

Article

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

Ping Wang <sup>1</sup> , Xingdong Deng <sup>2,3,\*</sup>, Yang Liu <sup>2,3</sup>, Liang Guo <sup>2,3</sup>, Jun Zhu <sup>1</sup>, Lin Fu <sup>1</sup>, Yakun Xie <sup>1</sup>, Weilian Li <sup>1</sup> and Jianbo Lai <sup>1</sup>

- <sup>1</sup> Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610031, China; wp@my.swjtu.edu.cn (P.W.); zhujun@swjtu.edu.cn (J.Z.); vge\_fulin@my.swjtu.edu.cn (L.F.); yakunxie@my.swjtu.edu.cn (Y.X.); vgewilliam@my.swjtu.edu.cn (W.L.); ljb2244@my.swjtu.edu.cn (J.L.)
- <sup>2</sup> Guangzhou Urban Planning and Design Survey Research Institute, Guangzhou 510060, China; liuyang@gzpi.com.cn (Y.L.); guoliang@gzpi.com.cn (L.G.)
- <sup>3</sup> Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning, Guangzhou 510060, China
- \* Correspondence: dengxingdong@gzpi.com.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 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 K-core decomposition and Louvain algorithms. In this method, author keywords are used as nodes to construct a weighted co-occurrence network, and a pruning standard value is defined as 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. The results show that the run time of the proposed method is significantly less than that of the traditional method.

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



**Citation:** Wang, P.; Deng, X.; Liu, Y.; Guo, L.; Zhu, J.; Fu, L.; Xie, Y.; Li, W.; Lai, J. A Knowledge Discovery Method for Landslide Monitoring Based on K-Core Decomposition and the Louvain Algorithm. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 217. <https://doi.org/10.3390/ijgi11040217>

Academic Editor: Wolfgang Kainz

Received: 10 January 2022

Accepted: 21 March 2022

Published: 22 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Landslide monitoring, as an important method of disaster prevention and reduction, has attracted extensive attention from academia in recent decades [1–3]. Today, it has become an important branch of natural disaster assessment and prevention [4–7]. There are many studies on landslide monitoring. The discovery and summary of subject levels and fine-scale knowledge, such as research objects, key technologies and disaster-causing factors in landslide monitoring, are helpful in understanding the research status and hot spots in landslide monitoring. Additionally, a reference for scientific analyses; disaster prevention and mitigation; and disaster monitoring is provided.

To analyze the knowledge on domain topics, scholars have attempted to depict their internal structures [8–10]. In these studies, keyword co-occurrence networks can quantitatively reflect the development process and structural relationship of scientific knowledge through community structure and network centrality trends [11–14]. The Louvain algorithm is a community mining algorithm based on modularity and is suitable for the

partitioning of small and medium-sized networks [15]. However, its computing cost is high [16–18]. Rich text semantic relations can produce dense topics for knowledge discovery [19]. For some networks with a small number 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 [20–23] have generally set thresholds to screen keywords according to the word frequency or edge weights, but these methods do not consider the possible effect of semantic association between two keywords. Seidman [24] 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 [25–31]. 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 knowledge discovery method for landslide monitoring based on K-core decomposition and the Louvain algorithm. The core co-occurrence relationship is decomposed by a k-core, and the dense community structure is efficiently detected with the Louvain algorithm. The remainder of this paper is organized as follows. Section 2 introduces related works, including knowledge discovery in the landslide field and network community knowledge discovery. In Section 3, the methods, including the overall research concept, are introduced, and subgraph extraction and the community detection process are discussed. Section 4 provides an analysis of the experimental results, and the data sources and experimental environment are introduced; additionally, a comparison of methods is performed. Section 5 discusses the study's conclusions and future research prospects.

## 2. Related Work

### 2.1. Landslide Domain Knowledge Discovery

Knowledge discovery finds credible and useful knowledge from data and is presented in an easy-to-understand method or pattern, first in the 1980s. Data mining is one of its steps. Common methods include association, classification and sequencing. With the development and application of machine learning and knowledge maps, the method and connotation of knowledge discovery have been further expanded and have become an important data-driven discipline knowledge discovery research topic. Currently, there are three common landslide monitoring knowledge discovery methods, including summary, data investigation and statistical analysis of the literature in landslide monitoring.

The first method is an overview of the landslide monitoring field. This method is a landslide risk assessment method based on expert knowledge. The literature summary method reviews and analyzes landslide monitoring objects, monitoring equipment, monitoring methods and technologies and related research from a qualitative point of view. Guzzetti et al. [32] reviewed the experience in the design, implementation, management and verification of landslide early warning systems. Whiteley et al. [9] evaluated the latest technology applied to geophysical monitoring of humidity-induced landslides. Michel et al. [33] reviewed the different lidar applications in landslides, rockfalls and debris flows. Chae et al. [34] reviewed the latest progress and methods of the basic components of landslide disaster assessment.

The second method is data research, summarizing and discovering knowledge. Phengsuwan et al. [35] proposed a knowledge discovery method for the relationship between landslide disasters and earth observations and urban data sources. Sufi et al. [36] used machine learning to reveal the hidden knowledge of a series of complex scenes created by the attributes of five landslide elements. Data investigation is used to monitor landslide site and deformation characteristics by using equipment monitoring. Angeli et al. [37] discussed some of the main issues in the installation and management of monitoring equipment used to study landslides. Zhang et al. [38] provided basic data by monitoring

the long-term performance of anti-skid piles in the system and evaluating the long-term stability of stable landslides under reservoir operation.

The third method is knowledge discovery based on statistical characteristics. The statistical analysis method summarizes the landslide early warning system monitoring strategy based on statistical results in the literature. Aydinoglu et al. [39] used statistics and machine learning technology to study landslide-prone areas. Solari et al. [8] analyzed its application in monitoring from 250 papers related to InSAR in Italian landslide research. In recent years, there has been an increasing number of quantitative analyses of landslide monitoring research topics and fine-scale knowledge from the perspective of statistics in the literature. Among them, the co-occurrence network provides a demonstrable quantitative form, which can mine knowledge from the aspects of network nature and community structure. In recent years, keyword co-occurrence networks have been widely used in various fields, such as exploring the knowledge structure and progress of epilepsy genetics, stem cell research and malaria [40–42].

## 2.2. Network Community Knowledge Discovery

Community structure is an important network feature. The entire network is composed of many community structures. The community has the characteristics of close connections between internal nodes and sparse connections between communities. As a collection of individuals with the same attributes, community structure plays an important role in network analysis. Landslide subject knowledge discovery can be roughly divided into social networks, co-occurrence networks and literature knowledge maps according to different network contents.

Social networks are used to deliver news and some valuable information. In the context of Big Data, they can also be used to monitor adverse events. Meng et al. [43] used Twitter data to establish a highly reliable disaster information monitoring system. Musaev et al. [44] introduced the concept of a virtual community to establish a comprehensive disaster information system of landslide social networks. Shehara et al. [45] modeled drought, flood, landslide, tsunami and cyclone disaster information and ranked them according to centrality to highlight the key nodes in the network.

Another method of network knowledge discovery is to generate a co-occurrence network through the literature. A document knowledge atlas is essentially a special co-occurrence network. Zhang et al. [46] constructed a recommendation system among users, scenes and data of landslide disaster environments based on knowledge maps and deep neural networks. Gizzi et al. [47] constructed a co-occurrence network based on the literature title data extracted from Scopus, Web of Science and other databases and reviewed the research in the last 40 years. Yong et al. [48] used CiteSpace to analyze papers with landslide susceptibility as the theme word in the Web of Science database from 1991 to 2020 and revealed the current situation of fields including scientific achievements, research community structure and international cooperation.

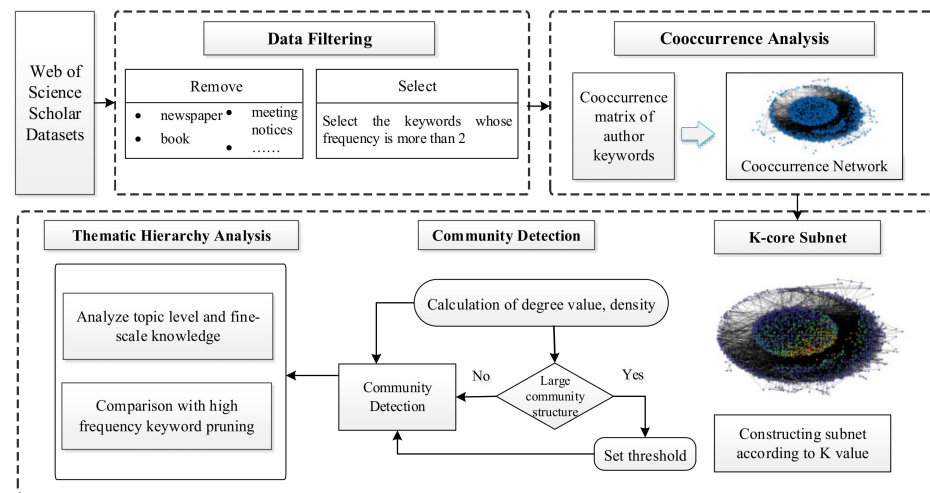
Under the background of landslide disasters, network analysis focuses on the current situation of relevant fields and the theme of disaster monitoring information systems. It can be observed that effective monitoring and analysis require a large quantity of relevant data as support. As a common method, a co-occurrence network is used not only for relational knowledge modeling but also for analyzing and managing domain knowledge. With the rise and application of Big Data and knowledge maps, the connotation and extension of co-occurrence networks have been further expanded, and relevant network analysis has become an important research topic of knowledge discovery.

## 3. Method

### 3.1. Overall Research Concept

The technical route of knowledge discovery in landslide monitoring is shown in Figure 1. The Web of Science dataset is preprocessed by using data filtering to reduce invalid and noisy data. 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 analyzed, and the threshold value of  $\Delta Q$  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 analyze 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.



**Figure 1.** Technical research route.

### 3.2. Construction of the K-Sub Map of the Cooccurrence Network of Landslide Monitoring

#### 3.2.1. Calculation of the Pruning Standard Based on the K-Core

After constructing the co-occurrence network, due to the large number of network nodes, it is difficult to clearly display knowledge and identify and extract information at the theme level in landslide monitoring. Additionally, the Louvain algorithm is characterized by high complexity when detecting network community structures; thus, it is necessary to prune the network. Before pruning the network, the density and degree distributions of the network need to be calculated. For degree distribution, it is necessary to calculate the degree of correlation between each keyword and all other nodes. In a semantic undirected graph with  $n$  keyword nodes, the degree centrality of a node keyword is the total number of direct correlations between this keyword and  $n-1$  keyword nodes other than itself. The degree distribution  $d_k$  is the proportion of nodes with network degree  $k$ . Generally,  $\sum_{k=0}^n d_k = 1$ . Network density is used to describe the density of keyword associations in the network. In graph  $G$  with  $n$  nodes and  $m$  edges, the network density is as follows.

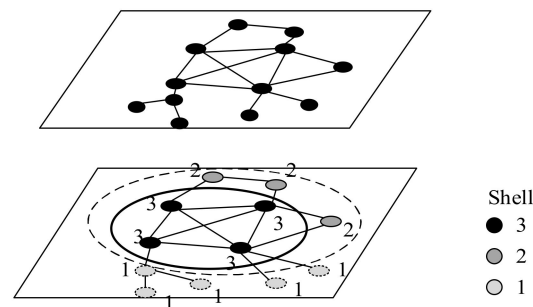
$$L(G) = \frac{2m}{n(n-1)} \quad (1)$$

Compared to degree-irrelevant networks, the degree distribution and density of keyword co-occurrence networks are more complex; thus, the  $k$ -core pruning standard needs to be considered differently. In graph  $G$ , which is composed of keyword nodes and edges of keyword co-occurrence relationships, kernel number  $k$  represents the degree of keywords in the network, and any two vertices in the same  $k$ -core subgraph are connected by  $k$  edges. 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 Equation (2):

$$K = \frac{\sum_i k_i n_i}{m} \quad (2)$$

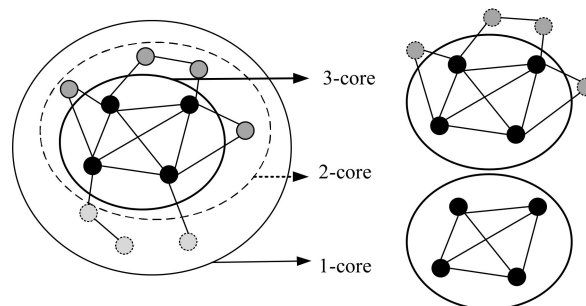
where  $k_i$  represents the K value of each shell,  $n_i$  is the number of shells,  $M$  is the total number of nodes and  $i$  is the shell for each  $k$  value. When the value of node  $k$  is less than  $K$ , some of the nodes can be deleted; otherwise, all nodes should be reserved. As shown in Figure 2, the network consists of three shells that contain 12 nodes. Equation (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.



**Figure 2.** Decomposing the keyword network based on  $K$  ( $k > K$ ).

### 3.2.2. 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 Figure 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 cooccurrence 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 run time while ensuring high quality results.



**Figure 3.** Process of generating K-subnets by pruning.

## 3.3. Community Topic Hierarchy and Fine-Scale Knowledge Discovery

### 3.3.1. 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 [15]. 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 modularity Q according to the input node and edge set. The calculation for initial modularity is shown in Equation (5). Each key node in the network is regarded as an independent community, and the weight of a community and the weighted sum of the connected 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, threshold t is determined according to the degree of network analysis. The corresponding formulas are described as follows:

$$\Delta Q = \frac{w_{i,in}}{2m} - \frac{\sum_{tot} w_i}{2m^2} \quad (3)$$

$$f(x) = \begin{cases} \Delta Q > 0 \\ \Delta Q > t \end{cases} \quad (4)$$

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 edges connected to node i.  $\sum_{tot}$  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, modularity is calculated, and the modularity gain values are compared. If  $\Delta Q$  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 that is smaller than the original is reconstructed, and the community partition states when the Q value is optimal and when the modularity values are outputted. By setting the critical value of network modularity, the degree of internal contact among some communities can be increased.

### 3.3.2. 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 [31,49,50]. 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 Equation (5):

$$Q = \frac{1}{2n} \sum_{w_i, w_j} \left[ A_{w_i, w_j} - \frac{k_{w_i} k_{w_j}}{2n} \right] \delta(c_{w_i}, c_{w_j}) \quad (5)$$

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, 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 to 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.

## 4. Experiments and Analysis of Results

### 4.1. Data Collection and Preprocessing

This study uses the Web of Science (<http://isiknowledge.com/wos> (accessed on 1 January 2021)) 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, 12,193 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 12,193 keywords is 148,669,249, it is difficult to create a large dataset, 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 19,305 co-occurrence semantic relationships are obtained.

**Table 1.** Cited frequency, betweenness and degree ranking of keywords in landslide monitoring research (1950–2020).

No.	Citation Frequency	Keywords	Betweenness	Keywords	Degree	Keywords
1	187	‘remote sensing’	118,378.172	‘remote sensing’	328	‘remote sensing’
2	155	‘InSAR’	94,352.602	‘rainfall’	260	‘rainfall’
3	147	‘rainfall’	93,265.320	‘slope stability’	250	‘slope stability’
4	143	‘slope stability’	61,160.973	‘debris flow’	228	‘InSAR’
5	129	‘GPS’	52,054.520	‘InSAR’	190	‘debris flow’
6	112	‘debris flow’	48,070.133	‘deformation’	158	‘GPS’
7	109	‘deformation monitoring’	44,625.008	‘slope engineering’	131	‘earthquake’
8	99	‘early warning’	42,595.09	‘GIS’	121	‘numerical simulation’
9	99	‘GIS’	40,889.879	‘early warning’	113	‘field monitoring’

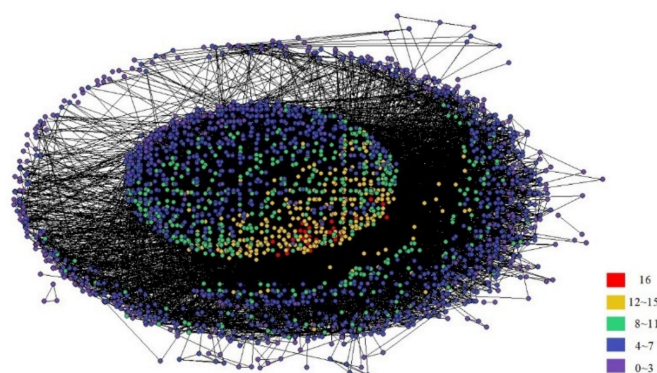
### 4.2. Experimental Environment

The experiment was run and tested on a desktop terminal. The terminal was equipped with an Lenovo 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 Windows 10 OS, Microsoft Edge, JetBrains PyCharm 5.0.3 and UCI6.

### 4.3. Analysis of Experimental Results

#### 4.3.1. Construction of the K-Nucleon Diagram

Based on the effective literature dataset, 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 Figure 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 [51]. The average degree of the network is approximately 18, which indicates that each node is connected to 18 other nodes on average.



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

The overall central trend of the network can be measured by statistical indexes such as citation frequency, betweenness and degree. Among them, an analysis of keywords with high betweenness can obtain research hotspots in the field. The citation frequency, betweenness and degree value ranking of keywords in landslide monitoring research from 1950 to 2020 are shown in Table 1. Generally, the disciplines mainly involved in landslide monitoring are ‘remote sensing’, ‘GIS’ and ‘geomorphology’; the technologies used are ‘GPS’, ‘InSAR’, ‘numerical simulation’ and ‘field monitoring’; and landslide monitoring mainly focuses on ‘slope stability’, ‘deformation’ and ‘early warning’. In addition, ‘rainfall’, ‘debris flow’ and ‘earthquake’ have also received extensive attention as the main factors causing landslides.

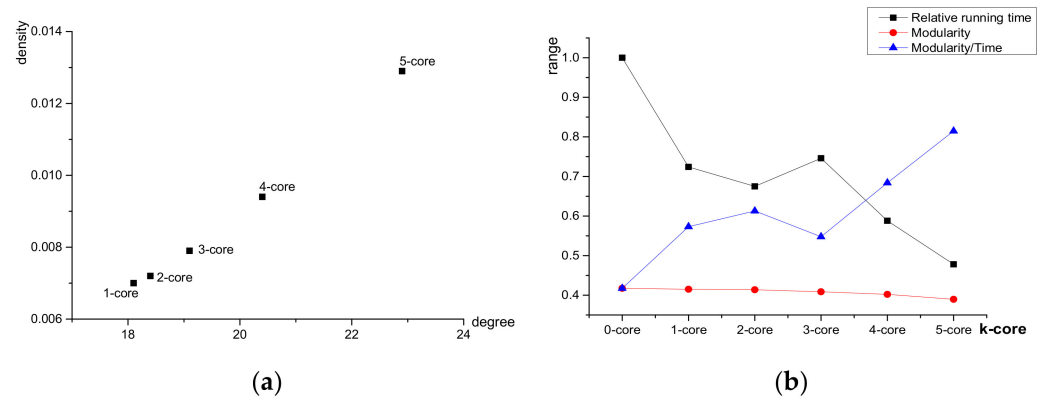
According to the Equation (2), the K value is 5.77. Using the above method, nodes with K-values greater than or equal to 5 are selected to construct the keyword co-occurrence network subgraph of landslide monitoring. Shells with K-values less than 5 are removed, and the numbers of nodes and connecting edges are shown in Table 2. 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 2.** Changes in network nodes and edges with K-value.

K-Value ( $\geq$ )	Number of Keywords	Number of Links
0-core	2589	19,305
1-core	2582	19,262
2-core	2541	19,009
3-core	2419	18,291
4-core	2180	16,955
5-core	1782	15,317

The nodes in the K-core subnet are associated with at least k nodes [29]. Figure 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 upwards 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.

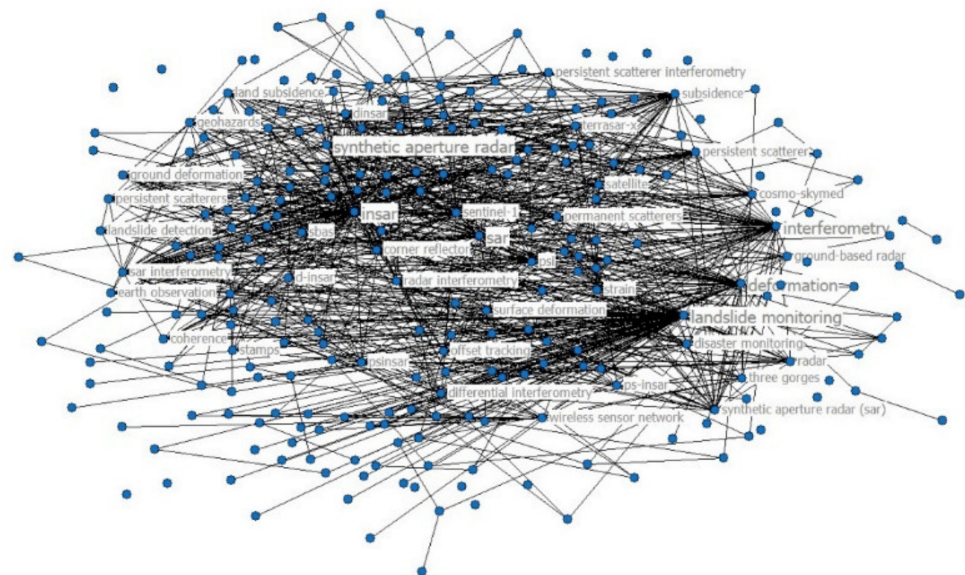




**Figure 5.** Variations in the network with the K-value. (a) Degree and density. (b) Modularity and relative running time.

#### 4.3.2. Community Theme Mining

A community can reflect the closeness among nodes and hierarchical relationships among types of fine-scale knowledge. The five-core subgraph is selected, and 17 community structures were obtained by using community division, with a modularity of 0.3895. The larger the proportion of community nodes, the richer the knowledge. The community with the largest proportion of nodes is selected for analysis (Figure 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 using 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’.



**Figure 6.** K-core co-occurrence network subcommunity for landslide monitoring ( $k \geq 5$ ).

Ten communities covering 86.5% of all nodes were selected, and the representative keywords of each community were selected according to their frequency or degree, as shown in Table 3. The results for community 1 indicate that landslide monitoring uses ‘InSAR’, ‘Earth observation’ and ‘offset tracking’ techniques and focuses on ‘deformation’. For communities 6 and 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 risk analysis, mainly using machine learning; and community 5 mainly encompasses the application of ‘electrical resistance tomography’ in time series analysis. By conducting community division, the subject types and fine-scale knowledge associated with landslide monitoring can be clearly obtained.

**Table 3.** Keywords associated with the landslide monitoring communities ( $K \geq 5$ ).

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  $\Delta Q$ , when parameter  $t$  is greater than 0.00003, the nodes can be split to form more than 17 communities, and 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 were obtained, and the modularity was 0.3807. Appropriately setting the  $\Delta Q$  threshold makes the nodes within the community closely connected, which is convenient for analyses of landslide monitoring domain knowledge.

The changes in the betweenness value of the subject keywords over the years reflect the changes in the influence of the research direction represented by the keywords in the discipline. The data downloaded from the Web of Science have time identification; thus, the literature is grouped by year. Then, the keyword betweenness in the literature in the last ten years is calculated, and the bar chart of keyword betweenness evolution is made, as shown in Figure 7. The research hotspots in landslide monitoring are ‘D-InSAR’ and ‘machine learning’, of which two peaks appear in 2017 and 2019; ‘earthquake’, ‘debris flow’ and ‘rainfall’ have a high influence on landslide monitoring over the years; and ‘numerical simulation’ and ‘rainfall’ grow almost synchronously, which is consistent with the association found by the subject community. In addition, we found that the centrality of ‘debris flow’ and ‘tertiary laser scanning’ also increased linearly, indicating a correlation between research topics.

#### 4.3.3. Comparative Evaluation of Methods

The abovementioned community structure detection method is evaluated using the same high-frequency keyword subnet as the k-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 (Figure 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 using the node degree as the threshold. The graph shows that the largest network degree values are associated with ‘landscape monitoring’, ‘InSAR’, ‘interaction’ and ‘synthetic aperture radar’, and these results are basically consistent with the K-core subgraph’s results.

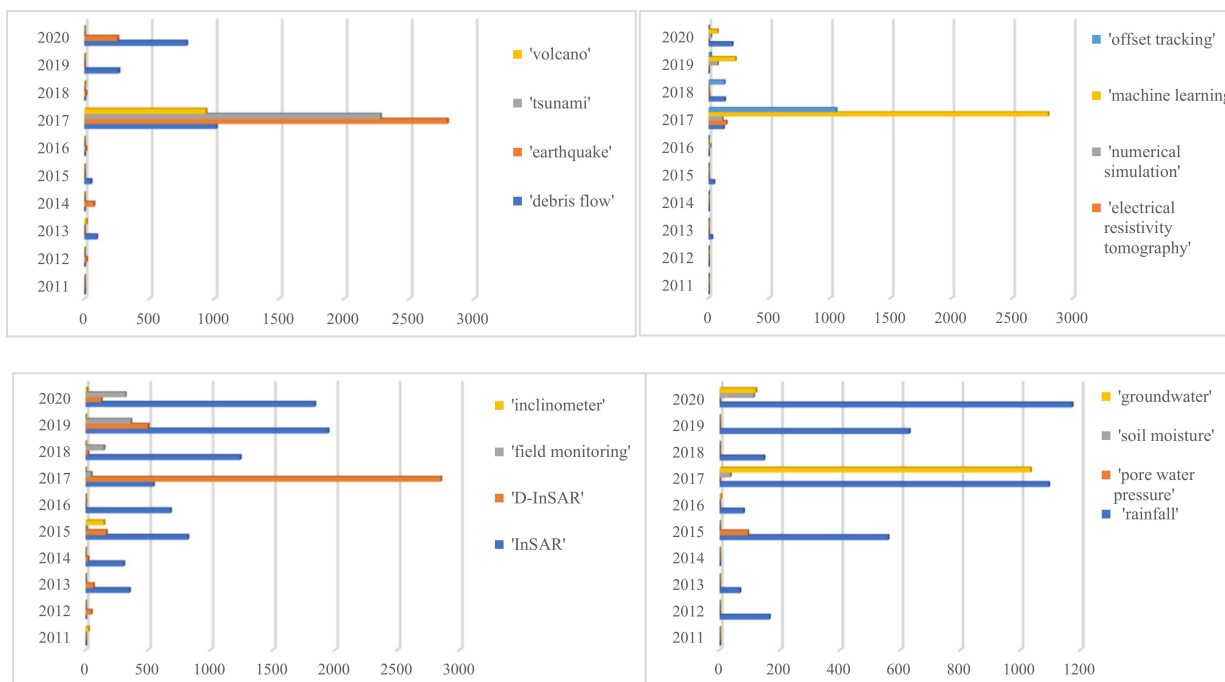


Figure 7. Betweenness evolution of subject keywords.

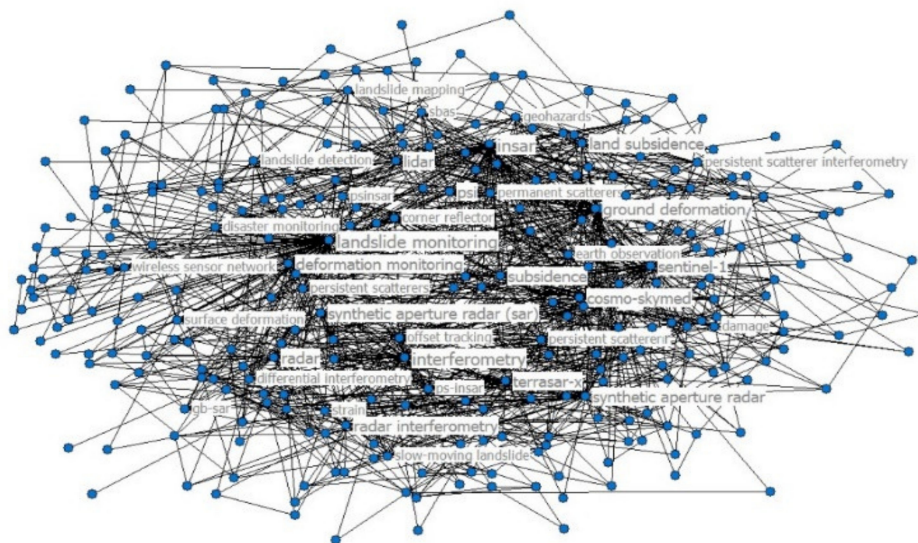


Figure 8. Subcommunities of the high-frequency co-occurrence network for landslide monitoring.

To quantitatively analyze our proposed method, we use two indicators, the network modularity  $Q$  value and relative running time, to evaluate the results, as shown in Table 4. The relative run time refers to the ratio of the community detection time after pruning to that before pruning. Compared with the existing high-frequency keyword pruning methods, the modularity of our method is greater than or equal to that of the high-frequency keyword pruning network, and its relative running time is significantly lower. Therefore, our method significantly reduces the run time while ensuring the effect of network partitioning. 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.

**Table 4.** Relative run time and modularity.

K-Core	Our Method		High-Frequency Keyword Pruning	
	Modularity	Time	Modularity	Time
0-core	0.417	1	0.417	1
1-core	0.414	0.724	0.413	0.904
2-core	0.413	0.675	0.413	0.877
3-core	0.409	0.746	0.409	0.807
4-core	0.403	0.588	0.403	0.746
5-core	0.389	0.478	0.385	0.539

## 5. Conclusions 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 analyzed. Our multilevel analyses determined the research theme, emerging technologies, related disasters and their correlation with landslide monitoring. The experimental results reveal that our method effectively reduces the run time of the Louvain community partitioning algorithm while ensuring network partition results. The main contributions of this paper are summarized as follows.

1. To explore topic hierarchy and fine-scale knowledge in the landslide monitoring field, the degree value characteristics, subgraph density, betweenness and community structure of nodes in the keyword co-occurrence network are quantitatively analyzed. Using time series to analyze the central changes in keywords, the hot spots in landslide monitoring are identified. Compared with existing research, we quantitatively reveal the subject structure, research status and hot spots of landslide monitoring by using the central trend of the co-occurrence network and community structure and obtained 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 analyzing the subject-level and fine-scale knowledge in the landslide monitoring field. In the process of community partitioning, the  $\Delta Q$  threshold is set according to the resolution's degree. During processing, if the modularity value is greater than the threshold, community division occurs so that the internal nodes of the community are composed of 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.

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 for 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.

**Author Contributions:** Conceptualization, Ping Wang and Xingdong Deng; methodology, Ping Wang, Xingdong Deng, Yang Liu, Liang Guo, Weilian Li, Lin Fu and Yakun Xie; validation, Jun Zhu, Jianbo Lai and Yakun Xie; funding acquisition, Xingdong Deng and Jun Zhu. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper was supported by the Key-Area Research and Development Programme of Guangdong Province (Grant Nos. 2020B121202019) and the Sichuan Science and Technology Programme (Grant No. 2020JDTD0003).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are openly available in Web of Science at <http://isiknowledge.com/wos> (accessed on 1 January 2021).

**Conflicts of Interest:** The authors have no conflicts of interest to declare that are relevant to the content of this article.

## References

- Whiteley, J.S.; Chambers, J.E.; Uhlemann, S.; Boyd, J.; Cimpoiasu, M.O.; Holmes, J.L.; Inauen, C.M.; Watlet, A.; Hawley-Sibbett, L.R.; Sujitapan, C.; et al. Landslide monitoring using seismic refraction tomography—The importance of incorporating topographic variations. *Eng. Geol.* **2020**, *268*, 105525–105551. [[CrossRef](#)]
- Lollino, P.; Giordan, D.; Allasia, P.; Fazio, N.L.; Perrotti, M.; Cafaro, F. Assessment of post-failure evolution of a large earthflow through field monitoring and numerical modelling. *Landslides* **2020**, *17*, 2013–2026. [[CrossRef](#)]
- Hou, W.; Lu, X.; Wu, P.; Xue, A.; Li, L. An Integrated Approach for Monitoring and Information Management of the Guanling Landslide (China). *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 79. [[CrossRef](#)]
- Li, W.; Zhu, J.; Zhang, Y.; Fu, L.; Gong, Y.; Hu, Y.; Cao, Y. An on-demand construction method of disaster scenes for multilevel users. *Nat. Hazards* **2020**, *101*, 409–428. [[CrossRef](#)]
- Del Soldato, M.; Solari, L.; Raspini, F.; Bianchini, S.; Ciampalini, A.; Montalti, R.; Ferretti, A.; Pellegrineschi, V.; Casagli, N. Monitoring Ground Instabilities Using SAR Satellite Data: A Practical Approach. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 307. [[CrossRef](#)]
- Huang, Q.; Wang, Y.; Xu, J.; Nishyirimbere, A.; Li, Z. Geo-Hazard Detection and Monitoring Using SAR and Optical Images in a Snow-Covered Area: The Menyuan (China) Test Site. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 293. [[CrossRef](#)]
- Xu, Y.; George, D.L.; Kim, J.; Lu, Z.; Riley, M.; Griffin, T.; de la Fuente, J. Landslide monitoring and runout hazard assessment by integrating multi-source remote sensing and numerical models: An application to the Gold Basin landslide complex. *Landslides* **2021**, *18*, 1131–1141. [[CrossRef](#)]
- Solari, L.; Del Soldato, M.; Raspini, F.; Barra, A.; Bianchini, S.; Confuorto, P.; Casagli, N.; Crosetto, M. Review of satellite interferometry for landslide detection in Italy. *Remote Sens.* **2020**, *12*, 1351. [[CrossRef](#)]
- Whiteley, J.S.; Chambers, J.E.; Uhlemann, S.; Wilkinson, P.B.; Kendall, J.M. Geophysical monitoring of moisture-induced landslides: A review. *Rev. Geophys.* **2019**, *57*, 106–145. [[CrossRef](#)]
- Aubaud, C.; Athanase, J.E.; Clouard, V.; Barras, A.V.; Sedan, O.A. A review of historical lahars, floods, and landslides in the precheur river catchment. *Bull. Soc. Geol. Fr.* **2019**, *184*, 137–154. [[CrossRef](#)]
- Small, H. Co-citation in the scientific literature: A new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci.* **1973**, *24*, 265–269. [[CrossRef](#)]
- Forliano, C.; Bernardi, P.D.; Yahiaoui, D. Entrepreneurial universities: A bibliometric analysis within the business and management domains. *Technol. Forecast. Soc. Chang.* **2021**, *165*, 120522. [[CrossRef](#)]
- Niu, J.; Tang, W.; Xu, F.; Zhou, X.; Song, Y. Global Research on Artificial Intelligence from 1990–2014: Spatially-Explicit Bibliometric Analysis. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 66. [[CrossRef](#)]
- Kessler, M.M. Bibliographic coupling between scientific papers. *Am. Doc.* **1963**, *14*, 10–25. [[CrossRef](#)]
- Zhang, J.; Ma, Z.; Sun, Q.; Yan, J. Research Review on Algorithms of Community Detection in Complex Networks. *J. Phys. Conf. Ser.* **2018**, *1069*, 012124. [[CrossRef](#)]
- Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory Exp.* **2008**, *2008*, 1008–1021. [[CrossRef](#)]
- Orman, G.K.; Labatut, V.; Cherifi, H. Qualitative comparison of community detection algorithms. *Digit. Inf. Commun. Technol. Its Appl.* **2011**, *167*, 265–279.
- De Meo, P.; Ferrara, E.; Fiumara, G.; Provetti, A. Generalized Louvain method for community detection in large networks. In Proceedings of the 2011 11th International Conference on Intelligent Systems Design and Applications, Córdoba, Spain, 22–24 November 2011.
- Daud, A.; Muhammad, F. Group topic modeling for academic knowledge discovery. *Appl. Intell.* **2012**, *36*, 870–886. [[CrossRef](#)]
- Dai, S.; Duan, X.; Zhang, W. Knowledge map of environmental crisis management based on keywords network and co-word analysis, 2005–2018. *J. Clean. Prod.* **2020**, *262*, 121168. [[CrossRef](#)]
- Li, Q.; Zhang, H.; Hong, X. Knowledge structure of technology licensing based on co-keywords network: A review and future directions. *Int. Rev. Econ. Financ.* **2020**, *66*, 154–165. [[CrossRef](#)]
- Xiao, L.; Chen, G.; Sun, J.; Han, S.; Zhang, C. Exploring the topic hierarchy of digital library research in China using keyword networks: A k-core decomposition approach. *Scientometrics* **2016**, *108*, 1085–1101. [[CrossRef](#)]
- Zhao, S.X.; Zhang, P.L.; Jiang, L.; Tan, A.M.; Ye, F.Y. Abstracting the core subnet of weighted networks based on link strengths. *J. Assoc. Inf. Sci. Tech.* **2014**, *65*, 984–994. [[CrossRef](#)]
- Seidman, S.B. Network structure and minimum degree. *Soc. Netw.* **1983**, *5*, 269–287. [[CrossRef](#)]



25. Ai, J.; Liu, Y.; Su, Z.; Zhao, F.; Peng, D. K-core decomposition in recommender systems improves accuracy of rating prediction. *Int. J. Mod. Phys. C* **2021**, *32*, 2150087. [[CrossRef](#)]
26. Kong, Y.X.; Shi, G.Y.; Wu, R.J.; Zhang, Y.C. K-core: Theories and applications. *Phys. Rep.* **2019**, *832*, 1–32. [[CrossRef](#)]
27. Sun, S.; Liu, X.; Wang, L.; Xia, C. New link attack strategies of complex networks based on k-core decomposition. *IEEE Trans. Circuits Syst. II Express Briefs* **2020**, *67*, 3157–3161. [[CrossRef](#)]
28. Wu, Y.; Zhao, J.; Sun, R.; Chen, C.; Wang, X. Efficient Personalized Influential Community Search in Large Networks. *Data Sci. Eng.* **2020**, *6*, 310–322. [[CrossRef](#)]
29. Kitsak, M.; Gallos, L.K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H.E.; Makse, H.A. Identification of influential spreaders in complex networks. *Nat. Phys.* **2010**, *6*, 888–893. [[CrossRef](#)]
30. Morone, F.; Ferraro, G.D.; Makse, H.A. The k-core as a predictor of structural collapse in mutualistic ecosystems. *Nat. Phys.* **2019**, *15*, 95–102. [[CrossRef](#)]
31. Orman, G.K.; Labatut, V. A comparison of community detection algorithms on artificial networks. *Algorithms Artif. Netw.* **2009**, *5805*, 242–256.
32. Guzzetti, F.; Gariano, S.L.; Peruccacci, S. Geographical landslide early warning systems. *Earth-Sci. Rev.* **2019**, *200*, 102973. [[CrossRef](#)]
33. Jaboyedoff, M.; Oppikofer, T.; Abellán, A.; Derron, M.H.; Loye, A.; Metzger, R.; Pedrazzini, A. Use of lidar in landslide investigations: A review. *Nat. Hazards* **2010**, *61*, 5–28. [[CrossRef](#)]
34. Chae, B.G.; Park, H.J.; Catani, F.; Simoni, A.; Berti, M. Landslide prediction, monitoring and early warning: A concise review of state-of-the-art. *Geosci. J.* **2017**, *21*, 1033–1070. [[CrossRef](#)]
35. Phengsuwan, J.; Shah, T.; James, P.; Thakker, D.; Barr, S.; Ranjan, R. Ontology-based discovery of time-series data sources for landslide early warning system. *Computing* **2019**, *102*, 745–763. [[CrossRef](#)]
36. Sufi, F.K.; Alsulami, M. Knowledge Discovery of Global Landslides Using Automated Machine Learning Algorithms. *IEEE Access* **2021**, *9*, 131400–131419. [[CrossRef](#)]
37. Angeli, M.G.; Pasuto, A.; Silvano, S. A critical review of landslide monitoring experiences. *Eng. Geol.* **2010**, *55*, 133–147. [[CrossRef](#)]
38. Zhang, Y.; Hu, X.; Tannant, D.D.; Zhang, G.; Tan, F. Field monitoring and deformation characteristics of a landslide with piles in the three gorges reservoir area. *Landslides* **2018**, *15*, 581–592. [[CrossRef](#)]
39. Aydinoglu, A.C.; Alturk, G. Producing Landslide Susceptibility Maps Using Statistics and Machine Learning Techniques: The Rize-Taslidere Basin Example 1. *J. Geogr.* **2021**, *43*, 159–176.
40. Yang, J.; Liu, X.J.; Qu, Z.; Chang, H. Retrieval keywords complex networks for analyzing legal complexity of stem cell research. *EPL (Europhys. Lett.)* **2020**, *130*, 68001. [[CrossRef](#)]
41. Gan, J.; Cai, Q.; Galer, P.; Ma, D.; Chen, X.; Huang, J.; Bao, S.; Luo, R. Mapping the knowledge structure and trends of epilepsy genetics over the past decade: A co-word analysis based on medical subject headings terms. *Medicine* **2019**, *98*, e16782. [[CrossRef](#)] [[PubMed](#)]
42. Fu, H.; Hu, T.; Wang, J.; Feng, D.; Fang, H.; Wang, M.; Tang, S.; Yuan, F.; Feng, Z. A bibliometric analysis of malaria research in china during 2004–2014. *Malar. J.* **2015**, *14*, 195. [[CrossRef](#)]
43. Hou, Q.; Han, M. Incorporating content beyond text: A high reliable twitter-based disaster information system. *Int. Conf. Comput. Data Soc. Netw.* **2019**, *11917*, 282–292.
44. Musaev, A.; Hou, Q. Gathering High Quality Information on Landslides from Twitter by Relevance Ranking of Users and Tweets. In Proceedings of the IEEE International Conference on Collaboration & Internet Computing, Pittsburgh, PA, USA, 1–3 November 2016.
45. Shehara, P.L.; Siriwardana, C.S.; Amaratunga, D.; Haigh, R. Application of Social Network Analysis (SNA) to Identify Communication Network Associated with Multi-Hazard Early Warning (MHEW) in Sri Lanka. In Proceedings of the 2019 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 3–5 July 2019.
46. Zhang, Y.; Zhu, J.; Zhu, Q.; Xie, Y.; Li, W.; Fu, L.; Zhang, J.; Tan, J. The construction of personalized virtual landslide disaster environments based on knowledge graphs and deep neural networks. *Int. J. Digit. Earth* **2020**, *13*, 1637–1655. [[CrossRef](#)]
47. Gizzi, F.T.; Potenza, M.R. The scientific landscape of november 23rd, 1980 irpinia-basilicata earthquake: Taking stock of (almost) 40 years of studies. *Geosciences* **2020**, *10*, 482. [[CrossRef](#)]
48. Yong, C.; Jinlong, D.; Fei, G.; Bin, T.; Tao, Z.; Hao, F.; Li, W.; Qinghua, Z. Review of landslide susceptibility assessment based on knowledge mapping. *Stoch. Environ. Res. Risk Assess.* **2022**, 1436–3240. [[CrossRef](#)]
49. Karimi-Majd, A.M.; Fathian, M.; Amiri, B. A hybrid artificial immune network for detecting communities in complex networks. *Computing* **2015**, *97*, 483–507. [[CrossRef](#)]
50. Yuan, C.; Rong, C.; Yao, Q. Boundary-connection deletion strategy based method for community detection in complex networks. *Appl. Intell.* **2020**, *50*, 3570–3589. [[CrossRef](#)]
51. Freeman, L.C. Centrality in Social Networks: Conceptual clarification. *Soc. Netw.* **1979**, *1*, 215–239. [[CrossRef](#)]