four parts: (1) green, (2) yellow, (3) orange and (4) red. Correspondingly, these four levels are consistent with aforementioned four risk types from the monitoring data green-general risk, yellow-significant risk, orange-high risk and red-serious risk (Fig. 3).

4.4.2.1. Analysis for red alert of surface settlement. The target network for surface settlement are presented by degree and betweenness are presented in Fig. 12. The ranking of the nodes in the surface settlement network by centrality metrics (indegree, closeness and betweenness) are presented in Fig. 13. Here most nodes in the surface settlement network are identified in green and yellow areas such as *{A, D, J}* ({*Yellow-Surface Settlement, Yellow-Building Settlement, and Yellow-Structural Settlement*}). Notably, some nodes are in orange such as {*D, M, N*} ({*Yellow-Building Settlement, Orange-Concrete Strut Axial Forces, Red-Concrete Strut Axial Forces*}). Only few nodes are in red area such as *{M} ({Orange-Concrete Strut Axial Forces})*, which suggest that they are of high risk of surface settlement.

Through the analysis of centrality metrics (indegree, closeness and betweenness), most nodes are ranked approximately the same. Among these risk factors, the centrality metrics of *{M} ({Orange-Concrete Strut Axial Forces})* is the highest one with degree (45), closeness (0.639751553) and betweenness (681.631572). Noteworthy, the

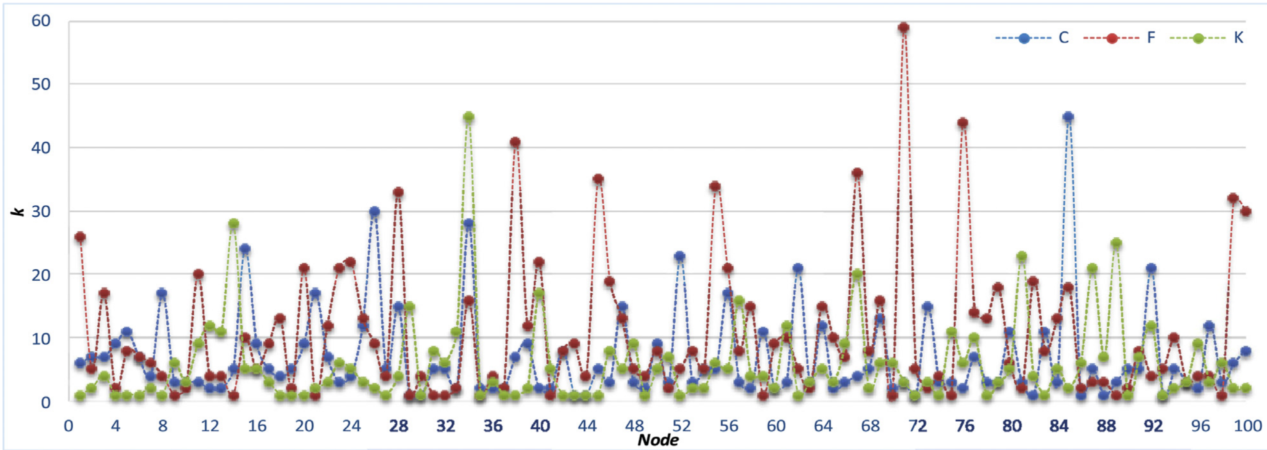


Fig. 8. Distributions of node degree in settlement monitoring network.

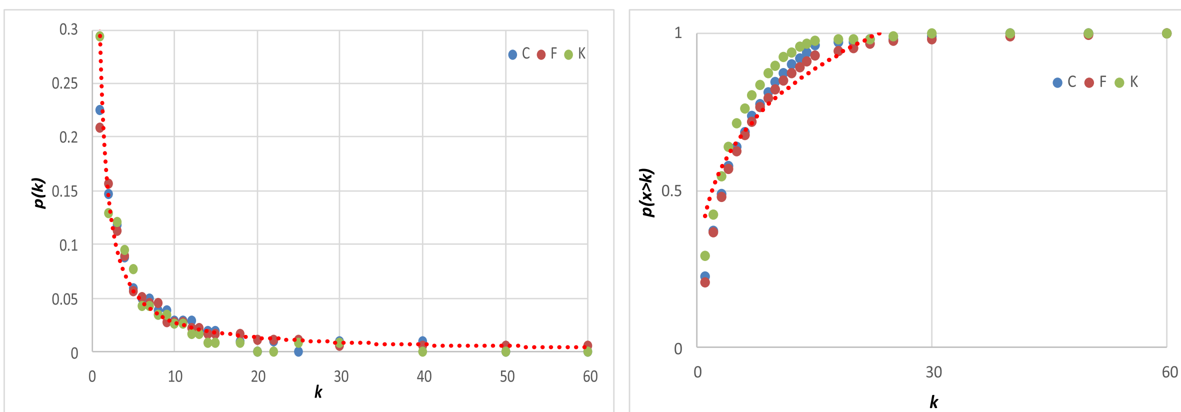


Fig. 9. Probability distribution and cumulative probability distribution of node degree.

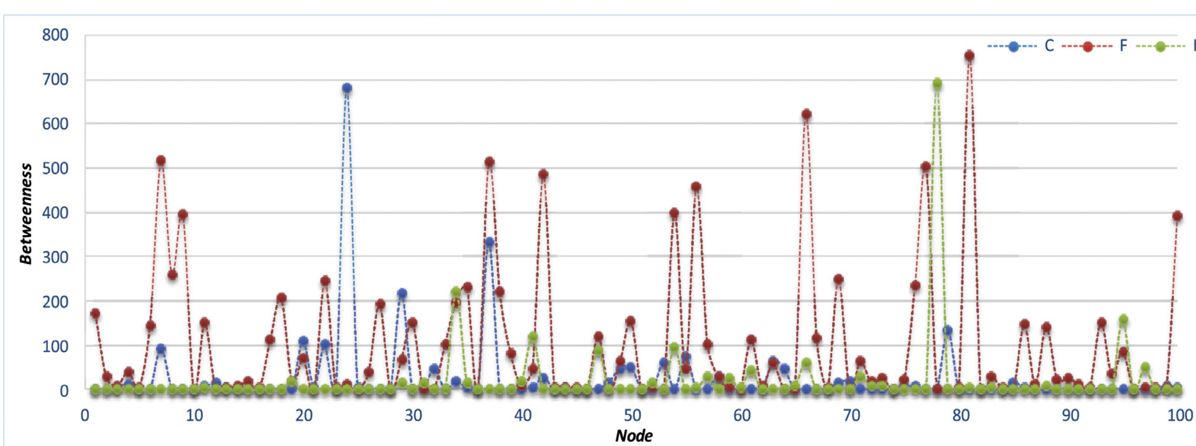


Fig. 10. Distributions of node betweenness in settlement monitoring network.

degree of *{M}* ({Orange-Concrete Strut Axial Forces}) is 45. This indicates there are 45 association rules in the entire network directly or indirectly related to this node. Hence, Orange-Concrete Strut Axial Forces are the most important node in the surface settlement network. Effective control of Orange-Concrete Strut Axial Forces can increase the average path length and slow down the diffusion efficiency of risk events throughout the surface settlement network.

4.4.2.2. Analysis for red alert of building settlement. The target network for building settlement are presented by degree and betweenness (Fig. 14). The ranking of the nodes in the building settlement network by centrality metrics (indegree, closeness and betweenness) are presented in Fig. 15. The most important feature of this network are the most nodes identified in yellow and orange. Several nodes are green and red such as *{A, C, J}* ({*Yellow-Surface Settlement, Red-Surface Settlement, Yellow-Structural Settlement*}) and *{C, D, K}* ({*Red-Surface Settlement, Yellow-Building Settlement, Red-Structural Settlement*}). As shown in Fig. 15, there are nodes that have a medium rank of centrality, while some nodes such as *{C, F, N}* ({*Red-Surface Settlement, Red-Building Settlement, Red-Concrete Strut Axial Forces*})

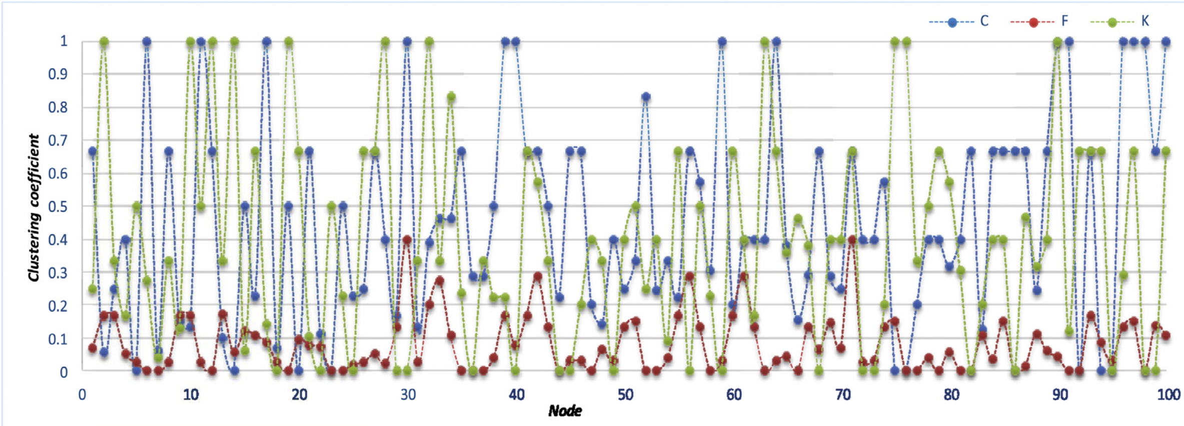


Fig. 11. Distributions of node clustering coefficient in settlement monitoring network.

Fig. 12. Target network of surface settlement by degree and betweenness, (a) degree, (b) betweeness.

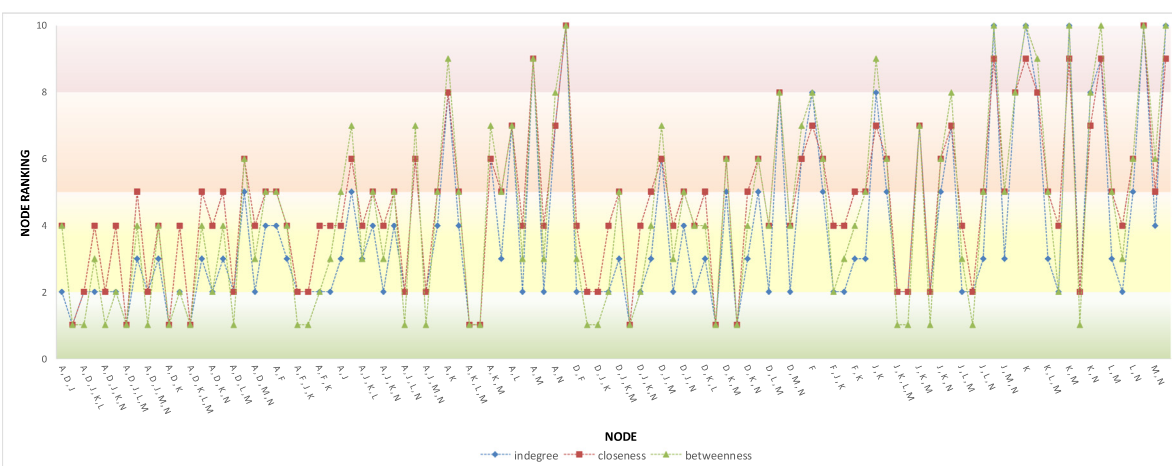
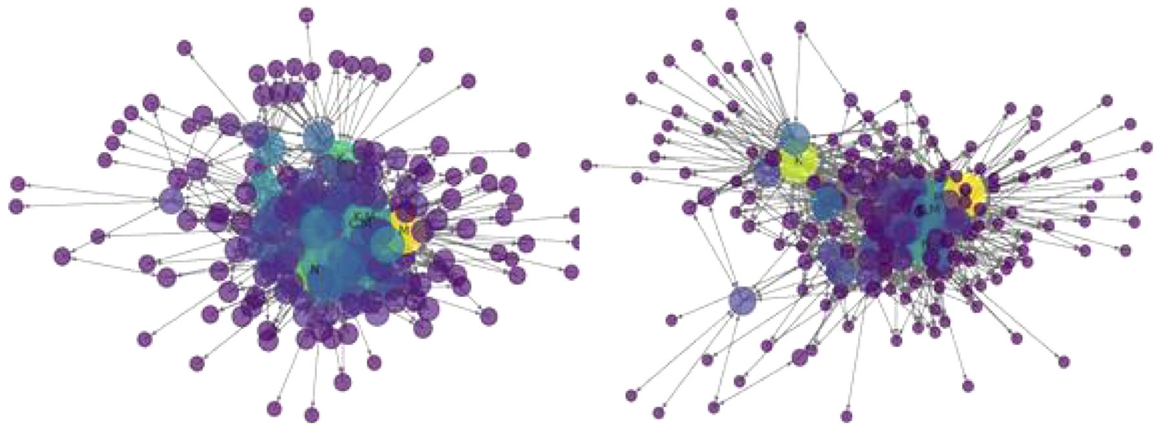


Fig. 13. Ranking of the nodes in the surface settlement network by centrality metrics (Node labels on the horizontal axis correspond to the node labels in the top panel).

have relatively high centrality metrics. The node {*C, N*} ({*Red-Surface Settlement, Red-Concrete Strut Axial Forces*}) is in the red area of Fig. 15 and has the highest betweenness (774.1695555) centrality, indicating that nearly 98% of the paths are connected to the node. The degree (56) and closeness (0.546321223) of this node is also very high amongst compared to others in the building settlement network.

This suggests that coupling exists between the surface settlement and the concrete axial force and a red alert is identified in this instance. In addition, this node has relatively high centrality metrics, which means that it is the hub of the network and have strong impact on the other nodes. Therefore, the combination of Red-Surface Settlement, Red-Concrete Strut Axial Forces should be controlled to mitigate risks



(a) Degree (b) Betweeness

Fig. 14. Target network of building settlement by degree and betweenness, (a) degree, (b) betweeness.

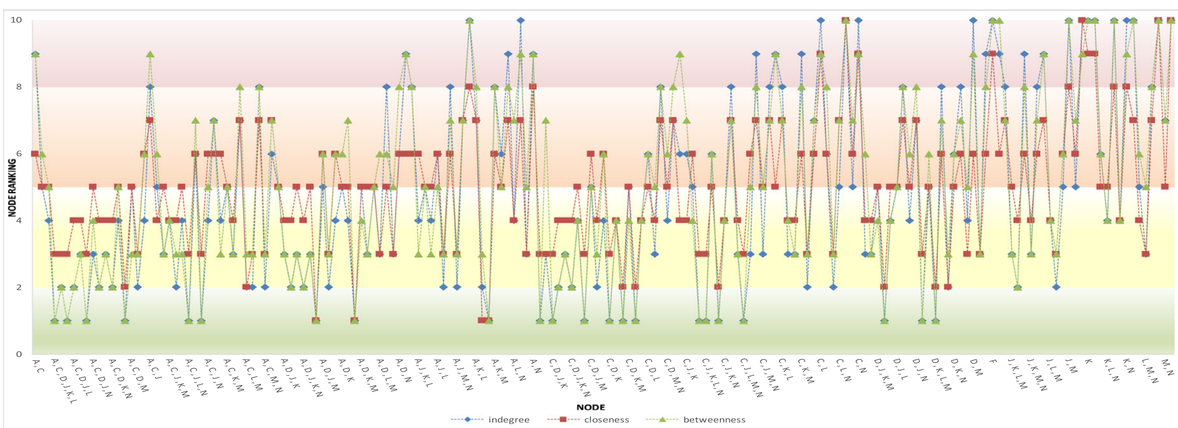


Fig. 15. Ranking of the nodes in the building settlement network by centrality metrics (Node labels on the horizontal axis correspond to the node labels in the top panel).

(a) Degree (b) Betweeness

Fig. 16. Target network of structural settlement by degree and betweenness, (a) degree, (b) betweeness.

across the building settlement network.

4.4.2.3. Analysis for red alert of structural settlement. The target networks for structural settlement are presented by degree and betweenness in Fig. 16. The nodes in the structural settlement network by centrality metrics (indegree, closeness and betweenness) are presented in Fig. 17. The centrality metrics of this network's nodes are low and are distributed in green and yellow area in Fig. 17. Only a limited number of nodes' centrality metrics are high, such as {*C, M*} ({*Red-Structural Settlement, Orange-Concrete Strut Axial Forces*}) and {*N*} ({*Red-Concrete Strut Axial Forces*}). This illustrates that the nodes with the highest centrality metrics play important role in structural settlement. The node {*N*} ({*Red-Concrete Strut Axial Forces*}) has the maximum 'betweenness' and 'closeness' centrality, as shown in Fig. 17. Effective control of *Red-Concrete Strut Axial Forces* can increase the average path length and slow down the diffusion efficiency of risk

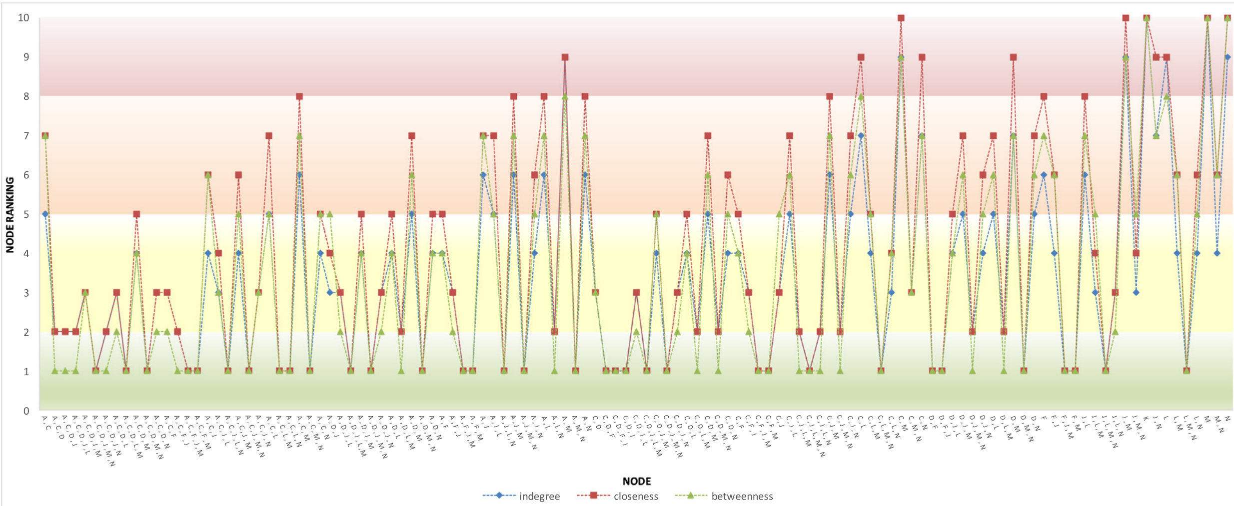


Fig. 17. Ranking of the nodes in the structural settlement network by centrality metrics (Node labels on the horizontal axis correspond to the node labels in the top panel).

events throughout the network, which can also reduce the risk transfer of the building settlement network.

4.5. Results

Based on the above analysis, the coupling relationships of risk factors of deep foundation pits during construction have been revealed. The research has illustrated that coupling relationships exists between the surface settlement, building settlement and the concrete axial force, and enabled the generation of rules such as {*J, M, D, A*⇒ *C, L, K*} ({*Yellow-Structural Settlement, Orange-Concrete Strut Axial Forces, Yellow-Building Settlement, Yellow-Surface Settlement*⇒ *Red-Surface Settlement, Yellow-Concrete Strut Axial Forces, Red-Structural Settlement*}) (Support = 67.5%, Confidence = 80.4%, Lift = 1.26). Risk factors of surface, building and structural settlement networks are found to be both scare-free and small word network. The building settlement's risk factors characteristics were more conforms to scale-free network, whereas surface and structural settlement networks were more akin to a small world. The centrality metrics of the three settlement networks presented different characteristics. Each network has its own most vital node that can have significant impact on its neighbors and even the entire network. Thus, different control points should be effectively monitored for different settlements.

During construction, the data collected at the Li Miao-lu station were used to closely monitor the surrounding conditions. A building settlement monitoring point F057 experienced significant settlement 13.17 mm, the surface settlement monitoring point DB02 nearby experienced extreme settlement 166.39 mm and the concrete strut axial forces of monitoring point TZL05 nearby is 308.82 kn It is important to highlight that the values for the concrete strut's axial forces at the Li Miao-lu station were between −25–40 kn. Therefore, the concrete strut axial forces 308.82 kn of the monitoring point TZL05 was an anomaly compared with others. Noteworthy, the values for the building settlement, surface settlement, and concrete strut axial forces were in the red alert range. Based on the analysis and results presented, the risks associated with surface settlement, building settlement and the concrete axial force are coupled. From the node influence analysis of three different settlements, it is revealed that the monitoring of the axial force of the support structure is a major issue that needs to be monitored when constructing deep foundation pits. The research presented in this paper can provide managers with the knowledge needed to monitor, control and predict the coupling of risks. In this instance, managers can obtain timely feedback in the event of an abnormal situation and therefore ensure necessary engineering emergency measures are undertaken in a timely manner. This may require adjustments to the method and process of construction or modifications to a project's design to ensure to ensure safety.

5. Conclusion

During the process of excavating deep foundation pits, the support structure, soil, adjacent structures and the surrounding environment need to be monitored and systemically analyzed. In comparison with other risk management methods that have been developed, our proposed approach provides a powerful tool to analyze coupled risks associated with the construction of deep foundation pits. The developed association rules in conjunction with CNs enable the determination of frequent itemsets between the uncertain factors that cause an unexpected safety event to occur. As a result, the coupling relationships between risk factors can also be identified. By monitoring during the construction of deep foundation pits, managers are better positioned to make informed decisions that can contribute to improving safety as well as projects overall productivity and performance. The developed method combining ARM with CNs was also verified on a real-life project.

The data mining undertaken in this research provides a decision support aide to identify risks during the construction of subway infrastructure. Using the data acquired from an actual project, meaningful association rules between monitoring type anomalies. However, the quality of mining that is undertaken and the extraction of association rules is dependent on the size of the sample data and number of parameters. Therefore, the parameters need to be made solicited from experts and managers who are engaged in safety management.

There are several possible extensions to this research. Firstly, additional parameters can be introduced to enrich the content of the association rules. In this research, the parameters *Support, Confidence* and *Lift* were used in mining of relationships between monitoring types. Other parameters *Conviction* and *Leverage* can be used in future research to further verify the combinations of the monitoring types once further combinations have been obtained after applying this proposed method in real constructions. In addition, the research focused on analyzing the relationship between monitoring data anomalies and therefore future research should consider examining the coupling of accident risks and monitoring anomalies. Secondly, automatic data collection technology to simplify data collection the data can be pre-processed before performing the association analysis. Monitoring data is often affected by the construction process and environmental factors. For example, if a heavy vehicle passes a deep foundation pit, the ground settlement data may be subjected to transient variation. However, this variation cannot be used to explain the effects of the construction process. Therefore, data cleaning technologies can be adopted to pre-process the actual monitoring data to improve the accurate rate of mining results.

Acknowledgments

The presented work has been supported by the National Science Foundation of China (NSFC) through grant 71471072 and Hubei Technical Innovation Project through grant 2017ACA186.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ress.2019.02.013.

Reference

[1] Marco GN, Matthew A. Turnerb subways and urban growth: evidence from earth. J Urban Econ 2018;108:85–106https://doi.org/10.1016/j.jue.2018.09.002.
[2] Zhou Y, Luo H, Yang Y. Implementation of augmented reality for segment displacement inspection during tunneling construction. Autom Constr 2017(82):112–21https://doi.org/10.1016/j.autcon.2017.02.007.
[3] Das R, Singh PK, Kainthola A. Numerical analysis of surface subsidence in asymmetric parallel highway tunnels. J Rock Mech Geotech Eng 2017;9(1):170–9https://doi.org/10.1016/j.jrmge.2016.11.009.
[4] Zhong Y. Technical innovation for Sanyang Road Cross-river Tunnel Project. Front Eng Manag 2018;5(3):411–5https://doi.org/10.15302/J-FEM-2018206.
[5] Khakzad N, Khan F, Paltrinieri N. On the application of near accident data to risk analysis of major accidents. Reliab Eng Syst Saf 2014;126:116–25https://doi.org/10.1016/j.ress.2014.01.015.
[6] Zhou H, Wang H, Zeng W. Smart construction site in mega construction projects: a case study on island tunneling project of Hong Kong-Zhuhai-Macao Bridge. Front Eng Manag 2018;5(1):78–87https://doi.org/10.15302/J-FEM-2018075.
[7] Woochul N, Ki-Yong O, Bogdan IE. Evolution of the dynamic response and its effects on the serviceability of offshore wind turbines with stochastic loads and soil degradation. Reliab Eng Syst Saf 2018:1–13https://doi.org/10.1016/j.ress.2018.03.017.
[8] Chao F, Franck M, Enrico Z, Jean C. Network theory-based analysis of risk interactions in large engineering projects. Reliab Eng Syst Saf 2012;106:1–10https://doi.org/10.1016/j.ress.2012.04.005.
[9] Fink O, Zio E, Weidmann U. Predicting component reliability and level of degradation with complex-valued neural networks. Reliab Eng Syst Saf 2014;121:198–206https://doi.org/10.1016/j.ress.2013.08.004.
[10] Ki-Chang H, Sangyoon M, Hangseok C, Jeongjun P, In-Mo L. Risk analysis using fault-tree analysis (FTA) and analytic hierarchy process (AHP) applicable to shield TBM tunnels. Tunn Undergr Space Technol 2015;49:121–9https://doi.org/10.1016/j.tust.2015.04.007.
[11] Wu X, Liu H, Zhang L, Skibniewski M, Deng Q, Teng J. A dynamic Bayesian network based approach to safety decision support in tunnel construction. Reliab Eng Syst Saf 2015;134:157–68https://doi.org/10.1016/j.ress.2014.10.021.
[12] Sihombing F, Torbol M. Parallel fault tree analysis for accurate reliability of complex systems. Struct Saf 2018;72:41–53https://doi.org/10.1016/j.strusafe.2017.12.003.
[13] Mahdevari S, Haghighat HS, Torabi SR. A dynamically approach based on svm algorithm for prediction of tunnel convergence during excavation. Tunn Undergr Space Technol 2013;38(9):59–68https://doi.org/10.1016/j.tust.2013.05.002.
[14] Wang Y, Zhao X, Wang B. LS-SVM and Monte Carlo methods based reliability analysis for settlement of soft clayey foundation. J Rock Mech Geotech Eng 2013;5(4):312–7https://doi.org/10.1016/j.jrmge.2012.06.003.
[15] Dai H, Cao Z. A wavelet support vector machine-based neural network metamodel for structural reliability assessment 32. John Wiley & Sons Inc; 2017. p. 334–57https://doi.org/10.1111/mice.12257.
[16] Wang J, Liu X, Ni Y. A Bayesian probabilistic approach for acoustic emission-based rail condition assessment. Comput Aided Civ Infrastruct Eng 2017;31(1):1–14https://doi.org/10.1111/mice.12316.
[17] Zhao L, Wang X, Qian Y. Analysis of factors that influence hazardous material transportation accidents based on Bayesian networks: a case study in China. Saf Sci 2012;50(4):1049–55https://doi.org/10.1016/j.ssci.2011.12.003.
[18] Zhou J, Shi X, Du K, Qiu X, Li X, Mitri HS. Feasibility of random-forest approach for prediction of ground settlements induced by the construction of a shield-driven tunnel. Int J Geomech 2017;17(6):04016129https://ascelibrary.org/doi/10.1061/(ASCE)GM.1943-5622.0000817.
[19] Zhou J, Li X, Mitri H S. Classification of rockburst in underground projects: comparison of ten supervised learning methods. J Comput Civil Eng 2016;30(5):04016003https://doi.org/10.1061/(ASCE)CP.1943-5487.0000553.
[20] Zhou Y, Li C, Zhou C, Lou H. Using Bayesian network for safety risk analysis of diaphragm wall deflection based on field data. Reliab Eng Syst Saf 2018(180):152–67https://doi.org/10.1016/j.ress.2018.07.014.
[21] Zhou Y, Li S, Zhou C, Luo H. Intelligent approach based on random forest for safety risk prediction of deep foundation pit in subway stations. J Comput Civil Eng 2019;33(1):05018004https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CP.1943-5487.0000796.
[22] Xin P, Khan F, Ahmed S. Dynamic hazard identification and scenario mapping using bayesian network. Process Saf Environ Prot 2017;105:143–55https://doi.org/10.1016/j.psep.2016.11.003.
[23] Zhou Y, Su W, Ding L, Luo H, Love PED. Predicting safety risks in deep foundation pits in subway infrastructure projects: a support vector machine approach. J Comput Civil Eng 2017;31(5):04017052https://doi.org/10.1061/(ASCE)CP.1943-5487.0000700.
[24] Yu Z, Henry VB, Han S, Mehrdad S. A machine learning framework for assessing post-earthquake structural safety. Struct Saf 2018;72:1–16https://doi.org/10.1016/j.strusafe.2017.12.001.
[25] Khakzad N, Khan F, Amyotte P. Safety analysis in process facilities: comparison of fault tree and Bayesian network approaches. Reliab Eng Syst Saf 2011;96(8):925–32https://doi.org/10.1016/j.ress.2011.03.012.
[26] Ioanna D, Demos CA. Risk-based maintenance scheduling using monitoring data for moored floating breakwaters. Struct Saf 2013;41(3):107–18https://doi.org/10.1016/j.strusafe.2012.10.007.
[27] Jorge I C, Hosseinian S, Sergiy B. Network-based optimization techniques for wind farm location decisions. Front Eng Manag 2018;5(4):533–40https://doi.org/10.15302/J-FEM-2018025.
[28] Fan YB, Liu C, Wang JB. Integrating multi-granularity model and similarity measurement for transforming process data into different granularity knowledge. Adv Eng Inf 2018;37:88–102https://doi.org/10.1016/j.aei.2018.04.012.
[29] Xue YG, Zhang XL, Li SC, Qiu DH, Su MX, Li LP, Li ZQ, Tao YF. Analysis of factors influencing tunnel deformation in loess deposits by data mining: a deformation prediction model. Eng Geol 2018;232:94–103https://doi.org/10.1016/j.enggeo.2017.11.014.
[30] Stolzer AJ, Friend MA, Truong D, William A, Tuccio MA. Measuring and evaluating safety management system effectiveness using data envelopment analysis. Saf Sci 2018;104:55–69https://doi.org/10.1016/j.ssci.2017.12.037.
[31] Noureldin M, Allinson WG, Cinar Y, Baz H. Coupling risk of storage and economic metrics for ccs projects. Int J Greenhouse Gas Control 2017;60:59–73https://doi.org/10.1016/j.ijggc.2017.03.008.
[32] Sakdirat K, Chatpong C. Railway track inspection and maintenance priorities due to dynamic coupling effects of dipped rails and differential track settlements. Eng Fail Anal 2018;93:157–71https://doi.org/10.1016/j.engfailanal.2018.07.009.
[33] Skorupski J. The simulation-fuzzy method of assessing the risk of air traffic accidents using the fuzzy risk matrix. Saf Sci 2016;88:76–87https://doi.org/10.1016/j.ssci.2016.04.025.
[34] Hong Y, Kan Y, Zeng J, Lin H. Multiphysics coupling model for computing pier scour upon simulation and experiment. J Comput Civil Eng 2018;32(1):04017068https://doi.org/10.1061/(ASCE)CP.1943-5487.0000716.
[35] Mia M, Henning T, Costello S. Application of fault tree analysis to identify moisture damage potential in flexible road pavements. J Infrastruct Syst 2017;23(4):04017026https://doi.org/10.1061/(ASCE)IS.1943-555X.0000387.
[36] Zhou HB, Zhang H. Risk assessment methodology for a deep foundation pit construction project in Shanghai, China. J Constr Eng Manag 2011;137(12):1185–94https://doi.org/10.1061/(ASCE)CO.1943-7862.0000391.
[37] Hu X, Yan X. Construction safety evaluation of construction project based on risk coupling mechanism. Saf Environ Eng 2015;22(6):134–8https://doi.org/10.13578/j.cnki.issn.1671-1556.2015.06.026.
[38] Wu X, Wu K, Shen M. Study on coupling risk of metro construction safety based on N-K model. China Saf Sci J 2016;26(4):96–101https://doi.org/10.16265/j.cnki.issn1003-3033.2016.04.018.
[39] Abhishek V, Sudha Das K, Maiti J, Krishna OB. Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports. Saf Sci 2014;70:89–98https://doi.org/10.1016/j.ssci.2014.05.007.
[40] Sousa H, Santos LO, Chryssanthopoulos M. Quantifying monitoring requirements for predicting creep deformations through Bayesian updating methods. Struct Saf 2019;76:40–50https://doi.org/10.1016/j.strusafe.2018.06.002.
[41] Zhou NQ, Vermeer PA, Lou RX, Tang YQ, Jiang SM. Numerical simulation of deep foundation pit dewatering and optimization of controlling land subsidence. Eng Geol 2010;114(3):251–60https://doi.org/10.1016/j.enggeo.2010.05.002.
[42] Pastén D, Torres F, Toledo B, Muñoz V, Rogan J, Valdivia J. Non-universal critical exponents in earthquake complex networks. Physica A 2018;491:445–52https://doi.org/10.1016/j.physa.2017.09.064.
[43] Long W, Guan L, Shen J, Song L, Cui L. A complex network for studying the transmission mechanisms in stock market. Physica A 2017;484:345–57https://doi.org/10.1016/j.physa.2017.04.043.
[44] Scabini L, Fistarol D, Cantero S, Gonçalves W, Machado B, Rodrigues J. Angular descriptors of complex networks: a novel approach for boundary shape analysis. Expert Syst Appl 2017;89:362–73https://doi.org/10.1016/j.eswa.2017.08.009.
[45] Cats O, Koppeno G, Warnierc M. Robustness assessment of link capacity reduction for complex networks: application for public transport systems. Reliab Eng Syst Saf 2017;167:544–53https://doi.org/10.1016/j.ress.2017.07.009.
[46] Feng J, Li XM, Mao BH, Xu Q, Bai Y, Dawson KA. Weighted complex network analysis of the beijing subway system: train and passenger flows. Physica A 2017;474:213–23https://doi.org/10.1016/j.physa.2017.01.085.
[47] Telikani A, Shahbahrami A. Data sanitization in association rule mining: an analytical review. Expert Syst Appl 2018;96:406–26https://doi.org/10.1016/j.eswa.2017.10.048.
[48] Song K, Lee K. Predictability-based collective class association rule mining. Expert

Syst Appl 2017;79:1–7https://doi.org/10.1016/j.eswa.2017.02.024.
[49] Martín D, Alcalá-Fdez J, Rosete A, Herrera F. Nicgar: a niching genetic algorithm to mine a diverse set of interesting quantitative association rules. Inf Sci 2016;355-356:208–28https://doi.org/10.1016/j.ins.2016.03.039.
[50] Luna J, Cano A, Sakalauskas V, Ventura S. Discovering useful patterns from multiple instance data. Inf Sci 2016;357:23–38https://doi.org/10.1016/j.ins.2016.04.007.
[51] Sowan B, Dahal K, Hossain MA, Zhang L, Spencer L. Fuzzy association rule mining approaches for enhancing prediction performance. Expert Syst Appl 2013;40(17):6928–37https://doi.org/10.1016/j.eswa.2013.06.025.
[52] Hassan T, Karim A, Kim J, Jeon M. Cdim: document clustering by discrimination information maximization. Inf Sci 2015;316(5):87–106https://doi.org/10.1016/j.ins.2015.04.009.
[53] Li N, Zeng L, He Q, Shi Z. Parallel implementation of Apriori algorithm based on MapReduce. 13th ACIS international conference on software engineering, artificial intelligence, networking and parallel/distributed computing. 1. IEEE; 2012. p. 236–41https://ieeexplore.ieee.org/document/6299286.
[54] Mahyar H, Hasheminezhad R, Ghalebi KE, Ali N, Radu G, Ali M, Hamid RR. Compressive sensing of high betweenness centrality nodes in networks. Phys A Stat Mech Appl 2018;497:166–84https://doi.org/10.1016/j.physa.2017.12.145.
[55] Liao H, Mariani M, Medo M, Zhang Y, Zhou M. Ranking in evolving complex networks. Phys Rep 2017;689:1–52https://doi.org/10.1016/j.physrep.2017.05.001.