A Knowledge Discovery Method for Landslide Monitoring Based on K-core Decomposition and the Louvain Algorithm

Ping Wang (wp@my.swjtu.edu.cn)
Southwest Jiaotong University https://orcid.org/0000-0001-6394-8374

Jun Zhu
Southwest Jiaotong University

Wei-lian Li
Southwest Jiaotong University

Ya-kun Xie
Southwest Jiaotong University

Fu Lin
Southwest Jiaotong University

Research Article

Keywords: landslide monitoring, co-occurrence network, K-core decomposition, Louvain algorithm, knowledge discovery

Posted Date: June 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-556511/v1

Version of Record: A version of this preprint was published at ISPRS International Journal of Geo-Information on March 22nd, 2022. See the published version at https://doi.org/10.3390/ijgi11040217.

A knowledge discovery method for landslide monitoring based on K-core decomposition and the Louvain algorithm

Ping Wang [a], Jun Zhu[a*], Wei-lian Li[a], Ya-kun Xie[a], Lin Fu[a]

a Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610031, China;

** Correspondence: zhujun@swjtu.edu.cn;*

Abstract: Landslide monitoring plays an important role in predicting, forecasting and preventing landslides. Quantitative explorations at the subject level and fine-scale knowledge in landslide monitoring research can be used to provide information and references for landslide monitoring status analysis and disaster management. In the context of the large amount of keyword co-occurrence network information, it is difficult to clearly determine and display the domain topic hierarchy and knowledge structure. This paper proposes a landslide monitoring knowledge discovery method that combines the K-core decomposition and Louvain algorithms. In this method, author keywords from the literature are used as nodes to construct a weighted co-occurrence network, and a pruning standard value is defined for K. The K-core approach is used to decompose the network into subgraphs. Combined with the unsupervised Louvain algorithm, subgraphs are divided into different topic communities by setting a modularity change threshold, which is used to establish a topic hierarchy and identify fine-scale knowledge related to landslide monitoring. Based on the Web of Science, a comparative experiment involving the above method and a high-frequency keyword subgraph method for landslide monitoring knowledge discovery is performed. In the resulting 5-core network subgraph of landslide monitoring keyword co-occurrence, 17 community structures can be identified, and the degree value and density of subcommunities are analysed by taking the community with the largest proportion of nodes as an example. The results show that the retention time of the proposed method is significantly lower than that of the traditional method.

Keywords: *landslide monitoring; co-occurrence network; K-core decomposition; Louvain algorithm; knowledge discovery*

JUN ZHU received his M.S. degree in geodesy and survey engineering from Southwest Jiaotong University, Chengdu, China, in 2003 and his Ph.D. degree in cartography and geographic information systems from the Chinese Academy of Sciences, Beijing, China, in 2006. From 2007 to 2008, he was a Postdoctoral Research Fellow at the Chinese University of Hong Kong, Shatin, Hong Kong. Currently, he is a Professor with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University. His research interests include computer vision, 3-D GIS technology and virtual geographic environments. E-mail: zhujun@swjtu.edu.cn;

Ping Wang received her B.S. degree in geographic information science from the College of Environment and Planning, Henan University in 2017, and her M.S. degree in cartography and geographic information from the Faculty of Geomatics, Lanzhou Jiaotong University in 2020. She is currently pursuing her Ph.D. degree at Southwest Jiaotong University, Chengdu, China. Her research interests include disaster scene visualization and virtual geographic environment. E-mail: wp@my.swjtu.edu.cn;

Wei-lian Li received the B.S. degree in survey engineering from Tianjin Chengjian University, Tianjin, China, in 2015 his Ph.D. degree in geodesy and survey engineering from Southwest Jiaotong University, in 2021. His research interests include virtual geographic environments and disaster scene visualization. E-mail: vgewilliam@my.swjtu.edu.cn

Ya-kun Xie received his B.S. degree in survey engineering from the School of Survey Engineering, Henan University of Urban Construction, Pingdingshan, China, in 2015 and his M.S. degree in geodesy and survey engineering from the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China, in 2018, where he is currently pursuing his Ph.D. degree. His research interests include intelligent fire control, computer vision, and remote sensing image processing. E-mail: yakunxie@my.swjtu.edu.cn;

Lin Fu received the B.S. degree in geographic information science from the School of Land and Resources, China West Normal University, NanChong, China, in 2017. He is currently working toward the Ph.D. degree at Southwest Jiaotong University, Chengdu, China. His research interests include 3-D GIS technology, deep learning, and virtual geographic environments. E-mail: fulin_swjt@163.com.

Declarations

Funding: This paper was supported by the National Natural Science Foundation of China (Grant Nos. U2034202 and 41871289), Sichuan Science and Technology Program (Grant No.2020JDTD0003).

Conflicts of interest/Competing interests: The authors have no conflicts of interest to declare that are relevant to the content of this article.

Data availability statement: The data that support the findings of this study are openly available in Web of Science at [http://isiknowledge.com/wos].

A knowledge discovery method for landslide monitoring based on K-core decomposition and the Louvain algorithm

Ping Wang[1], Jun Zhu[1,*], Wei-lian Li[1],Ya-kun Xie[1], Lin Fu[1]

1 Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610031, China;

* Correspondence: zhujun@swjtu.edu.cn;

Abstract: Landslide monitoring plays an important role in predicting, forecasting and preventing landslides. Quantitative explorations at the subject level and fine-scale knowledge in landslide monitoring research can be used to provide information and references for landslide monitoring status analysis and disaster management. In the context of the large amount of keyword co-occurrence network information, it is difficult to clearly determine and display the domain topic hierarchy and knowledge structure. This paper proposes a landslide monitoring knowledge discovery method that combines the K-core decomposition and Louvain algorithms. In this method, author keywords from the literature are used as nodes to construct a weighted co-occurrence network, and a pruning standard value is defined for K. The K-core approach is used to decompose the network into subgraphs. Combined with the unsupervised Louvain algorithm, subgraphs are divided into different topic communities by setting a modularity change threshold, which is used to establish a topic hierarchy and identify fine-scale knowledge related to landslide monitoring. Based on the Web of Science, a comparative experiment involving the above method and a high-frequency keyword subgraph method for landslide monitoring knowledge discovery is performed. In the resulting 5-core network subgraph of landslide monitoring keyword co-occurrence, 17 community structures can be identified, and the degree value and density of subcommunities are analysed by taking the community with the largest proportion of nodes as an example. The results show that the retention time of the proposed method is significantly lower than that of the traditional method.

Keywords: landslide monitoring; co-occurrence network; K-core decomposition; Louvain algorithm; knowledge discovery

Introduction

Landslide monitoring provides strong technical support for understanding landslide evolution processes and is an important approach for disaster prevention and reduction (Whiteley et al. 2020; Xu et al. 2021; Lollino et al. 2020; Schlgel et al. 2015). Currently, information from different subject levels and fine-scale knowledge related to landslide monitoring research are being obtained; additionally, key technologies and disaster-causing factors are being in landslide monitoring to provide a reference for scientific analyses, disaster prevention and mitigation, and disaster monitoring.

Landslide monitoring is a popular topic in the field of landslide research, and there are several ways to discover landslide monitoring knowledge. One way involves literature reviews of landslide monitoring studies, equipment, methods and technologies from a qualitative perspective (Solari et al. 2020; Whiteley et al. 2019; Aubaud et al. 2013). In addition, on-site investigations can be used to obtain actual surveys of landslide and deformation characteristics (Angeli et al. 2010; Zhang et al. 2018). Another less common approach involves summarizing the monitoring strategies used by landslide warning systems through various statistical methods (Pecoraro et al. 2018). However, these studies are based on classic literature and seldom involve quantitative analyses of landslide monitoring research or fine-scale knowledge.

A keyword co-occurrence network is a network formed by keywords and their co-occurrence relationships in the field of bibliometrics; such networks can quantitatively reflect the development process of scientific knowledge and corresponding structural relationships (Small 1973; Forliano et al. 2021; Weeds et al. 2005; Kessler 1963). In recent years, keyword co-occurrence networks have been

widely used in various fields, such as stem cell research (Yang et al. 2020), epilepsy genetics (Gan et al. 2019), crop gene information mining based on the basic characteristics of soil and plants (Li et al. 2020), and malaria research (Fu et al. 2015). The Louvain algorithm is a computationally expensive and time-consuming algorithm (Blondel et al. 2008; Orman et al. 2011; Meo et al. 2011) that is suitable for the division of small and medium-sized networks. Rich text semantic relations can produce dense topics for knowledge discovery (Daud et al. 2012). For some networks with small numbers of nodes, the topic hierarchy can be effectively determined with the Louvain algorithm, but for networks with abundant information or unclear expressions, pruning is needed to determine and display the topic hierarchy. Previous studies (Xiao et al. 2016; Kadi et al. 2017; Zhao et al. 2014) generally set thresholds to screen keywords according to the word frequency or edge weights, but these methods did not consider the possible effect of semantic association between two keywords. Seidman (1983) proposed the K-core approach to express the specific hierarchical structure properties and hierarchical characteristics of networks, and this method has been widely applied to hierarchical decomposition networks (Zhang et al. 2008; Kong et al. 2019; Kitsak et al. 2010; Orman et al. 2009). Notably, the K-core approach can be used to decompose core co-occurrence relationships and can be combined with the Louvain algorithm to efficiently detect the community structure and explore the subject-level and fine-scale information related to landslide monitoring.

This paper presents a combined quantitative and qualitative method to explore the subject hierarchy and fine-scale knowledge in the research field of landslide monitoring and to analyse the degree, density and community division results for the resulting subnetworks. The remainder of this paper is organized as follows. In the first section, the methods, including the overall research concept, are introduced, and the extraction of subgraphs and process of community detection are discussed. The second section provides an analysis of the experimental results, and the data sources and experimental environment are introduced; additionally, a comparison of methods is performed. The final section discusses the study conclusions and future research prospects.

Section 1: Method

1.1 Overall research concept

The technical route of knowledge discovery in the field of landslide monitoring is shown in Fig. 1. The Web of Science preprocesses data through data filtering to reduce invalid data and noise in the original product. According to the word frequency and co-occurrence relationships among the extracted keywords, the co-occurrence matrix is obtained, and a co-occurrence network of weighted keywords related to landslide monitoring is constructed. The pruning index is defined, and a co-occurrence network subgraph is generated based on the structure of the peripheral nodes; the core nodes are retained, and some nodes are removed according to their K-values. The degree and density of subcommunities are analysed, and the threshold value of ΔQ is set; this value increases the degree of tightness in some communities. Finally, the community structure of the subgraph is determined with the Louvain algorithm to analyse the subject-level and fine-scale knowledge in the landslide monitoring field, and the modularity, partitioning time and hierarchy results are compared for different high-frequency keyword subgraphs.

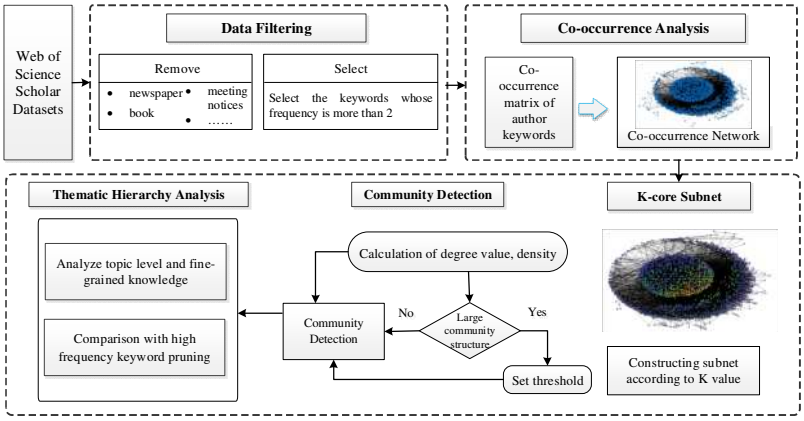


Fig. 1 Technical research route

1.2 Construction of the K-sub map of the co-occurrence network of landslide monitoring

1.2.1 Calculation of the pruning standard based on the K-core

If the number of network nodes is large, it can be difficult to clearly display knowledge and identify and extract information at the theme level in the field of landslide monitoring. Additionally, the Louvain algorithm is characterized by high complexity when detecting network community structures, so it is necessary to prune the network. We retain the main structure of the co-occurrence network through pruning to reduce time and ensure quality, and this process includes three steps. First, the K-value of the entire network node is calculated. Second, the K-value is used to define the pruning subgraph evaluation function and identify the core nodes in the network. Finally, the hierarchical structure based on the K-values of nodes is used to simplify the network. The graph G= (V, E) is obtained, where node n=|V| and edge m=|E|. If a subgraph S satisfies S= (W, E|W) and any node degree value V (V belongs to S) =k, S is the K-shell of graph G. We assess the pruning standard by measuring the strength of the K-value in the main part of the network. The K-value can be calculated as shown in Eq. 1.

$$K = \frac{\sum_i k_i n_i}{m} \tag{1}$$

where k_i represents the K value of each shell, n_i is the number of shells, M is the total number of nodes, and i is the shell for each k value. When the value of node k is less than K, some of the nodes can be deleted; otherwise, all nodes should be reserved. As shown in Fig. 2, the network consists of three shells that contain 12 nodes. Eq. 1 shows that some nodes in shell 1 need to be removed. By defining the K-value, the standard of the pruning generation subgraph is defined. In the next section, the process of generating K-core subgraphs for landslide monitoring is introduced.

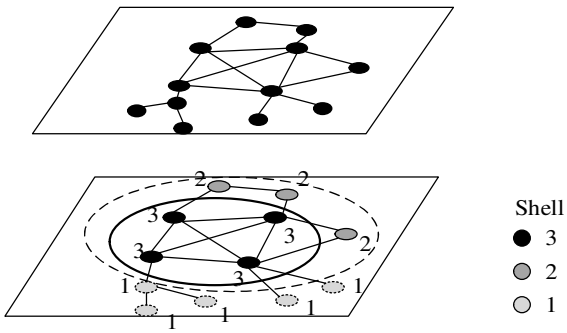


Fig. 2 Decomposing the keyword network based on K (k>K)

1.2.2 Generating a K-core map for landslide monitoring

The process of decomposing the keyword co-occurrence network according to the K-value is shown

in Fig. 3. The K-core subgraph is the union of all shells with k-values greater than or equal to K. According to the K value of each node, the relationship between the node and the co-occurrence matrix of landslide monitoring is assessed, and some nodes can be removed. In this study, we briefly discuss the influence of the proposed method and the high-frequency nodes on the community structure detection algorithm applied to the landslide monitoring co-occurrence network. For networks with the same amount of node information and fewer edge connections than k-subgraphs, the proposed method can significantly reduce the run time while ensuring high quality.

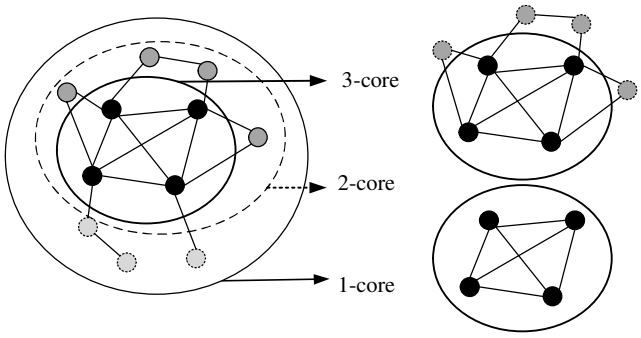


Fig. 3 The process of generating K-subnets by pruning

1.3 Community topic hierarchy and fine-scale knowledge discovery

1.3.1 Knowledge detection among landslide monitoring communities

Communities are characterized by very close relationships among internal nodes and relatively sparse relationships with other communities. Therefore, communities in landslide monitoring keyword co-occurrence networks represent a collection of closely related words with the same cognitive structure related to the same topic. Based on the Louvain algorithm, this paper studies community division and topic detection for landslide monitoring keyword co-occurrence networks. The objective of the algorithm is to first treat a single node as a community and then continuously move the nodes among communities to increase the Q value of the modularity function (Blondel et al. 2008). In the iterative process of the Louvain algorithm, the most time-consuming step is to divide a single node into communities (i.e., the first stage). Therefore, the K-core algorithm is needed to prune and retain the main community structure. After pruning, the process of knowledge discovery based on the corresponding landslide monitoring co-occurrence network is as follows.

The first stage involves calculating the modularity Q according to the input node and edge set. The calculation for initial modularity is shown in Eq. 4. Each key node in the network is regarded as an independent community, and the weight of a community and the weighted sum of the connecting edges of the nodes inside the community are calculated. In the second stage, the change in modularity is calculated, and this value is used to adjust the community ownership of nodes. Additionally, the threshold t is determined according to the degree of network analysis. The corresponding formulas are as follows.

$$\Delta Q = \frac{w_{i,in}}{2m} - \frac{\Sigma_{tot} w_i}{2m^2} \tag{2}$$

$$f(x) = \begin{cases} \Delta Q > 0 \\ \Delta Q > t \end{cases} \tag{3}$$

where $w_{i,in}$ is the sum of the edge weights of nodes in the community, m is the number of edges, and w_i is the sum of the weights of all the edges connected to node i . Σtot is the sum of the weights of the links among nodes in the community. If two nodes share an edge, they should be grouped into the same community. Then, the modularity is calculated, and the modularity gain values are compared. If

ΔQ is greater than the threshold, the result is divided into one class; if the modularity result is less than the threshold, no division occurs. The selection of the threshold value should be based on the number of community divisions and the changes in modularity. Finally, a community network with a smaller size than the original is reconstructed, and the community partition state when the Q value is optimal and the modularity value are output. By setting the critical value of network modularity, the degree of internal contact among some communities can be increased.

1.3.2 Evaluation index modularity Q

Modularity is used to measure the effect of community division and is applied in the comparison of algorithms in different fields (Orman et al. 2009; Karimi-Majd et al. 2015; Yuan et al. 2020). Notably, modularity is the difference obtained by subtracting the expected value of the proportion of the edges of keyword nodes in a community for a network with a uniform community structure and that for another network with random vertices. The corresponding calculation is shown in Eq. 4.

$$Q = \frac{1}{2n}\sum_{w_i w_j}[A_{w_i,w_j} - \frac{k_{w_i}k_{w_j}}{2n}]\delta(c_{w_i}, c_{w_j}) \tag{4}$$

where n is the total number of edges in the network, A_{w_i,w_j} represents the weight of an edge between keyword nodes, and k_{w_i} and k_{w_j} denote the total weights of all the edges associated with the two keywords. c_{w_i} is a Boolean function that depends on the keyword nodes in the current community. Generally, the larger the modularity value is, the better the division result. The range of modularity is [-0.5, 1); when this value is between 0.3 and 0.7, the clustering effect is good. Thus, modularity can be used reflect the community division effect for a landslide monitoring keyword co-occurrence network based on K-core decomposition and the corresponding high-frequency co-occurrence network.

Section 2: Experiments and analysis of results

2.1 Data collection and preprocessing

This study uses the Web of Science (http://isiknowledge.com/wos) as a data source and "landslide monitoring" as the subject. The selection period was from 1950 to 2020, and a total of 6212 search results were obtained. The search results were sorted, and newspaper articles, conference notices, book reviews and other irrelevant literature types were removed. A total of 5165 valid literature records were obtained. Then, 12193 keywords were obtained by extracting author keywords, which were used to construct a keyword co-occurrence network. As shown in Table 1, since the total number of co-occurrence relationships between 12193 keywords is 148669249, it is difficult to create a huge data set, and many single-frequency keywords are not associated with other keywords in the co-occurrence relationship set. Therefore, this paper selects 2589 keywords with frequencies greater than or equal to 2 to construct a keyword co-occurrence network for analysis, and a total of 19305 co-occurrence semantic relationships are obtained.

2.2 Experimental environment

The experiment was run and tested on a desktop terminal. The terminal was equipped with an AMD Ryzen 7 CPU @ 2.9 GHz with 16 GB of memory and an NVIDIA GeForce RTX2060 GPU with 8 GB of memory. The software installed on the terminal included a Windows 10 OS, Microsoft Edge, JetBrains PyCharm 5.0.3 and UCI6.

2.3 Analysis of experimental results

2.3.1 Construction of the K-nucleon diagram

Based on the effective literature data set, the co-occurrence frequencies for keywords can be calculated, and the co-occurrence matrix can be created. After K-core analysis, the keyword network was

divided into 25 levels, as shown in Fig. 4. The number of nodes connected to each node is called the node degree, and the average value of all node degrees is called the network average degree, which is used to represent the complexity of the network (Freeman 1979). As shown in Fig. 4, the average degree of the network is approximately 18, which indicates that each node is connected to 18 other nodes on average.

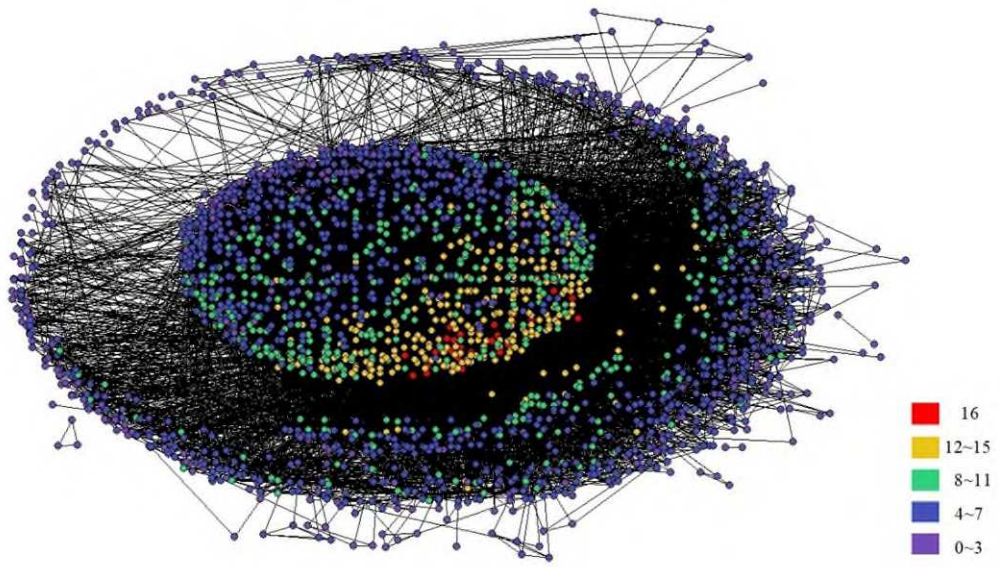


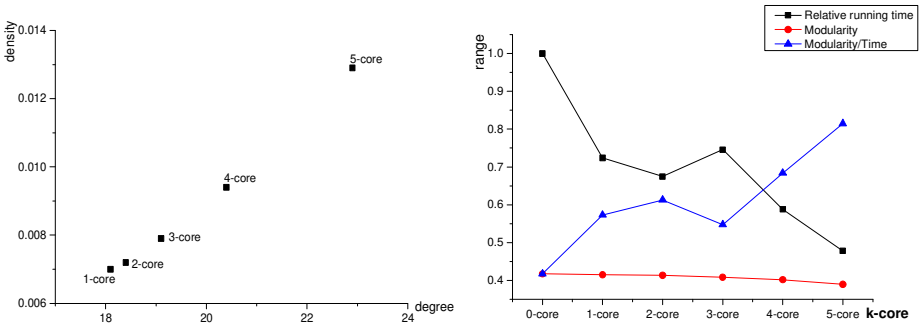
Fig. 4 K-cores of the keyword network of the landslide monitoring field

According to the Eq. 1, the K value is 5.77. Using the above method, nodes with K-values greater than or equal to 5 are selected to construct the keyword co-occurrence network subgraph of landslide monitoring. Shells with K values less than 5 are removed, and the numbers of nodes and connecting edges are shown in Table 1. Compared with the high-frequency keyword network, the new subnetwork considers the strong correlations between nodes. In addition, the K-core decomposition network contains some important keywords with low frequencies, which can be used to comprehensively study landslide monitoring.

Table 1 Changes in network nodes and edges with the K-value

K-value (≥)	Number of keywords	Number of links
0-core	2589	19305
1-core	2582	19262
2-core	2541	19009
3-core	2419	18291
4-core	2180	16955
5-core	1782	15317

The nodes in the K-core subnet are associated with at least k nodes (Kitsak et al 2010). Fig. 5 shows the changes in the density and degree of different K-core graphs. Among them, the relative run time is calculated in reference to the detection time for a network community with a K-value of 0. Notably, as the core value increases, the network degree and density display upward trends, which suggests that increasingly close relations exist between keyword nodes and core content. The run time of the K-core subgraph algorithm decreases with the number of cores used, and the modularity is greater than 0.3, which indicates that the clustering effect is good. When the core value is 5, the modularity/time ratio of the K-core pruning network community is the highest.



(a) Degree and density (b) Modularity and relative running time

Fig. 5 Variations in the network with the K-value

2.3.2 Community theme mining

A community can reflect the closeness among nodes and hierarchical relationships among types of fine-scale knowledge. The 5-core subgraph is selected, and 17 community structures are obtained through community division, with a modularity of 0.3895. The larger the proportion of community nodes is, the richer the knowledge is. The community with the largest proportion of nodes is selected for analysis (Fig. 6). The graph contains 263 nodes, accounting for 14.8% of all nodes, and 1850 edges. The network average degree value is 10.4, the average density is 0.0401, and the node label size is set according to the node degree as the threshold. The figure indicates that the largest network degree values are associated with 'landslide monitoring', 'InSAR', 'deformation', 'interaction', and 'synthetic aperture radar'.

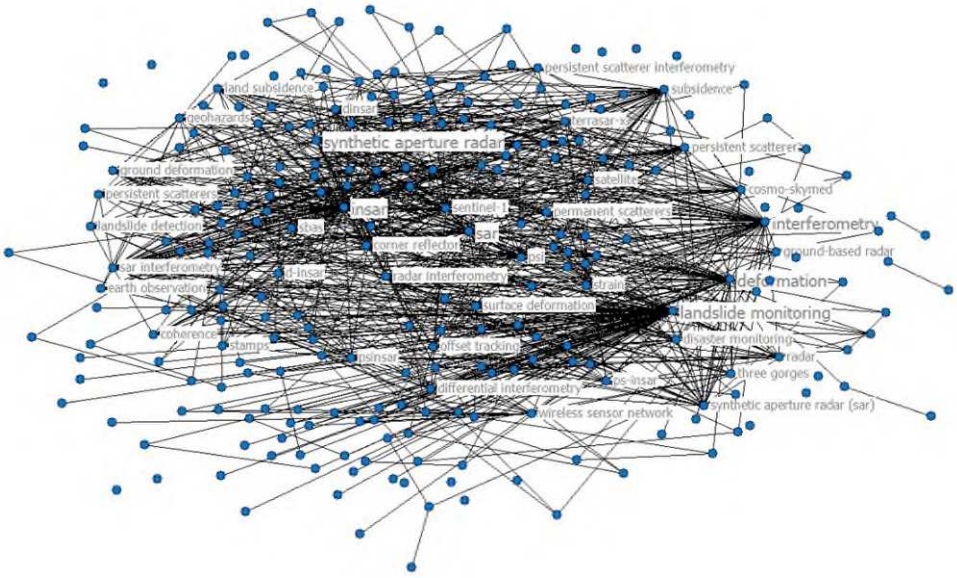


Fig. 6 K-core co-occurrence network subcommunity for landslide monitoring (k ≥ 5)

Ten communities covering 86.5% of all nodes were selected, and the representative keywords of each community were selected according to the frequency or degree, as shown in Table 2. The results for community 1 indicate that landslide monitoring uses 'InSAR' and 'Earth observation' techniques and focuses on 'deformation' and 'offset tracking'. For community 6 and community 8, landslides are related to 'debris flows', 'earthquakes' and 'tsunamis'. Community 4 focuses on the aspects that affect or lead to landslides, such as 'heavy rainfall' and 'rainfall information'. The theme of community 3 is slope engineering and deformation-triggering factors; community 2 is related to the discipline of landslide monitoring and related fields; community 9 focuses on landslide prediction and analysis technology and processes; and community 5 mainly encompasses monitoring instruments. Through community division,

the subject types and fine-scale knowledge associated with landslide monitoring can be clearly obtained.

Table 2 Keywords associated with the landslide monitoring communities (K ≥ 5)

Community	Keywords
1	'landslide monitoring', 'InSAR', 'deformation', 'interferometry', 'synthetic aperture radar', 'persistent scatterers', 'earth observation', 'offset tracking'
4	'slope stability', 'field monitoring', 'heavy rainfall', 'rainfall infiltration'
3	'rainfall', 'numerical simulation', 'stability', 'slope engineering', 'groundwater'
2	'remote sensing', 'lidar', 'risk assessment', 'change detection', 'photogrammetry'
0	'early warning system', 'deformation prediction', 'laser scanning', 'forecast'
6	'debris flow', 'erosion', 'climate change', 'soil moisture', 'permafrost'
9	'landslide prediction', 'machine learning', 'data processing', 'risk analysis'
5	'deformation monitoring', 'inclinometer', 'terrestrial laser scanning'
8	'earthquake', 'tsunami', 'dynamic monitoring', 'volcano', 'outburst flood'
11	'electrical resistivity tomography', 'time series analysis', 'tomography'

Based on the critical value of ΔQ, when the parameter t is greater than 0.00003, the nodes can be split to form more than 17 communities, and the modularity reaches a peak value at 0.000034. Therefore, the threshold is set to 0.000034, and the result of each iteration varies when the modularity of the newly divided community is greater than the threshold value. After community division, 21 community structures are obtained, and the modularity is 0.3807. The community with the largest proportion of nodes was selected as the representative community (Fig. 7) for analysis. The corresponding graph contains 347 nodes, accounting for 19.5% of all nodes, and 2778 edges. The label size is set according to the node degree value. The average network degree value is 11.7, and the average density is 0.0338. The nodes with the largest degree values are 'landslide monitoring', 'InSAR', 'interferometry', 'deformation monitoring' and 'GPS'. Appropriately setting the ΔQ threshold makes the nodes within the community closely connected, which is convenient for analyses of landslide monitoring domain knowledge.

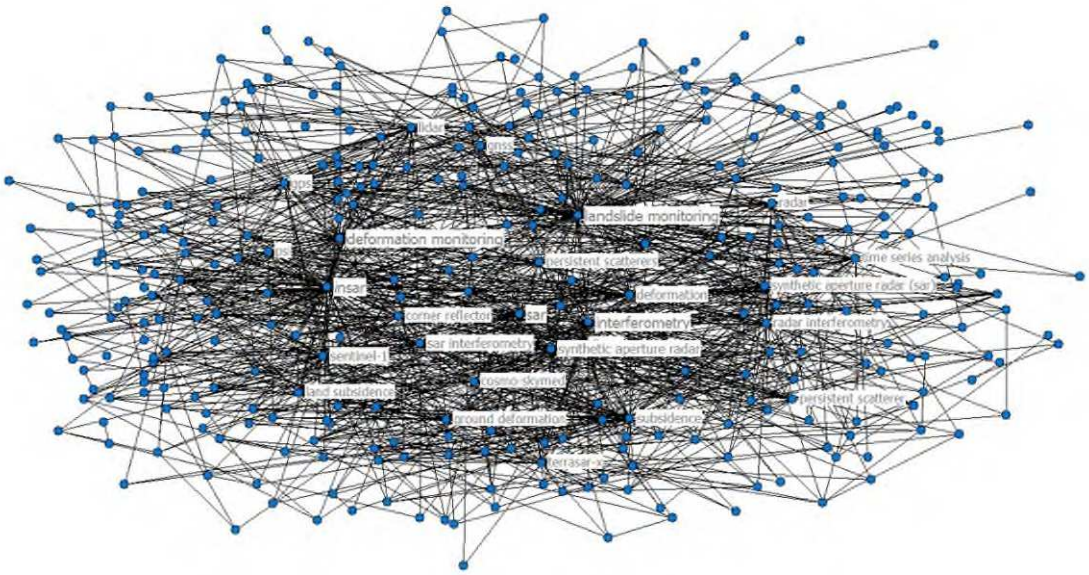


Fig. 7 Five-core co-occurrence network subcommunities in landslide monitoring

2.3.3 Comparative evaluation of methods

The abovementioned community structure detection method is evaluated through the same high-frequency keyword subnet as the 5-core node. After Louvain community division, 18 community structures were obtained, with a modularity of 0.3855. Additionally, the community with the largest proportion of nodes was selected as the representative community (Fig. 8) for analysis. The graph contains 298 nodes, accounting for 16.7% of all nodes, and 2668 edges. The average network degree is 12.7, and the node label size is set according to the node degree as the threshold. The graph shows that the largest values of network degree are associated with 'landscape monitoring', 'InSAR', 'interaction', and 'synthetic aperture radar', and these results are basically consistent with the K-core subgraph results.

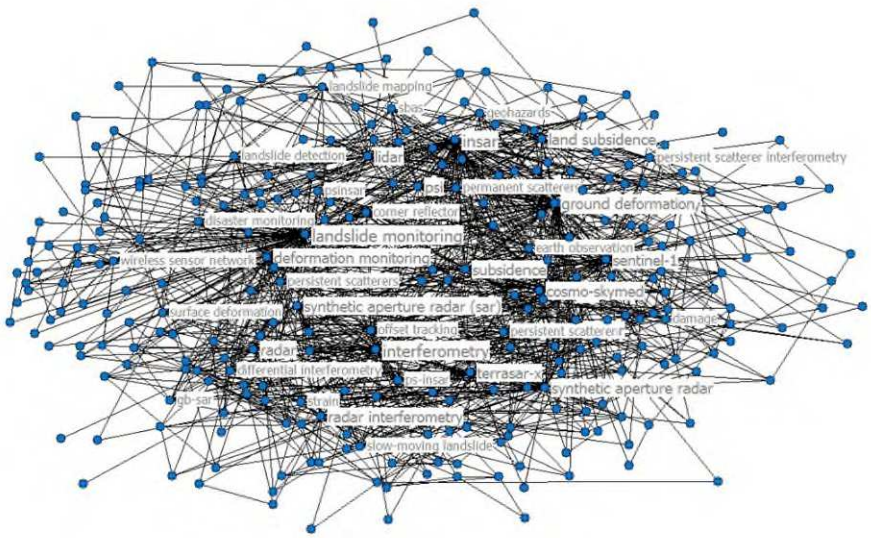


Fig. 8 Subcommunities of the high-frequency co-occurrence network for landslide monitoring

Ten communities encompassing 86.1% of all nodes were selected, and the representative keywords of each community were selected according to the frequency or degree, as shown in Table 3. Most nodes in communities 1, 2, 5 and 6 are the same as those in the K-subnet, which indicates that the keyword structure of the community is closely clustered; the corresponding research topics in landslide monitoring focus on technologies, disciplines, monitoring instruments and related disasters. The theme of community 4 is landslide simulations and modelling, the theme of community 3 is slope engineering and failure mechanisms, and the themes of communities 7 and 8 are monitoring data analysis techniques and methods involved in landslide prediction.

Table 3 Keywords of the landslide monitoring community (K ≥ 5)

Community	Keywords
1	'landslide monitoring', 'InSAR', 'interferometry', 'subsidence', 'deformation', 'synthetic aperture radar', 'sentinel-1','lidar'
2	'remote sensing', 'risk assessment', 'change detection', 'photogrammetry'
4	'slope stability', 'numerical modelling', 'rainfall infiltration', 'pore water pressure'
3	'rainfall', 'numerical simulation', 'stability', 'failure mechanism'
0	'early warning system', 'deformation prediction', 'laser scanning', 'forecast'

6	'debris flow', 'erosion', 'climate change', 'soil moisture', 'permafrost'
12	'slope monitoring', 'fiber Bragg grating', 'geotechnical engineering'
5	'GPS', 'deformation monitoring', 'inclinometer', 'terrestrial laser scanning'
7	'early warning', 'slope failure', 'real-time monitoring', 'data mining'
8	'slope engineering', 'groundwater', 'landslide prediction', 'machine learning'

The results of community detection based on high-frequency keyword pruning and the k-core method were evaluated based on the relative run time and modularity Q value. The relative run time refers to the ratio of the community detection time after pruning to that before pruning. The results shown in Fig. 9 indicate that the overall run time of the K-core pruning method is significantly lower than that of the high-frequency keyword feature selection method; the modularity of the K-core pruning method fluctuates, and that of the K-core pruning method is slightly higher than that of the high-frequency keyword feature selection method. When the core value is 5, the modularity of the K-core pruning network community structure is higher than that of the high-frequency keyword network structure.

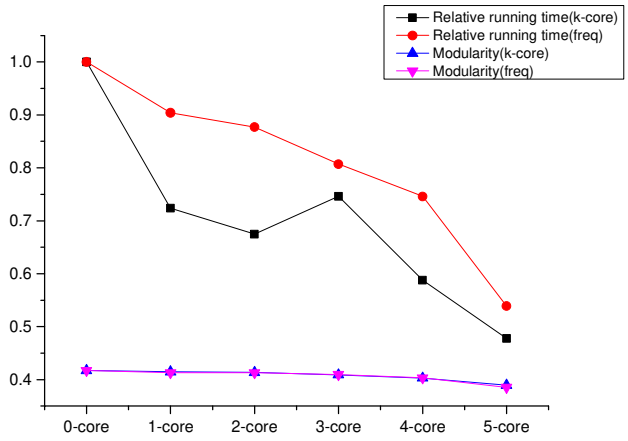


Fig. 9 Relative run time and modularity

Section 3: Conclusion and Prospects

From the perspective of quantitative analysis, we propose a method of knowledge discovery based on keyword co-occurrence network community division. By defining the pruning standard K, the keyword co-occurrence network of landslide monitoring research is simplified, and the degree values and community density characteristics of subcommunities are analysed. Landslide monitoring research focuses on related disciplines, technologies, monitoring instruments and related disasters. In general, the K-core pruning method effectively reduces the run time of the Louvain community partitioning algorithm and retains the relevant community structure. The main contributions of this paper are summarized as follows.

(1) To explore the topic hierarchy and fine-scale knowledge in the landslide monitoring field, the degree value characteristics, subgraph density and community structure of nodes in the keyword co-occurrence network are quantitatively analysed. Compared with existing research, we combine quantitative research with qualitative analysis, reveal the knowledge structure and theme levels of landslide monitoring research, explore new statistical analysis methods for theme discovery, and obtain rigorous and convincing research results.

(2) K-core decomposition is used to generate subgraphs, and the optimal subset is selected by considering the correlations among nodes through the pruning index value; this approach is convenient for analysing the subject-level and fine-scale knowledge in the landslide monitoring field. In the process

of community partitioning, the ΔQ threshold is set according to the resolution degree. During processing, if the modularity value is greater than the threshold, and community division occurs so that the internal nodes of the community are composed closely related topic keywords. Compared with methods in previous studies, such as the high-frequency keyword feature selection method, the proposed method considers the co-occurrence relationships among keyword nodes and the topic structures and fine-scale knowledge in different communities, retains the community structure, and reduces the overall run time.

The threshold t is adjustable and needs to be changed according to the modularity and community division results. In this study, the community division parameters are only applicable to the landslide monitoring co-occurrence network, and further analyses should be performed with other networks. In addition, this study focuses on the exploration and analysis of landslide monitoring at the subject level and fine-scale knowledge discovery methods; some new keywords and topics in the field are worthy of further discussion.

Reference

[1] Whiteley JS, Chambers JE, Uhlemann S et al (2020) Landslide monitoring using seismic refraction tomography – the importance of incorporating topographic variations. Engineering Geology 268:105525 -105551. https://doi.org/ 10.1016/j.enggeo.2020.105525

[2] Xu Y, George DL, Kim J et al (2021) Landslide monitoring and runout hazard assessment by integrating multi-source remote sensing and numerical models: an application to the Gold Basin landslide complex, northern Washington. Landslides 18:1131–1141. https://doi.org/10.1007/s10346-020-01533-0

[3] Lollino P, Giordan D, Allasia P et al (2020) Assessment of post-failure evolution of a large earthflow through field monitoring and numerical modelling. Landslides 17: 2013–2026. https://doi.org/10.1007/s10346-020-01415-5

[4] Schlgel R, Cécile Doubre, Malet JP, Frédéric Masson (2015) Landslide deformation monitoring with alos/palsar imagery: a d-insar geomorphological interpretation method. Geomorphology 231:314-330. https://doi.org/10.1016/j.geomorph.2014.11.031

[5] Solari L, Soldato MD, Raspini F et al (2020) Review of satellite interferometry for landslide detection in italy. Remote Sensing 12:1351-1380. https://doi.org/10.3390/rs12081351

[6] Whiteley JS, Chambers JE, Uhlemann S et al (2019) Geophysical monitoring of moisture-induced landslides: a review. Reviews of Geophysics 57: 106-145. https://doi.org/10.1029/2018RG000603

[7] Aubaud C, Athanase J E, Clouard V et al (2013) A review of historical lahars, floods, and landslides in the precheur river catchment (montagne pelee volcano, martinique island, lesser antilles). Bulletin de la Societe Geologique de France 184 :137-154. https://doi.org/10.2113/gssgfbull.184.1-2.137

[8] Angeli M G, Pasuto A, Silvano S (2010) A critical review of landslide monitoring experiences. Engineering Geology 55:133-147. https://doi.org/10.1016/S0013-7952(99)00122-2

[9] Zhang Y., Hu X, Tannant D D, Zhang G, F Tan (2018) Field monitoring and deformation characteristics of a landslide with piles in the three gorges reservoir area. Landslides 15:581-592. https://doi.org/10.1007/s10346-018-0945-9

[10] Pecoraro G, Calvello M, Piciullo L (2018) Monitoring strategies for local landslide early warning systems. Landslides 16: 213–231. https://doi.org/10.1007/s10346-018-1068-z

[11] Small H (1973) Co-citation in the scientific literature: a new measure of the relationship between two documents. Journal of the American Society for Information Science 24:265 - 269.

https://doi.org/10.1002/asi.4630240406

[12] Forliano C, Bernardi P D, Yahiaoui D (2021) Entrepreneurial universities: a bibliometric analysis within the business and management domains. Technological Forecasting and Social Change 165:120522. https://doi.org/10.1016/j.techfore.2020.120522

[13] Weeds J, Weir D (2005) Co-occurrence retrieval: a flexible framework for lexical distributional similarity. Computational Linguistics 31:439-475. https://doi.org/10.1162/089120105775299122

[14] Kessler MM (1963) Bibliographic coupling between scientific papers. American Documentation 14:10-25. https://doi.org/10.1002/asi.5090140103

[15] Yang J, Liu XJ, Qu Z, Chang H (2020) Retrieval keywords complex networks for analyzing legal complexity of stem cell research. EPL (Europhysics Letters) 130:68001-68008. https://doi.org/10.1209/0295-5075/130/68001

[16] Gan J, Cai QY, Galer P, et al (2019) Mapping the knowledge structure and trends of epilepsy genetics over the past decade: a co-word analysis based on medical subject headings terms. Medicine 98: 16782-16793. https://doi.org/10.1097/MD.0000000000016782

[17] Li F, Zhang S, Wang Y, Li Y, Han Y (2020) Rare fungus, mortierella capitata, promotes crop growth by stimulating primary metabolisms related genes and reshaping rhizosphere bacterial community. Soil Biology and Biochemistry 151:108017-108028. https://doi.org/10.1016/j.soilbio.2020.108017

[18] Fu H, Hu T, Wang JY et al (2015) A bibliometric analysis of malaria research in china during 2004-2014. Malaria Journal 14:195-202. https://doi.org/10.1186/s12936-015-0715-2

[19] Blondel VD, Guillaume J-L, Lambiotte R et al (2008) Fast unfolding of communities in large networks. Journal of Statistical Mechanics Theory & Experiment 2008:1008-1021. https://doi.org/1742-5468/2008/10/P10008

[20] Orman GK, Labatut V, Cherifi H (2011) Qualitative comparison of community detection algorithms. Digital Information and Communication Technology and Its Applications 167:265–279. http: //doi.org/10.1007/978-3-642-22027-2_23

[21] Meo DP, Ferrara, E., Fiumara G, Provetti A (2011) Generalized Louvain method for community detection in large networks. IEEE. http: //doi.org/10.1109/isda.2011.6121636

[22] Daud A, Muhammad F (2012) Group topic modeling for academic knowledge discovery. Applied Intelligence 36:870-886. http: //doi.org/10.1007/s10489-011-0302-3

[23] Xiao L, Chen G, Sun J, Han S, Zhang C (2016) Exploring the topic hierarchy of digital library research in china using keyword networks: a k-core decomposition approach. Scientometrics 108:1-17. http: //doi.org/10.1007/s11192-016-2051-x

[24] Kadi I, Idri A, Fernandez-Aleman JL (2017) Knowledge discovery in cardiology: a systematic literature review. International Journal of Medical Informatics 97:12-32. http: //doi.org/10.1016/j.ijmedinf.2016.09.005

[25] Zhao SX, Zhang PL, Jiang L, Tan AM, Ye FY (2014) Abstracting the core subnet of weighted networks based on link strengths. Journal of the Association for Information Sci & Tech 65:984-994. http: //doi.org/10.1002/asi.23030

[26] Seidman SB (1983) Network structure and minimum degree. Social Networks 5:269-287. http: //doi.org/10.1016/0378-8733(83)90028-x

[27] Zhang GQ, Yang QF, Cheng SQ, Zhou T (2008) Evolution of the Internet and its cores. New Journal of Physics 10:123027-123038. http: //doi.org/10.1088/1367-2630/10/12/123027

[28] Kong YX, Shi GY, Wu RJ, Zhang YC (2019) K -core: theories and applications. Physics Reports 832:1-32. https://doi.org/10.1016/j.physrep.2019.10.004

[29] Kitsak M, Gallos LK, Havlin S et al (2010) Identification of influential spreaders in complex networks. Nature Physics 6:888–893. https://doi.org/10.1038/nphys1746

[30] Orman GK, Labatut V (2009) A comparison of community detection algorithms on artificial networks. Algorithms on Artificial Networks 5805:242-256. https://doi.org/10.1007/978-3-642-04747-3_20

[31] Karimi-Majd AM, Fathian M, Amiri B (2015) A hybrid artificial immune network for detecting communities in complex networks. Computing 97:483-507. https://doi.org/10.1007/s00607-014-0433-6

[32] Yuan C, Rong C, Yao Q (2020) Boundary-connection deletion strategy based method for community detection in complex networks. Applied Intelligence. https://doi.org/10.1007/s10489-020-01762-9 Accessed 18 June 2020

[33] Freeman LC (1979) Centrality in Social Networks: Conceptual clarification. Social Networks 1:215-239. https://doi.org/10.1016/0378-8733(78)90021-7